Quantile Treatment Effects in Difference in Differences Models with Panel Data*

Brantly Callaway[†]
Tong Li[‡]

July 31, 2015

Abstract

The existing literature on Quantile Treatment Effects on the Treated (QTETs) in Difference in Differences models shows that QTETs are either (i) partially identified or (ii) point identified under strong restrictions on the distribution of unobservables over time. We show that having an additional period of pre-treatment data allows for point identification of QTETs under a different set of assumptions that may be more plausible in many applications. Second, we introduce a propensity score re-weighting technique that makes flexible estimation more feasible when the Difference in Differences Assumption holds only conditional on covariates. We compare the performance of our method to existing methods for estimating QTETs in Difference in Differences models using Lalonde (1986)'s job training dataset. Using this dataset, we find the perfomance of our method compares favorably to the performance of existing methods for estimating QTETs.

JEL Codes: C14, C20, C23
Keywords: Quantile Treatment Effect on the Treated, Difference in Differences, Panel Data, Propensity Score Re-weighting

1 Introduction

This paper provides conditions and assumptions needed to identify the Quantile Treatment Effect on the Treated (QTET) under a Distributional Difference in Differences assumption that is analogous to the most commonly used Mean Difference in Differences assumption used to identify the Average Treatment Effect on the Treated (ATT) (Abadie 2005). Importantly, this framework allows empirical researchers to invoke an assumption that is more familiar in applied work and, in many cases, may be more plausible than existing identifying assumptions for the QTET. To identify the ATT, Difference in Differences methods require at least two periods of data that can come from either a panel or repeated cross sections. Previous work that identifies the QTET has maintained the same setup as is needed to identify ATT under a Mean Difference in Differences assumption but has not used an analogous identifying assumption. Instead, previous work has imposed several other assumptions to identify the QTET with the same available data. Although these assumptions are not strictly stronger than the Distributional Difference in Differences assumption made in the current paper, they may be less plausible in many applications and are at least intuitively less familiar to most applied researchers. Fan and Yu (2012) show that the Distributional Difference in Differences assumption made in this paper does not point identify the QTET though it does partially identify the QTET in the conventional Difference in Differences setup. The current paper considers additional data requirements (particularly, at least 3 periods of panel data) and an auxilliary assumption (the Copula Stability Assumption discussed below) that point identify the QTET under the same Distributional Difference in Differences assumption.

Most applied research in economics studies the effect of one variable on the conditional mean of another variable. The reasons for focusing on conditional means include (i) in many cases, the mean may be the most important and easily interpreted feature of the conditional distribution; (ii) if the researcher believes that the effect of treatment is constant across agents (as is implicit, for example, in the linear regression model), then the average effect fully characterizes the distribution of the effect; (iii) its computational simplicity and feasibility in moderate sized datasets commonly used in economics research.

In many cases, however, researchers may be interested in studying effects that "go beyond the mean." In general, this class of treatment effect parameters is called distributional treatment effects, and we study one member of this class, the QTET, in this paper. Studying distributional treatment effects may be important when the effect of treatment is thought to be heterogeneous across agents in the population (Heckman and Robb 1985; Heckman, Smith, and Clements 1997). Distributional treatment effects explicitly recognize the possibility that the same treatment can affect different individuals in different ways. Treatment effect heterogeneity can be important in many contexts. For example, there may be cases where a policy that increases average income by less than an alternative policy may be preferred by policymakers because of differing distributional impacts of each policy. Studying average treatment effects does not allow for this sort of comparison between policies, but studying distributional treatment effects does.

It should be noted that the quantile treatment effects studied in this paper do not correspond to the distribution or quantile of the treatment effect itself. Because treated and untreated outcomes are never simultaneously observed for any individual, the distribution of the treatment effect is not directly identified. For the QTET, the distribution of treated outcomes for the treated group is compared to the counterfactual distribution of untreated outcomes for the treated group. Even when this counterfactual distribution is identified, unless there is some additional assumption on the dependence between these two distributions (Heckman, Smith, and Clements 1997; Fan and Park 2009) or some additional structure placed on the individual's decision on whether or not to be treated (Carneiro, K. Hansen, and Heckman 2003; Abbring and Heckman 2007) the distribution of the treatment effect is not identified. When social welfare evaluations do not depend on the identity of the individual – the anonymity condition – quantile treatment effects provide a complete summary of the welfare effects of a policy (Sen 1997; Carneiro, K. Hansen, and Heckman 2001). On the other hand, parameters that depend on the joint distribution of treated and untreated potential outcomes such as the fraction of the population that benefits from treatment are not identified. Additionally, one could construct examples of treatments that are heterogeneous across all individuals but where the quantile treatment effect is constant.

The focus of this paper is identifying the QTET under a Distributional Difference in Differences assumption. Difference in Differences assumptions are common in applied economic research, but they are most commonly used to estimate the ATT. The intuition underlying the Difference in Differences approach is that, even after possibly controlling for some covariates, treated agents and untreated agents may still differ from each other in unobserved ways that affect the outcome of interest. These differences render cross-sectional comparisons between individuals with the same covariates unable to identify the true effect of treatment. However, if the effect of these unobserved differences on outcomes are constant over time (this is the so-called "parallel trends" assumption), then the researcher can use the difference in the *change* in outcomes for the treated group and the untreated group (rather than differences in the *level* of outcomes) to identify the ATT.

Researchers interested in studying treatment effect heterogeneity have tried to use this same idea to study distributional treatment effects. Unfortunately, the analogy of adjusting last period's average outcome for the treated group by the average change in outcomes for the untreated group does not carry over straightforwardly to studying distributional treatment effects. We discuss why this is the case in more detail below. This problem has led to two different approaches being taken in the literature. The first approach is to make the most straightword Distributional Difference in Differences assumption. In this case, the QTET is partially identified (Fan and Yu 2012). In practice, these bounds can be quite wide. The other approach is to make alternative assumptions on the distribution of unobservables over time that can lead to point identification (Athey and Imbens 2006; Thuysbaert 2007; Bonhomme and Sauder 2011). Each of these papers makes a different set of identifying assumptions. In particular applications, any one of these set of identifying assumptions could potentially hold though, in our view, one limitation of this literature is that the identifying assumptions are not analogous to the Difference in Differences logic used to estimate ATT which

is well known to applied researchers.

The key insight of the current paper is to consider the straightforward Distributional Difference in Differences assumption made by Fan and Yu (2012) and add to their assumption additional data requirements (namely, having three periods of panel data rather than two periods of repeated cross sections) and an additional assumption on the dependence between (i) the distribution of the change in untreated potential outcomes for the treated group and (ii) the initial level of outcomes in the untreated state for the treated group. This combination of assumptions is likely to be more plausible in many applications than the assumptions made elsewhere in the literature and more similar to assumptions that applied researchers are accustomed to making. The data requirements are stronger than in the other papers mentioned above, but three or more periods of panel data may be available in many cases. In these cases, our method may provide a useful way to estimate the QTET.

The second contribution of the paper is to provide a convenient way to accommodate covariates in the Distributional Difference in Differences assumption using a propensity score re-weighting technique. There are many cases where observed characteristics may affect the path of the untreated outcomes. If the distribution of characteristics differs between the treated and untreated groups, then the unconditional "parallel trends" assumption is necessarily violated. One example of this phenomenom is the so-called Ashenfelter's dip (Ashenfelter 1978) where individuals entering a job training program are likely to have experienced a negative transitory shock to wages. Because the shock is transitory, a job training participant's wages are likely to recover even in the absence of job training which implies that using an unconditional Difference in Differences assumption will tend to overstate the effect of the job training program. Conditioning on lags of wages or unemployment histories could help alleviate this problem (Heckman, Ichimura, Smith, and Todd 1998; Heckman and Smith 1999; Abadie 2005). Additionally, if other background characteristics such as education or experience are distributed differently across the treated and untreated groups and the path of wages in the absence of treatment differs by these background characteristics, then an unconditional Difference in Differences assumption will be violated, but a conditional Difference in Differences assumption will be valid.

In contrast, existing methods for estimating QTETs are either (i) unavailable when the researcher desires to make the identifying assumptions conditional on covariates or (ii) require strong parametric assumptions on the relationship between the covariates and outcomes. Because the ATT can be obtained by integrating the QTET and is available under weaker assumptions, a researcher's primary interest in studying the QTET is likely to be in the shape of the QTET rather than the location of the QTET. In this regard, the parametric assumptions required by other methods to accommodate covariates are troubling because nonlinearities or misspecification of the parametric model could easily be confused with the shape of the QTET. This difference between our method and other methods appears to be fundamental. To our knowledge, there is no work on nonparametrically allowing for conditioning on covariates in alternative methods; and, at the least, doing so would be computationally challenging. Moreover, a similar propensity score re-weighting technique

to the one used in the current paper does not appear to be available for existing methods.

There are few empirical papers that have studied the QTET under a Difference in Differences assumption. Meyer, Viscusi, and Durbin (1995) studies the effect of worker's compensation laws on time spent out of work. That paper invokes an unconditional Difference in Differences assumption. To our knowledge, there are no empirical papers that invoke a conditional Difference in Differences assumption to identify the QTET.

Based on our identification results, estimation of the QTET is straightforward and computationally fast. The estimate of the QTET is consistent and $\sqrt{n}$-asymptotically normal. Without covariates, estimating the QTET relies only on estimating unconditional moments, empirical distribution functions, and empirical quantiles. When the identifying assumptions require conditioning on covariates, we estimate the propensity score in a first step. We discuss parametric, semiparametric, and nonparametric estimation of the propensity score which allows for some flexibility for applied researchers in choosing how to implement the method. The computational cost of our method increases with semiparametric or nonparametric estimation of the propensity score, but we show that under standard conditions the speed of convergence of our estimate of the QTET is not affected by the method chosen for the first stage estimation of the propensity score.

We conclude the paper by comparing the performance of our method with alternative estimators of the QTET based on a conditional independence assumption (Firpo 2007) and alternative Difference in Differences methods in an application to estimating the QTET with the job training dataset of LaLonde (1986). This dataset contains an experimental component where individuals were randomly assigned to a job training program and an observational component from the Panel Study of Income Dynamics (PSID). It has been used extensively in the literature to measure how well various observational econometric techniques perform in estimating various treatment effect parameters.

The outline of the paper is as follows. Section 2 provides some background on the notation and setup most commonly used in the treatment effects literature and discusses the various distributional treatment effect parameters estimated in this paper. Section 3 provides our main identification result in the case where the Distributional Difference in Differences assumption holds with no covariates. Section 4 extends this result to the case with covariates and provides a propensity score re-weighting procedure to make estimation more feasible. Section 5 details our estimation strategy and the asymptotic properties of our estimation procedure. Section 6 is the empirical example using the job training data. Section 7 concludes.

2   Background

This section begins by covering some background, notation, and issues in the treatment effects literature. It then discusses the most commonly estimated treatment effects parameters paying particular attention to distributional treatment effect parameters. Finally, we introduce some background on Difference in Differences: (i) the most common parameters estimated using a Dif-

ference in Differences assumption and (ii) the reason why a similar assumption only leads to partial identification of distributional treatment effects.

2.1 Treatment Effects Setup

The setup and notation used in this paper is common in the statistics and econometrics literature. We focus on the case of a binary treatment. Let $D_t = 1$ if an individual is treated at time t (we suppress an individual subscript i throughout to minimize notation). We consider a panel data case where the researcher has access to at least three periods of data for all agents in the sample. We also focus, as is common in the Difference in Differences literature, on the case where no one receives treatment before the final period which simplifies the exposition; a similar result for a subpopulation of the treated group could be obtained with little modification in the more general case. The researcher observes outcomes Y_t, Y_{t-1}, and Y_{t-2} for each individual in each time period. The researcher also possibly observes some covariates X which, as is common in the Difference in Differences setup, we assume are constant over time. This assumption could also be relaxed with appropriate strict exogeneity conditions.

Following the treatment effects literature, we assume that individuals have potential outcomes in the treated or untreated state: Y_{1t} and Y_{0t}, respectively. The fundamental problem is that exactly one (never both) of these outcomes is observed for a particular individual. Using the above notation, the observed outcome Y_t can be expressed as follows:

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

For any particular individual, the unobserved potential outcome is called the counterfactual. The individual's treatment effect, $Y_{1t} - Y_{0t}$ is therefore never available because only one of the potential outcomes is observed for a particular individual. Instead, the literature has focused on identifying and estimating various functionals of treatment effects and the assumptions needed to identify them. We discuss some of these treatment effect parameters next.

2.2 Common Treatment Effect Parameters and Identifying Assumptions

The most commonly estimated treatment effect parameters are the Average Treatment Effect (ATE) and the Average Treatment Effect on the Treated (ATT).[1] The unconditional on covariates versions of these are given below:

$$ATE = E[Y_{1t} - Y_{0t}]$$
$$ATT = E[Y_{1t} - Y_{0t}|D_t = 1]$$

---

[1] There are more treatment effect parameters such as the Local Average Treatment Effect (LATE) of Imbens and Angrist (1994) and the Marginal Treatment Effect (MTE) and Policy Relevant Treatment Effect (PRTE) of Heckman and Vytlacil (2005). Heckman, LaLonde, and Smith (1999) and Heckman and Vytlacil (2005) also discuss conditions when various parameters are of interest.

It is also common to estimate versions of ATE and ATT conditional on covariates X. The unconditional ATE and ATT can then be obtained by integrating out X. The parameters provide a summary measure of the average effect of treatment for a random individual in the population (ATE) or for an individual from the subgroup of the population that is treated (ATT).

Various assumptions can be used to identify ATE and ATT. These include random treatment assignment, selection on observables, instrumental variables, and Difference in Differences. Difference in Differences methods identify the ATT, but not the ATE. See Imbens and Wooldridge (2009) for an extensive review.

2.3 Quantiles and Quantile Treatment Effects

In cases where (i) the effect of a treatment is thought to be heterogeneous across individuals and (ii) understanding this heterogeneity is of interest to the researcher, estimating distributional treatment effects such as quantile treatment effects is likely to be important. For example, the empirical application in this paper considers the effect of a job training program on wages. If the researcher is interested in the effect of participating in the job training program on low wage individuals, studying the quantile treatment effect is more useful than studying the average effect of the job training program. Our analysis is consistent with the idea that the effect of a job training program on wages differs between relatively high wage individuals and relatively low wage individuals.

For a random variable X, the τ-quantile, x_τ, of X is defined as

$$x_\tau = G_X^{-1}(\tau) \equiv \inf\{x : F_X(x) \geq \tau\} \tag{1}$$

When X is continuously distributed, x_τ satisfies $P(X \leq x_\tau) = \tau$. An example is the 0.5-quantile – the median.[2] Researchers interested in program evaluation may be interested in other quantiles as well. In the case of the job training program, researchers may be interested in the effect of job training on low income individuals. In this case, they may study the 0.05 or 0.1-quantile. Similarly, researchers studying the effect of a policy on high earners may look at the 0.99-quantile.

Let $F_{Y_{1t}}(y)$ and $F_{Y_{0t}}(y)$ denote the distributions of Y_{1t} and Y_{0t}, respectively. Then, the Quantile Treatment Effect (QTE) is defined as

$$\text{QTE}(\tau) = F_{Y_{1t}}^{-1}(\tau) - F_{Y_{0t}}^{-1}(\tau) \tag{2}$$

Analogously to the case of identifying the ATE, QTE is not directly identified because the researcher cannot simultaneously observe Y_{1t} and Y_{0t} for any individual. When treatment is randomized, each distribution will be identified and the quantiles can be recoverd. Similarly, selection on observables also identifies QTE because the marginal distributions of Y_{1t} and Y_{0t} are identified (Firpo 2007).[3]

---

[2]In this paper, we study Quantile Treatment Effects. A related topic is quantile regression. See Koenker (2005).

[3]There are also several papers that identify versions of QTE when the researcher has an available instrument.

Researchers may also be interested in identifying the Quantile Treatment Effect on the Treated (QTET) defined by

$$\text{QTET}(\tau) = F^{-1}_{Y_{1t}|D_t=1}(\tau) - F^{-1}_{Y_{0t}|D_t=1}(\tau) \tag{3}$$

The QTET is the parameter studied in this paper. Difference in Differences methods are useful for studying treatment effect parameters for the treated group because they make use of observing untreated outcomes for the treated group in a time period before they become treated.

2.4 Partial Identification of the Quantile Treatment Effect on the Treated under a Distributional Difference in Differences Assumption

The most common nonparametric assumption used to identify the ATT in Difference in Differences models is the following:

Assumption 1 (Mean Difference in Differences)**.**

$$\text{E}[\Delta Y_{0t}|D_t = 1] = \text{E}[\Delta Y_{0t}|D_t = 0]$$

This is the "parellel trends" assumptions common in applied research. It states that, on average, the unobserved change in untreated potential outcomes for the treated group is equal to the observed change in untreated outcomes for the untreated group. To study the QTET, Assumption 1 needs to be strengthened because the QTET depends on the entire distribution of untreated outcomes for the treated group rather than only the mean of this distribution.

The next assumption due to Fan and Yu (2012) strengthens Assumption 1 and this is the assumption maintained throughout the paper.

Distributional Difference in Differences Assumption.

$$\text{P}(\Delta Y_{0t} \leq \Delta y|D_t = 1) = \text{P}(\Delta Y_{0t} \leq \Delta y|D_t = 0)$$

The Distributional Difference in Differences Assumption says that the distribution of the change in potential untreated outcomes does not depend on whether or not the individual belongs to the treatment or the control group. Intuitively, it generalizes the idea of "parallel trends" holding on average to the entire distribution. In applied work, the validity of using a Difference in Differences approach to estimate the ATT hinges on whether the unobserved trend for the treated group can be replaced with the observed trend for the untreated group. This is exactly the same sort of thought experiment that needs to be satisfied for the Distributional Difference in Differences Assumption to hold. Being able to invoke a standard assumption to identify the QTET stands in contrast to the existing literature on identifying the QTET in similar models which generally require less familiar assumptions on the relationship between observed and unobserved outcomes.

See Abadie, Angrist, and Imbens (2002), Chernozhukov and C. Hansen (2005), and Froelich and Melly (2013).

Using statistical results on the distribution of the sum of two known marginal distributions, Fan and Yu (2012) show that this assumption is not strong enough to point identify the counterfactual distribution $F_{Y_{0t}|D_t=1}(y)$, but it does partially identify it.[4] The resulting bounds are given by

$$F_{Y_{0t}|D_t=1}(s) \leq 1 + \min\left[\inf_y F_{(Y_{0t}-Y_{0t-1})|D_t=1}(y) + F_{Y_{0t-1}|D_t=1}(s-y) - 1, 0\right]$$

$$F_{Y_{0t}|D_t=1}(s) \geq \max\left[\sup_y F_{(Y_{0t}-Y_{0t-1})|D_t=1}(y) + F_{Y_{0t-1}|D_t=1}(s-y) - 1, 0\right] \tag{4}$$

One can show that these bounds are sharp. In other words, there exist dependence stuctures between the two marginal distributions so that the bounds $F_{Y_{0t}|D_t=1}(y)$ obtains either its upper or lower bound. This also means that one cannot improve these bounds without additional assumptions or restrictions on the data generating process. These bounds lead to bounds on the counterfactual quantiles of untreated potential outcomes for the treated group; which, in turn, leads to bounds on the QTET. In the next section, we provide one set of additional assumptions (and data requirements) that point identifies QTET and may be plausible in many cases.

3   Main Results: Identifying QTET in Difference in Differences Models

The main results in this section deal with point identification of QTET under a Distributional Difference in Differences assumption. Existing papers that point- or partially-identify the QTET include Athey and Imbens (2006), Thuysbaert (2007), Bonhomme and Sauder (2011), and Fan and Yu (2012). In general, these papers require stronger (or at least less intuitively familiar) distributional assumptions than are made in the current paper while requiring access to only two periods of repeated cross section data.

The main theoretical contribution of this paper is to impose a Distributional Difference in Differences Assumption plus additional data requirements and an additional assumption that may be plausible in many applications to identify the QTET. The additional data requirement is that the researcher has access to at least three periods of panel data with two periods preceding the period where individuals may first be treated. This data requirement is stronger than is typical in most Difference in Differences setups which usually only require two periods of repeated cross-sections (or panel) data. The additional assumption is that the dependence between (i) the distribution of

---

[4]More specifically, Fan and Yu (2012) write $F_{Y_{0t}|D_t=1}(y) = F_{\Delta Y_{0t}+Y_{0t-1}|D_t=1}(y) = g(F_{\Delta Y_{0t},Y_{0t-1}|D_t=1}(\Delta y, y))$ where $g(\cdot)$ is a known function of the joint distribution between the change in untreated potential outcomes and initial untreated potential outcome for the treated group. Under the Distributional Difference in Differences Assumption, the unknown distribution $F_{\Delta Y_{0t}|D_t=1}(\Delta y) = F_{\Delta Y_{0t}|D_t=0}(\Delta y)$ which is identified, and $F_{Y_{0t-1}|D_t=1}(y)$ is identified directly by the sampling process. This shows that $F_{Y_{0t}|D_t=1}(y)$ is function of an unknown joint distribution with known marginals which leads to partial identification. In the case where a researcher is only interested in the counterfactual mean, Abadie (2005) uses the fact that the sum of the two distributions does not depend on the joint distribution; rather it depends only on each known marginal distribution, and therefore the counterfactual mean can be identified.

$(\Delta Y_{0t}|D_t = 1)$ (the change in the untreated potential outcomes for the treated group) and (ii) the distribution of $(Y_{0t-1}|D_t = 1)$ (the initial untreated outcome for the treated group) is stable over time. This assumption does not say that these distributions themselves are constant over time; instead, only the dependence between the two marginal distributions is constant over time. We discuss this assumption in more detail and show how it can be used to point identify the QTET below.

Intuitively, the reason why a restriction on the dependence between the distribution of $(\Delta Y_{0t}|D_t = 1)$ and $(Y_{0t-1}|D_t = 1)$ is useful is the following. If the joint distribution $(\Delta Y_{0t}, Y_{0t-1}|D_t = 1)$ were known, then $F_{Y_{0t}|D_t=1}(y_{0t})$ (the distribution of interest) could be derived from it. The marginal distributions $F_{\Delta Y_{0t}|D_t=1}(\Delta y_{0t})$ (through the Distributional Difference in Differences assumption) and $F_{Y_{0t-1}|D_t=1}(y_{0t-1})$ (from the data) are both identified. However, because observations are observed separately for untreated and treated individuals, even though each of these marginal distributions are identified from the data, the joint distribution is not identified. Since, from Sklar's Theorem (Sklar 1959), joint distributions can be expressed as the copula function (capturing the dependence) of the two marginal distributions, the only piece of information that is missing is the copula.[5] We use the idea that the dependence is the same between period t and period $(t-1)$. With this additional information, we can show that $F_{Y_{0t}|D_t=1}(y_{0t})$ is identified.

The time invariance of the dependence between $F_{\Delta Y_{0t}|D_t=1}(\Delta y)$ and $F_{Y_{0t-1}|D_t=1}(y)$ can be expressed in the following way. Let $\mathrm{F}_{\Delta Y_{0t}, Y_{0t-1}|D_t=1}(\Delta y, y)$ be the joint distribution of $(\Delta Y_{0t}|D_t = 1)$ and $(Y_{0t-1}|D_t = 1)$. By Sklar's Theorem

$$\mathrm{F}_{\Delta Y_{0t}, Y_{0t-1}|D_t=1}(\Delta y, y) = C_{\Delta Y_{0t}, Y_{0t-1}|D_t=1}\left(F_{\Delta Y_{0t}|D_t=1}(\Delta y), F_{Y_{0t-1}|D_t=1}(y)\right)$$

where $C_{\Delta Y_{0t}, Y_{0t-1}|D_t=1}(\cdot, \cdot)$ is a copula function.[6] Next, we state the second main assummption which replaces the unknown copula with copula for the same outcomes but in the previous period which is identified because no one is treated in the periods before t.

Copula Stability Assumption.

$$C_{\Delta Y_{0t}, Y_{0t-1}|D_t=1}(\cdot, \cdot) = C_{\Delta Y_{0t-1}, Y_{0t-2}|D_t=1}(\cdot, \cdot)$$

The Copula Stability Assumption says that the dependence between the marginal distributions $F_{\Delta Y_{0t}|D_t=1}(\Delta y)$ and $F_{Y_{0t-1}|D_t=1}(y)$ is the same as the dependence between the distributions $F_{\Delta Y_{0t-1}|D_t=1}(\Delta y)$ and $F_{Y_{0t-2}|D_t=1}(y)$. It is important to note that this assumption does not require any *particular* dependence structure between the marginal distributions; rather, it requires that whatever the dependence structure is in the past, one can recover it and reuse it in the current period. It also does not require choosing any parametric copula. However, it may be helpful to consider a simple, more parametric example. If the copula of the distribution of $(\Delta Y_{0t-1}|D_t = 1)$

---

[5]Joe (1997), Nelsen (2007), and Joe (2015) are useful references for more details on copulas.
[6]The bounds in Fan and Yu (2012) arise by replacing the unknown copula function $C_{\Delta Y_{0t}, Y_{0t-1}|D_t=1}(\cdot, \cdot)$ with those that make the upper bound the largest and lower bound the smallest.

and the distribution of $(Y_{0t-2}|D_t = 1)$ is Gaussian with parameter ρ, the Copula Stability Assumption says that the copula continues to be Gaussian with parameter ρ in period t but the marginal distributions are allowed to change in unrestricted ways. Likewise, if the copula is Archimedean, the Copula Stability Assumption requires the generator function to be constant over time but the marginal distributions can change in unrestricted ways.

One of the key insights of this paper is that, in some particular situations such as the panel data case considered in the paper, we are able to observe the historical dependence between the marginal distributions. There are many applications in economics where the missing piece of information for identification is the dependence between two marginal distributions. In those cases, previous research has resorted to (i) assuming some dependence structure such as independence or perfect positive dependence or (ii) varying the copula function over some or all possible dependence structures to recover bounds on the joint distribution of interest. To our knowledge, we are the first to use historical observed outcomes to obtain a historical dependence structure and then assume that the dependence structure is stable over time.

Before presenting the identification result, we need some additional assumptions.

Assumption 2. *Let $\Delta \mathcal{Y}_{t|D_t=0}$ denote the support of the change in untreated outcomes for the untreated group. Let $\Delta \mathcal{Y}_{t-1|D_t=1}$, $\mathcal{Y}_{t-1|D_t=1}$, and $\mathcal{Y}_{t-2|D_t=1}$ denote the support of the change in untreated outcomes for the treated group in period $(t-1)$, the support of untreated outcomes for the treated group in period $(t-1)$, and the support of untreated outcomes for the treated goup in period $(t-2)$, respectively. We assume that*
   (a) $\Delta \mathcal{Y}_{t|D_t=0} \subseteq \Delta \mathcal{Y}_{t-1|D_t=1}$
   (b) $\mathcal{Y}_{t-1|D_t=1} \subseteq \mathcal{Y}_{t-2|D_t=1}$

Assumption 3. *Conditional on $D_t = d$, the observed data $(Y_{dt,i}, Y_{t-1,i}, Y_{t-2,i}, X_i)$ are independently and identically distributed.*

Assumption 4. *(Distribution of Y)*
   Each of the random variables ΔY_t for the untreated group and ΔY_{t-1}, Y_{t-1}, and Y_{t-2} for the treated group are continuously distributed on a compact support with densities that are bounded from above and bounded away from 0. The densities are also continuously differentiable and the derivative of each of the densities is bounded.

Theorem 1. *Under the Distributional Difference in Differences Assumption, the Copula Stability Assumption, Assumption 2, Assumption 3, and Assumption 4*

$$\begin{aligned}
&\mathrm{P}(Y_{0t} \leq y | D_t = 1) \\
&\quad = \mathrm{E}\left[\mathbb{1}\{ F^{-1}_{\Delta Y_{0t}|D_t=0}(F_{\Delta Y_{0t-1}|D_t=1}(\Delta Y_{0t-1})) \leq y - F^{-1}_{Y_{0t-1}|D_t=1}(F_{Y_{0t-2}|D_t=1}(Y_{0t-2})) \} | D_t = 1 \right]
\end{aligned}$$
(5)

Theorem 1 is the main identification result of the paper. It says that the counterfactual distribution of untreated outcomes for the treated group is identified. To provide some intuition, we provide

a short outline of the proof. First, notice that $P(Y_{0t} \leq y|D_t = 1) = E[\mathbb{1}\{\Delta Y_{0t} + Y_{0t-1} \leq y\}|D_t = 1]^7$
But ΔY_{0t} is not observed for the treated group because Y_{0t} is not observed. The Copula Stability Assumption effectively allows us to look at observed outcomes in the previous periods for the treated group and "adjust" them forward. Finally, the Distributional Difference in Differences Assumption allows us to replace $F^{-1}_{\Delta Y_{0t}|D_t=1}(\cdot)$ with $F^{-1}_{\Delta Y_{0t}|D_t=0}(\cdot)$ which is just the quantiles of the distribution of the change in (observed) untreated outcomes for the untreated group.

It can be estimated by plugging in the sample counterparts of the terms on the right hand side of Equation 5:

$$\frac{1}{n_T} \sum_{i \in \mathcal{T}} \left[\mathbb{1}\{\hat{F}^{-1}_{\Delta Y_{0t}|D_t=0}(\hat{F}_{\Delta Y_{0t-1}|D_t=1}(\Delta Y_{0t-1,i})) \leq y - \hat{F}^{-1}_{Y_{0t-1}|D_t=1}(\hat{F}_{Y_{0t-2}|D_t=1}(Y_{0t-2,i}))\} \right] \qquad (6)$$

This will be consistent and $\sqrt{n}$-asymptotically normal under straightforward conditions. Once this distribution is identified, we can easily use it to estimate its quantiles. We discuss more details of estimation in Section 6.

How to Interpret the Copula Stability Assumption

The Copula Stability Assumption is new to the treatment effect literature. As such, it is important to understand what models are compatible with the assumption. In this section, we show that a very general model of untreated potential outcomes for the treated group satisfies the Copula Stability Assumption. Consider the model

$$Y_{0it} = h(X_{it}, U_{it}) + \tau_t \qquad (7)$$
$$Y_{0it-1} = h(X_{it-1}, U_{it-1}) + \tau_{t-1}$$

where $h(\cdot)$ is a possibly nonseparable function of observable individual-specific covariates X_{is} and unobservables $U_{is} = \varsigma_i + \nu_{is}$ where ς_i is a vector of time invariant unobservables and ν_{is} are time varying unobservables. We assume that ν_{is} either (i) follows a unit root: $\nu_{is} = \nu_{is-1} + \epsilon_{is}$ with $\epsilon_{is} \perp\!\!\!\perp \nu_{is-1}$ and $\epsilon_{is} \sim_{iid} F_\epsilon(\cdot)$ or (ii) follow an AR(1) process: $\nu_{is} = \rho\nu_{is-1} + \epsilon_{is}$ with $-1 < \rho < 1$ and $\epsilon_{is} \perp\!\!\!\perp \nu_{is-1}$ and $\epsilon_{is} \sim_{iid} F_\epsilon(\cdot)$.

Proposition 1. *In the model of Equation 7 (Case 1 or Case 2), the Copula Stability Assumption is satisfied*

Proposition 1 is an important result because it says that the Copula Stability Assumption will hold in a wide variety of the most common econometric models.

This model generalizes many common econometric models. It allows for non-stationarity in outcomes. For example, in the empirical application on job training, aggregate time effects such as macroeconomic shocks are allowed in the model. Several other common models are special cases of

---

[7]Adding and substracting Y_{0t-1} is also the first step for showing that the Mean Difference in Differences Assumption identifies $E[Y_{0t}|D_t = 1]$.

this model. For example, the result covers the two-way fixed effects models with individual specific fixed effects and aggregate time fixed effects.

$$Y_{0it} = c_i + \theta_t + X_{it}\beta + \epsilon_{it}$$

where c_i is a time invariant fixed effect, θ_t is an aggregate time effect for the treated group, and ϵ_{it} is white noise. This result also covers a special case of the random trend model (Heckman and Hotz 1989).[8]

$$Y_{0it} = c_i + gt + X_{it}\beta + \epsilon_{it}$$

where c_i is a time invariant fixed effect, g_i is a random coefficient on a time trend, and ϵ_{it} white noise. As a final example, this result covers a unit root process

$$Y_{0it} = Y_{0it-1} + \epsilon_{it}$$

with ϵ_{it} white noise. Other models are also covered by Proposition 1.

It should be noted that the model in Equation 7 also provides a useful example for the case where the Conditional Distributional Difference in Differences Assumption (discussed in the next section) can combine with the Copula Stability Assumption for identification. Suppose that Equation 7 also generate outcomes for the untreated group and also suppose that all covariates are time invariant. Then, $F_{\Delta Y_{0t}|D_t=d}(\Delta) = P(h(X, U_{it}) - h(X, U_{it-1}) + \tau_t - \tau_{t-1} \leq \Delta)$. If the distribution of X is not the same for the treated and untreated groups, then the Distributional Difference in Differences Assumption will not hold but the Conditional Distributional Difference in Differences Assumption will hold.

Pre-Testing the Assumptions   Neither the Distributional Difference in Differences Assumption nor the Copula Stability Assumption are directly testable; however, the applied researcher can provide some additional tests to provide some evidence that the assumptions are more or less likely to hold.

The Copula Stability Assumption would be violated if the relationship between the change in untreated potential outcomes and the initial untreated potential outcome is changing over time. This is an untestable assumption. However, in the spirit of pre-testing in Difference in Differences models, with four periods of data, one could use the first two periods to construct the copula function for the third period; then one could compute the actual copula function for the third period using the data and check if they are the same. This would provide some evidence that the

---

[8]The Conditional Distributional Difference in Differences Assumption does not hold in the more general case where g is allowed to be individual specific g_i. To provide some intuition, consider the case where $Y_i = c_i + g_i t$. In this case the distribution of the change in outcomes is constant over time – it is just given by the distribution of g_i, and an individual's rank in the distribution remains the same over time. However, individual's with large values of g_i will increase their rank in the level of the outcome. This means dependence between the change in untreated potential outcomes and initial level of untreated potential outcome will increase over time

copula function is stable over time.

Additionally, the Distributional Difference in Differences Assumption is untestable though a type of pre-testing can also be done for this assumption. Using data from the previous period, the researcher can estimate both of the following distributions: $F_{\Delta Y_{0t-1}|D_t=1}(\Delta y)$ and $F_{\Delta Y_{0t-1}|D_t=0}(\Delta y)$. Then, one can check if the distributions are equal using, for example, a Kolmogorov-Smirnoff type test. This procedure does not provide a test that the Distributional Difference in Differences Assumption is valid, but when the assumption holds in the previous period, it does provide some evidence that that the assumption is valid in the period under consideration. Unlike the pre-test for the Copula Stability Assumption mentioned above, this pre-test of the Distributional Difference in Differences Assumption does not require access to additional data because three periods of data are already required to implement the method.

4   Allowing for covariates

The second theoretical contribution of this paper is to make identification of the QTET more feasible in applications. In many economic applications, Distributional Difference in Differences Assumption is likely to hold only after conditioning on some covariates X, and therefore accommodating covariates is important to make the type of method discussed in this paper viable in empirical applications (Abadie 2005).

For example, Fan and Yu (2012) suggest estimating conditional distributions, integrating out X, and then inverting to obtain the unconditional quantiles. Following this type of procedure is likely to be quite difficult to implement and computationally challenging especially when the dimension of X is greater than two or three. Alternatively, Athey and Imbens (2006) suggest specifying a parametric model and then performing a type of residualization to recover the QTET. Though this type of procedure is likely to be feasible in applications, using a linear model is likely to be unsatisfactory for studying treatment effect heterogeneity because nonlinearities or model misspecification are likely to be confused with the shape of the QTET.

In this section, we propose a propensity score re-weighting estimator similar to Abadie (2005) in the case of Mean Difference in Differences and to Firpo (2007) in the case of Quantile Treatment Effects under selection on observables. This procedure allows the researcher to estimate the propensity score in a first stage and then re-weight observations based on the propensity score as an intermediate step to estimating the QTET. This type of propensity score re-weighting technique does not appear to be available in the case of other available methods to estimate the QTET under some type of Difference in Differences assumption.

Using a propensity score re-weighting technique also gives the researcher some flexibility in choosing the best way implement our method. The propensity score can be specified parametrically which requires strong functional form assumptions but is easy to compute and feasible in medium sized samples. At the other extreme, the propensity score could be estimated nonparametrically without invoking functional form assumptions but is more difficult to compute and may suffer

from slower convergence depending on the assumptions on the smoothness of the propensity score. Finally, semiparametric methods are available such as Ichimura (1993) and Klein and Spady (1993) that offer some additional flexibility relative to parametric models and computational advantages relative to nonparametric methods.

It should be noted that interest still centers on the unconditional QTET rather than a QTET conditional on X. The role of the covariates is to make the Distributional Difference in Differences Assumption valid. One reason for this focus is that the unconditional QTET is easily interpreted while a conditional QTET may be difficult to interpret and estimate especially when X contains a large number of variables. Though we do not explicitly consider the case, our method could easily be adapted to the case where a researcher is interested in QTETs conditional on X_k where $X_k \subset X$. One example where this approach could be useful is in evaluating the distributional impacts separately by gender.

Conditional Distributional Difference in Differences Assumption.

$$\Delta Y_{0t} \perp\!\!\!\perp D_t | X$$

After conditioning on covariates X, the distribution of the change in untreated potential outcomes for the treated group is equal to the change in untreated potential outcomes for the untreated group.

By invoking the Conditional Distributional Difference in Differences Assumption rather than the Distributional Difference in Differences Assumption, it is important to note that, for the purpose of identification, the only part of Theorem 1 that needs to be adjusted is the identification of $F_{\Delta Y_{0t}|D_t=1}(\Delta y)$. Under the Distributional Difference in Differences Assumption, this distribution could be replaced directly by $F_{\Delta Y_{0t}|D_t=0}(\Delta y)$; however, now we utilize a propensity score re-weighting technique to replace this distribution with another object (discussed more below). Importantly, all other objects in Theorem 1 can be handled in exactly the same way as they were previously. Particularly, the Copula Stability Assumption continues to hold without needing any adjustment such as conditioning on X. The Copula Stability Assumption is an assumption on the dependence between $F_{Y_{0t-1}|D_t=1}(y)$ (which is observed) and $F_{\Delta Y_{0t}|D_t=1}(\Delta y)$ which we next show is identified under Conditional Distributional Difference in Differences Assumption. With these two distributions in hand, which do not depend on X, we can once again invoke the same Copula Stability Assumption to obtain identification in the same way as Theorem 1.

One thing to notice is that we have written the Conditional Distributional Difference in Differences Assumption in terms of time invariant covariates X, but the assumption can be extended to the case where covariates can change over time denoted X_{it} under standard assumptions. In particular, this extension would require a strict exogeneity assumption such as $P(Y_{0is} \leq y|X_i, c_i) = P(Y_{0is} \leq y|X_{it}, c_i)$ where s represents a generic time period, $X_i = (X_{it}, X_{it-1}, X_{it-2})$ is the vector of covariates across all periods, and c_i is an individual specific fixed effect. The strict exogeneity assumption says that conditional on the individual fixed effect and current period values of the covariates, the distribution of untreated potential outcomes does not depend

on the value of covariates in other periods. Under the strict exogeneity assumption, a natural version of the Conditional Distributional Difference in Differences Assumption would be $P(\Delta Y_{0it} \leq \Delta | X_i, c_i, D_t = 1) = P(\Delta Y_{0it} \leq \Delta | X_{it}, X_{it-1}, D_t = 1)$. Then, in the subsequent analysis, one should replace X with $\tilde{X} = (X_C, X_{it}, X_{it-1})$ where X_C are the covariates that do not change over time. We proceed throughout the rest of the paper, however, using the notation for time invariant X.

We also require several additional standard assumptions for identification. We state these first.

Assumption 5. $p \equiv P(D_t = 1) > 0$ *and* $p(x) \equiv P(D_t = 1 | X = x) < 1$.

This assumption says that there is some positive probability that individuals are treated, and that for an individual with any possible value of covariates x, there is some positive probability that he will be treated and a positive probability he will not be treated.

Theorem 2. *Under Copula Stability Assumption, Assumption 2, Assumption 3, and Assumption 4, Conditional Distributional Difference in Differences Assumption, and Assumption 5*

$$P(Y_{0t} \leq y | D_t = 1)$$
$$= E\left[\mathbb{1}\{F^{-1}_{\Delta Y_{0t}|D_t=1}(F_{\Delta Y_{0t-1}|D_t=1}(\Delta Y_{0t-1})) \leq y - F^{-1}_{Y_{0t-1}|D_t=1}(F_{Y_{0t-2}|D_t=1}(Y_{0t-2}))\} | D_t = 1\right]$$

where

$$F_{\Delta Y_{0t}|D_t=1}(y) = E\left[\frac{1 - D_t}{p}\frac{p(X)}{1 - p(X)}\mathbb{1}\{\Delta Y_t \leq \Delta y\} | D_t = 0\right] \tag{8}$$

This result is very similar to the main identification result in Theorem 1. The only difference is that $F_{\Delta Y_{0t}|D_t=1}(\cdot)$ is no longer identified by the distribution of untreated potential outcomes for the untreated group; instead, it is replaced by the re-weighted distribution in Equation 8. Equation 8 can be understood in the following way. It is a weighted average of the distribution of the change in outcomes experienced by the untreated group. The $\frac{p(X)}{1 - p(X)}$ term weights up untreated observations that have covariates that make them more likely to be treated. Equation 8 is almost exactly identical to the re-weighting estimators given in Hirano, Imbens, and Ridder (2003), Abadie (2005), and Firpo (2007); the only difference is the term $\mathbb{1}\{\Delta Y_t \leq \Delta y\}$ in our case is given by Y_t, ΔY_t, and $\mathbb{1}\{Y_t \leq y\}$ in each of the other cases, respectively.

This moment can be easily estimated in two steps: (i) estimate the propensity score to obtain $\hat{p}(x)$ and (ii) plug in the estimated propensity score into the sample analog of the moment:

$$\hat{P}(\Delta Y_{0t} \leq \Delta y | D_t = 1) = \frac{1}{n}\sum_{i=1}^{n}\left[\frac{(1 - D_t)}{p}\frac{\hat{p}(X_i)}{1 - \hat{p}(X_i))}\mathbb{1}\{\Delta Y_{it} \leq \Delta y\}\right] \tag{9}$$

This analog of the distribution of the change in untreated potential outcomes for the treated group can then be combined with estimates of the other distributions in Theorem 2 to estimate the QTET.

5 Estimation Details

In this section, we outline the estimation procedure. Then, we provide results on consistency and asymptotic normality of the estimators.

We estimate

$$\hat{\text{QTET}}(\tau) = \hat{F}^{-1}_{Y_{1t}|D_t=1}(\tau) - \hat{F}^{-1}_{Y_{0t}|D_t=1}(\tau)$$

The first term is estimated directly from the data using the order statistics of the treated outcome for the treated group.

$$\hat{F}^{-1}_{Y_{1t}|D_t=1}(\tau) = Y_{t|D_t=1}(\lceil n_T \tau \rceil)$$

where $X(k)$ is the kth order statistic of $X_1, \dots, X_n$, n_T is the number of treated observations, and the notation $\lceil s \rceil$ rounds s up to the closest, larger integer.

The estimator for $\hat{F}^{-1}_{Y_{0t}|D_t=1}(\tau)$ is more complicated. The distribution $\hat{F}_{Y_{0t}|D_t=1}(y_{0t})$ is identified by Distributional Difference in Differences Assumption or as in Theorem 2 depending on the situation. We use this result to provide an estimator of the quantiles of that distribution in the following way:

$$\hat{F}^{-1}_{Y_{0t}|D_t=1}(\tau) = \left\{ \hat{F}^{-1}_{\Delta Y_{0t}|D_t=1} \left(\hat{F}_{\Delta Y_{0t-1}|D_t=1}(\Delta Y_{0t-1|D_t=1}) \right) \right.$$
$$\left. + \hat{F}^{-1}_{Y_{0t-1}|D_t=1} \left(\hat{F}_{Y_{0t-2}|D_t=1}(Y_{0t-2|D_t=1}) \right) \right\} (\lceil n_T \tau \rceil)$$

Here, once again, we compute the quantiles of $(Y_{0t}|D_t = 1)$ using order statistics, but now they must be adjusted. We plug in estimates of the quantiles and distribution functions for the distributions in Theorem 1. It should be noted the order statistics are taken for the treated group (after adjusting the values based on the sample quantiles and distributions noted above).

The sample quantiles that serve as an input into estimating $F^{-1}_{Y_{0t}|D_t=1}(\tau)$ are estimated with the order statistics (with one exception mentioned below). The sample distributions are estimated using the empirical distribution:

$$\hat{F}_X(x) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}\{X_i \le x\}$$

The final issue is estimating $F^{-1}_{\Delta Y_{0t}|D_t=1}(\nu)$ when identification depends on covariates as in Section 4. Using the identification result in Section 4, we can easily construct an estimator of the distribution function

$$\hat{F}_{\Delta Y_{0t}|D_t=1}(\Delta y_{0t}) = \frac{1}{n} \sum_{i=1}^{n} \frac{(1 - D_{it})}{p} \frac{\hat{p}(X_i)}{(1 - \hat{p}(X_i))} \mathbb{1}\{\Delta Y_{t,i} \le \Delta y_{0t}\}$$

Then, an estimator of $F^{-1}_{\Delta Y_{0t}|D_t=1}(\nu)$ can be obtained in the following way. Let $\Delta Y_{t,i}(n)$ denote the ordered values of the change in outcomes from smallest to largest, and let $\Delta Y_{t,i}(j)$ denote the jth value of $\Delta Y_{t,i}$ in the ordered sequence. Then, $\hat{F}^{-1}_{\Delta Y_{0t}|D_t=1}(\nu) = \Delta Y_{t,i}(J^*)$ where $J^* = \inf\{J : \frac{1}{n}\sum_{j=1}^{J}\frac{(1-D_{jt})}{p}\frac{\hat{p}(X_j)}{(1-\hat{p}(X_j))} \geq \nu\}$.

When identification depends on covariates X, then there must be a first step estimation of the propensity score. In applied work, there are several possibilities for researchers to consider: (i) parametric propensity score, (ii) semi-parametric propensity score, and (iii) nonparametric propensity score. The tradeoff between these three involves trading off stronger assumptions (the parametric case) for more challenging computational issues (the nonparametric case). Below we show consistency and asymptotic normality for the parametric and nonparametric cases; additional results for the semiparametric case are available upon request. The estimator is $\sqrt{n}$-consistency and asymptotically normal in each case even though the propensity score itself converges at a slower than $\sqrt{n}$-rate when it is estimated nonparametrically. We also implement both approaches in the empirical application.

5.1 Inference

This section considers the asymptotic properties of the estimator. First, it focuses on the case with no covariates and then extends the results to the case where the Distributional Difference in Differences Assumption holds conditional on covariates. The proof for each of the theorems in this section is given in the Appendix.

5.1.1 No Covariates Case

In the case with no covariates, the following result holds

Theorem 3. *Consistency under Distributional Difference in Differences Assumption*
 Under Assumption 2, Assumption 3, and Assumption 4

$$Q\hat{T}ET(\tau) = \hat{F}^{-1}_{Y_{1t}|D_t=1}(\tau) - \hat{F}^{-1}_{Y_{0t}|D_t=1}(\tau) \xrightarrow{p} F^{-1}_{Y_{1t}|D_t=1}(\tau) - F^{-1}_{Y_{0t}|D_t=1}(\tau) = QTET(\tau)$$

To show asymptotic normality, we introduce some additional notation. Let

$$\mu_{12}(z;y) = E_{Y_{0t-2}|D_t=1}\left[\left(\mathbb{1}\{z \leq (y - F^{-1}_{Y_{0t-1}|D_t=1}(F_{Y_{0t-2}|D_t=1}(Y_{0t-2})))\}\right.\right.$$
$$\left.\left. - F_{\Delta Y_{0t}|D_t=0}(y - F^{-1}_{Y_{0t-1}|D_t=1}(F_{Y_{0t-2}|D_t=1}(Y_{0t-2})))\right)\right]$$

$$\mu_{22}(z;y) = E_{Y_{0t-2}|D_t=1}\left[\left(\mathbb{1}\{z \leq F^{-1}_{\Delta Y_{0t-1}|D_t=1}(F_{\Delta Y_{0t}|D_t=0}(y - F^{-1}_{Y_{0t-1}|D_t=1}(F_{Y_{0t-2}|D_t=1}(Y_{0t-2}))))\}\right.\right.$$
$$\left.\left. - F_{\Delta Y_{0t}|D_t=0}(y - F^{-1}_{Y_{0t-1}|D_t=1}(F_{Y_{0t-2}|D_t=1}(Y_{0t-2})))\right)\right]$$

$$\mu_{32}(z;y) = E_{Y_{0t-2}|D_t=1}\left[\left(\frac{f_{\Delta Y_{0t}|D_t=0}(y - F^{-1}_{Y_{0t-1}|D_t=1}(F_{Y_{0t-2}|D_t=1}(Y_{0t-2})))}{f_{Y_{0t-1}|D_t=1}(F_{Y_{0t-2}|D_t=1}(Y_{0t-2}))}\right.\right.$$

18

$$\times \left(\mathbb{1}\{z \le \mathrm{F}^{-1}_{Y_{0t-1}|D_t=1}(\mathrm{F}_{Y_{0t-2}|D_t=1}(Y_{0t-2}))\} \right.$$
$$\left. - \mathrm{F}_{Y_{0t-1}|D_t=1}(\mathrm{F}^{-1}_{Y_{0t-1}|D_t=1}(\mathrm{F}_{Y_{0t-2}|D_t=1}(Y_{0t-2}))) \right) \right]$$

$$\mu_{42}(z;y) = \mathrm{E}_{Y_{0t-2}|D_t=1}\left[\left(\frac{f_{\Delta Y_{0t}|D_t=0}(y - \mathrm{F}^{-1}_{Y_{0t-1}|D_t=1}(\mathrm{F}_{Y_{0t-2}|D_t=1}(Y_{0t-2})))}{f_{Y_{0t-1}|D_t=1}(\mathrm{F}_{Y_{0t-2}|D_t=1}(Y_{0t-2}))} \right. \right.$$
$$\left. \left. \times \left(\mathbb{1}\{z \le Y_{0t-2}\} - \mathrm{F}_{Y_{0t-2}|D_t=1}(Y_{0t-2})) \right) \right]$$

$$\mu_5(z_1,z_2;y) = \left(\mathbb{1}\{\mathrm{F}^{-1}_{\Delta Y_{0t}|D_t=0}(\mathrm{F}_{\Delta Y_{0t-1}|D_t=1}(z_1)) \le y - \mathrm{F}^{-1}_{Y_{0t-1}|D_t=1}(\mathrm{F}_{Y_{0t-2}|D_t=1}(z_2))\} \right.$$
$$\left. - \mathrm{E}\left[\mathbb{1}\{\mathrm{F}^{-1}_{\Delta Y_{0t}|D_t=0}(\mathrm{F}_{\Delta Y_{0t-1}|D_t=1}(z_1)) \le y - \mathrm{F}^{-1}_{Y_{0t-1}|D_t=1}(\mathrm{F}_{Y_{0t-2}|D_t=1}(z_2))\} \right) |D_t = 1 \right]$$

$$\psi(z;y) = (\mathbb{1}\{z \le y\} - \tau)$$

and

$$\lambda_{30}(y,v) = y - \mathrm{F}^{-1}_{Y_{0t-1}|D_t=1}(\mathrm{F}_{Y_{0t-2}|D_t=1}(v))$$

$$\lambda_{10}(y,v) = \mathrm{F}^{-1}_{\Delta Y_{0t-1}|D_t=1}(\mathrm{F}_{\Delta Y_{0t}|D_t=0}(\lambda_{30}(y,v))), v)$$

and

$$f_{Y_{0t}|D_t=1}(y) = \int_{\mathcal{Y}_{0t-2}|D_t=1} f_{\Delta Y_{0t-1}, Y_{0t-2}|D_t=1}(\lambda_{10}(y,v),v) \frac{f_{\Delta Y_{0t}|D_t=0}(\lambda_{30}(y,v))}{f_{\Delta Y_{0t-1}|D_t=1}(\lambda_{10}(y,v))} \, \mathrm{d}v$$

Theorem 4. *Asymptotic Normality*

Under the Distributional Difference in Differences Assumption, the Copula Stability Assumption, Assumption 2, Assumption 3, and Assumption 4

$$\sqrt{n}(Q\hat{T}ET(\tau) - QTET(\tau)) \xrightarrow{d} N(0,V)$$

where

$$V = \frac{1}{\left\{ f_{Y_{0t}|D_t=1}(F^{-1}_{Y_{0t}|D_t=1}(\tau)) \right\}^2} V_0$$
$$+ \frac{1}{\left\{ f_{Y_{1t}|D_t=1}(F^{-1}_{Y_{1t}|D_t=1}(\tau)) \right\}^2} V_1$$

$$-\frac{2}{f_{Y_{0t}|D_t=1}(F^{-1}_{Y_{0t}|D_t=1}(\tau))\cdot f_{Y_{1t}|D_t=1}(F^{-1}_{Y_{1t}|D_t=1}(\tau))}V_{01}$$

and

$$V_0 = \frac{1-p}{p^2}\,\mathrm{E}\left[\mu_{12}(\Delta Y_t; F^{-1}_{Y_{0t}|D_t=1}(\tau))^2|D_t=0\right]$$

$$+\frac{1}{p}\,\mathrm{E}\left[\Big(\mu_{22}(\Delta Y_{t-1}; F^{-1}_{Y_{0t}|D_t=1}(\tau)) + \mu_{32}(Y_{t-1}; F^{-1}_{Y_{0t}|D_t=1}(\tau))\right.$$

$$\left.+\mu_{42}(Y_{t-2}; F^{-1}_{Y_{0t}|D_t=1}(\tau)) + \mu_{5}(\Delta Y_{t-1}, Y_{t-2}; F^{-1}_{Y_{0t}|D_t=1}(\tau))\Big)^2|D_t=1\right]$$

$$V_1 = \frac{1}{p}\,\mathrm{E}\left[\psi(Y_{1t}; F^{-1}_{Y_{1t}|D_t=1}(\tau))^2|D_t=1\right]$$

$$V_{01} = \frac{1}{p}\,\mathrm{E}\left[\psi(Y_{1t}; F^{-1}_{Y_{1t}|D_t=1}(\tau))\Big(\mu_{22}(\Delta Y_{t-1}; F^{-1}_{Y_{0t}|D_t=1}(\tau)) + \mu_{32}(Y_{t-1}; F^{-1}_{Y_{0t}|D_t=1}(\tau))\right.$$

$$\left.+\mu_{42}(Y_{t-2}; F^{-1}_{Y_{0t}|D_t=1}(\tau)) + \mu_{5}(\Delta Y_{t-1}, Y_{t-2}; F^{-1}_{Y_{0t}|D_t=1}(\tau))\Big)|D_t=1\right]$$

5.1.2 Distributional Difference in Differences Assumption holds conditional on covariates

This section shows consistency and asymptotic normality of our estimator in the case where the Distributional Difference in Differences Assumption holds conditional on covariates. We first show these results in the case where the propensity score is estimated nonparametrically by using series logit methods. We also provide a result for the case where the propensity score is estimated parametrically using, for example, parametric logit or probit specifications. We make the following additional assumptions

Assumption 6. $\mathrm{E}[\mathbb{1}\{\Delta Y_{0t} \le y\}|X, D_t=0]$ *is continuously differentiable for all $x \in \mathcal{X}$.*

Assumption 7. *(Distribution of X)*

(i) The support $\mathcal{X}$ of X is a Cartesian product of compact intvervals; that is, $\mathcal{X} = \prod_{j=1}^{r}[x_{lj}, x_{uj}]$ where r is the dimension of X and x_{lj} and x_{uj} are the smallest and largest values in the support of the j-th dimension of X.

(ii) The density of X, $f_X(\cdot)$, is bounded away from 0 on $\mathcal{X}$.

Assumption 8. *(Assumptions on the propensity score)*

(i) $p(x)$ is continuously differentiable of order $s \ge 7r$ where r is the dimension of X.

(ii) There exist $\underline{p}$ and $\bar{p}$ such that $0 < \underline{p} \le p(x) \le \bar{p} < 1$.

Assumption 9. *(Series Logit Estimator)*

For nonparametric estimation of the propensity score, $p(x$ is estimated by series logit where the power series of the order $K = n^{\nu}$ for some $\frac{1}{4(s/r-1)} < \nu < \frac{1}{9}$.

Note that the restriction on derivatives in Assumption 8 (i) guarantees the existence of ν that satisfies the conditions of Assumption 9.

Theorem 5. *Consistency under Conditional Distributional Difference in Differences Assumption and when the propensity score is estimated nonparametrically*

Under the Conditional Distributional Difference in Differences Assumption, the Copula Stability Assumption, Assumption 2, Assumption 3, Assumption 4, Assumption 6, Assumption 7, and Assumption 8

$$Q\hat{T}ET(\tau) = \hat{F}^{-1}_{Y_{1t}|D_t=1}(\tau) - \hat{F}^{-1}_{Y_{0t}|D_t=1}(\tau) \xrightarrow{p} F^{-1}_{Y_{1t}|D_t=1}(\tau) - F^{-1}_{Y_{0t}|D_t=1}(\tau) = QTET(\tau)$$

We also introduce the following additional notation. In addition to μ_{22}, μ_{32}, μ_{42}, μ_5, and ψ defined above, for $z = (x, \Delta, d)$, let

$$\Psi_{N12}(z; y) = \mathrm{E}_{Y_{0t-2}|D_t=1}\left\{ \frac{p(x)}{(1-p(x))}\mathbb{1}\{\Delta \le y - F^{-1}_{Y_{0t-1}|D_t=1}(F_{Y_{0t-2}|D_t=2}(Y_{0t-2}))\} \right. \tag{10}$$

$$\left. - \mathrm{E}_{X,\Delta Y_t|D_t=0}\left[\frac{p(X)}{(1-p(X))}\mathbb{1}\{\Delta Y_t \le y - F^{-1}_{Y_{0t-1}|D_t=1}(F_{Y_{0t-2}|D_t=2}(Y_{0t-2}))\}\right] \right\} \tag{11}$$

$$\Psi_{N22}(z; y) = \mathrm{E}_{Y_{0t-2}|D_t=1}\left[\frac{\mathrm{E}[\mathbb{1}\{\Delta Y_t \le y - F^{-1}_{Y_{0t-1}|D_t=1}(F_{Y_{0t-2}|D_t=2}(Y_{0t-2}))\}|X=x, D_t=0]}{(1-p(x))} \right] \tag{12}$$

$$\times (d - p(x)) \tag{13}$$

and also replace the definition of $f_{Y_{0t}|D_t=1}(y)$ above with the following

$$f_{Y_{0t}|D_t=1}(y) = \int_{\mathcal{Y}_{0t-2}|D_t=1} f_{\Delta Y_{0t-1}, Y_{0t-2}|D_t=1}(\lambda_{10}(y,v), v)\frac{f_{\Delta Y_{0t}|D_t=1}(\lambda_{30}(y,v))}{f_{\Delta Y_{0t-1}|D_t=1}(\lambda_{10}(y,v))}\ \mathrm{d}v$$

where the difference from the previous definition is that $f_{\Delta Y_{0t}|D_t=0}(\cdot)$ (which is identified directly from the data) is replaced by $f_{\Delta Y_{0t}|D_t=1}(\Delta)$ which is obtained from the propensity score reweighted distribution derived above and is equal to

$$f_{\Delta Y_{0t}|D_t=1}(\Delta) = \mathrm{E}\left[\frac{p(X)}{1-p(X)}f_{\Delta Y_{0t}|D_t=0}(\Delta)|\Delta Y_{0t} = \Delta, D_t = 0 \right]$$

Theorem 6. *Asymptotic Normality under Conditional Distributional Difference in Differences Assumption and when the propensity score is estimated nonparametrically*

Under the Conditional Distributional Difference in Differences Assumption, the Copula Stability Assumption, Assumption 2, Assumption 3, Assumption 4, Assumption 6, Assumption 7, and Assumption 8

$$\sqrt{n}(Q\hat{T}ET(\tau) - QTET(\tau)) \xrightarrow{d} N(0, V_N)$$

where

$$V = \frac{1}{\left\{f_{Y_{0t}|D_t=1}(F^{-1}_{Y_{0t}|D_t=1}(\tau))\right\}^2} V_{0N}$$

$$+ \frac{1}{\left\{f_{Y_{1t}|D_t=1}(F^{-1}_{Y_{1t}|D_t=1}(\tau))\right\}^2} V_{1N}$$

$$- \frac{2}{f_{Y_{0t}|D_t=1}(F^{-1}_{Y_{0t}|D_t=1}(\tau)) \cdot f_{Y_{1t}|D_t=1}(F^{-1}_{Y_{1t}|D_t=1}(\tau))} V_{01N}$$

and

$$V_{0N} = \frac{1}{p^2} \, \mathrm{E}\left[\left(\Psi_{N22}(Z_i; F^{-1}_{Y_{0t}|D_t=1}(\tau)) + (1-D_t)\Psi_{N12}(Z_i; F^{-1}_{Y_{0t}|D_t=1}(\tau)\right.\right.$$

$$+ D_t\mu_{22}(Y_{t-2}; F^{-1}_{Y_{0t}|D_t=1}(\tau)) + D_t\mu_{32}(Y_{t-2}; F^{-1}_{Y_{0t}|D_t=1}(\tau))$$

$$\left.\left. + D_t\mu_{42}(Y_{t-2}; F^{-1}_{Y_{0t}|D_t=1}(\tau)) + D_t\mu_5(\Delta Y_{t-1}, Y_{t-2}; F^{-1}_{Y_{0t}|D_t=1}(\tau))\right)^2\right]$$

$$V_{1N} = \frac{1}{p} \, \mathrm{E}\left[\psi(Y_{1t}; F^{-1}_{Y_{1t}|D_t=1}(\tau))^2 | D_t = 1\right]$$

$$V_{01N} = \frac{1}{p^2} \, \mathrm{E}\left[D_t\psi(Y_{1t}; F^{-1}_{Y_{1t}|D_t=1}(\tau))\right.$$

$$\times \left(\Psi_{N22}(Z_i; F^{-1}_{Y_{0t}|D_t=1}(\tau)) + \mu_{22}(Y_{t-2}; F^{-1}_{Y_{0t}|D_t=1}(\tau)) + \mu_{32}(Y_{t-2}; F^{-1}_{Y_{0t}|D_t=1}(\tau))\right.$$

$$\left.\left. + \mu_{42}(Y_{t-2}; F^{-1}_{Y_{0t}|D_t=1}(\tau)) + \mu_5(\Delta Y_{t-1}, Y_{t-2}; F^{-1}_{Y_{0t}|D_t=1}(\tau))\right)\right]$$

Remark. When the true propensity score is known up to a finite set of parameters so that $p(x) = G(x^\top\gamma_0)$, then consistency and asymptotic normality continue to hold. The proof is identical to the nonparametric case with the following exceptions. First, the propensity score $p(x)$ should be replaced everywhere by $G(x^\top\gamma_0)$. Second, the following assumption replaces Assumption 8 and Assumption 9.

Assumption 10. *(Parametric Propensity Score)*
 (i) γ_0 is an interior point of a compact set $\Gamma \subset \mathbb{R}^d$
 (ii) $E[XX^\top]$ is non-singular
 (iii) Let $\upsilon = \{x^\top\gamma : x \in \mathcal{X}, \gamma \in \Gamma\}$. Then, for $\upsilon \in \Upsilon$, $G(\upsilon)$ is bounded away from 0 and 1, strictly increasing, and continuously differentiable with derivative $g(\upsilon)$ that is bounded away from zero and infinity.

 Third, $\Psi_{P22}(Z_i; \mathrm{F}^{-1}_{Y_{0t}|D_t=1}(\tau))$ replaces $\Psi_{N22}(Z_i; \mathrm{F}^{-1}_{Y_{0t}|D_t=1}(\tau))$ where

$$\Psi_{P22}(z;y) = \mathbb{E}_{Y_{0t-2}|D_t=1}\left\{\mathbb{E}_{X,\Delta Y_t|D_t=0}\left[\frac{1}{1-G(X^\top\gamma_0)}\left(1+\frac{G(X^\top\gamma_0)}{1-G(X^\top\gamma_0)}\right)\right.\right.$$

$$\left.\left.\times\ \mathbb{1}\{\Delta Y_t \le y - F^{-1}_{Y_{0t-1}|D_t=1}(F_{Y_{0t-2}|D_t=1}(Y_{0t-2}))\}g(X^\top\gamma_0)X^\top\right]\right\}$$

$$\times\ \mathbb{E}_{X,D_t}\left[\left(\frac{D_t - G(X^\top\gamma_0)}{G(X^\top\gamma_0)(1-G(X^\top\gamma_0))}\right)^2 g(X^\top\gamma_0)^2 XX^\top\right]^{-1}$$

$$\times\ \frac{d - G(x^\top\gamma_0)}{G(x^\top\gamma_0)(1-G(x^\top\gamma_0))}g(x^\top\zeta_0)x \tag{14}$$

Remark. The key step in the the proof for the case without covariates is to show the counterfactual distribution of untreated potential outcomes can be written in the following way

$$\sqrt{n}(\hat{\mathbb{P}}(Y_{0t} \le y|D_t = 1) - \mathbb{P}(Y_{0t} \le y|D_t = 1)) = \sqrt{n}(\hat\mu_1 + \hat\mu_2 + \hat\mu_3 + \hat\mu_4 + \hat\mu_5) + o_p(1) \tag{15}$$

where

$$\hat\mu_1 = \frac{1}{n_T n_C}\sum_{i\in\mathcal{T}}\sum_{j\in\mathcal{C}}\mathbb{1}\{\Delta Y_{0t,j} \le (y - F^{-1}_{Y_{0t-1}|D_t=1}(F_{Y_{0t-2}|D_t=1}(Y_{0t-2,i})))\}$$

$$- F_{\Delta Y_{0t}|D_t=0}(y - F^{-1}_{Y_{0t-1}|D_t=1}(F_{Y_{0t-2}|D_t=1}(Y_{0t-2,i})))$$

$$\equiv \frac{1}{n_T n_C}\sum_{i\in\mathcal{T}}\sum_{j\in\mathcal{C}}\mu_1(Y_{0t-2,i}, \Delta Y_{0t,j})$$

$$\hat\mu_2 = \frac{1}{n_T^2}\sum_{i\in\mathcal{T}}\sum_{j\in\mathcal{T}}\mathbb{1}\{\Delta Y_{0t-1,j} \le F^{-1}_{\Delta Y_{0t-1}|D_t=1}(F_{\Delta Y_{0t}|D_t=0}(y - F^{-1}_{Y_{0t-1}|D_t=1}(F_{Y_{0t-2}|D_t=1}(Y_{0t-2,i}))))\}$$

$$- F_{\Delta Y_{0t-1}|D_t=1}(F^{-1}_{\Delta Y_{0t-1}|D_t=1}(F_{\Delta Y_{0t}|D_t=0}(y - F^{-1}_{Y_{0t-1}|D_t=1}(F_{Y_{0t-2}|D_t=1}(Y_{0t-2,i})))))]$$

$$\equiv \frac{1}{n_T^2}\sum_{i\in\mathcal{T}}\sum_{j\in\mathcal{T}}\mu_2(Y_{t-2,i}, \Delta Y_{t-1,j})$$

$$\hat\mu_3 = \frac{1}{n_T^2}\sum_{i\in\mathcal{T}}\sum_{j\in\mathcal{T}}\frac{f_{\Delta Y_{0t}|D_t=0}(y - F^{-1}_{Y_{0t-1}|D_t=1}(F_{Y_{0t-2}|D_t=1}(Y_{0t-2,i})))}{f_{Y_{0t-1}|D_t=1}(F_{Y_{0t-2}|D_t=1}(Y_{0t-2,i}))}$$

$$\times\left(\mathbb{1}\{Y_{0t-1,j} \le F^{-1}_{Y_{0t-1}|D_t=1}(F_{Y_{0t-2}|D_t=1}(Y_{0t-2,i}))\}\right.$$

$$\left. - F_{Y_{0t-1}|D_t=1}(F^{-1}_{Y_{0t-1}|D_t=1}(F_{Y_{0t-2}|D_t=1}(Y_{0t-2,i})))\right)$$

$$\equiv \frac{1}{n_T^2}\sum_{i\in\mathcal{T}}\sum_{j\in\mathcal{T}}\mu_3(Y_{t-2,i}, Y_{t-1,j})$$

$$\hat\mu_4 = \frac{1}{n_T^2}\sum_{i\in\mathcal{T}}\sum_{j\in\mathcal{T}}\frac{f_{\Delta Y_{0t}|D_t=0}(y - F^{-1}_{Y_{0t-1}|D_t=1}(F_{Y_{0t-2}|D_t=1}(Y_{0t-2,i})))}{f_{Y_{0t-1}|D_t=1}(F^{-1}_{Y_{0t-1}|D_t=1}(F_{Y_{0t-2}|D_t=1}(Y_{0t-2,i})))}$$

$$\times \left(\mathbb{1}\{Y_{0t-2,j} \leq Y_{0t-2,i}\} - F_{Y_{0t-2}|D_t=1}(Y_{0t-2,i})]\right)$$

$$\equiv \frac{1}{n_T^2} \sum_{i \in \mathcal{T}} \sum_{j \in \mathcal{T}} \mu_4(Y_{t-2,i}, Y_{t-2,j})$$

$$\hat{\mu}_5 = \frac{1}{n_T} \sum_{i \in \mathcal{T}} \left[\mathbb{1}\{F_{\Delta Y_{0t}|D_t=0}^{-1}(F_{\Delta Y_{0t-1}|D_t=1}(\Delta Y_{0t-1,i})) \leq y - F_{Y_{0t-1}|D_t=1}^{-1}(F_{Y_{0t-2}|D_t=1}(Y_{0t-2,i}))\} \right]$$

$$- \mathrm{E}\left[\mathbb{1}\{F_{\Delta Y_{0t}|D_t=0}^{-1}(F_{\Delta Y_{0t-1}|D_t=1}(\Delta Y_{0t-1,i})) \leq y - F_{Y_{0t-1}|D_t=1}^{-1}(F_{Y_{0t-2}|D_t=1}(Y_{0t-2,i}))\}|D_t = 1\right]$$

$$\equiv \frac{1}{n_T} \sum_{i \in \mathcal{T}} \mu_5(\Delta Y_{t-1,i}, Y_{0t-2,i})$$

Then, using standard results on V-statistics, Equation 15 can be written as

$$\sqrt{n}(\hat{\mathrm{P}}(Y_{0t} \leq y|D_t = 1) - \mathrm{P}(Y_{0t} \leq y|D_t = 1))$$

$$= \sqrt{n}\left(\frac{1}{n} \sum_{i=1}^{n} \left\{ \frac{(1-D_t)}{(1-p)} \mu_{12}(\Delta Y_t; F_{Y_{0t}|D_t=1}^{-1}(\tau)) \right.\right.$$

$$+ \frac{D_t}{p}\left[\mu_{22}(\Delta Y_{t-1}; F_{Y_{0t}|D_t=1}^{-1}(\tau)) + \mu_{32}(Y_{t-1}; F_{Y_{0t}|D_t=1}^{-1}(\tau)) \right.$$

$$\left.\left.\left. + \mu_{42}(Y_{t-2}; F_{Y_{0t}|D_t=1}^{-1}(\tau)) + \mu_5(\Delta Y_{t-1}, Y_{t-2}; F_{Y_{0t}|D_t=1}^{-1}(\tau)) \right]\right\}\right) + o_p(1)$$

Then, the result follows by accounting for estimating quantiles instead of distribution functions and the Central Limit Theorem.

In the case with covariates, the result follows from combining the results in the unconditional case with the results on two step propensity score weighting where the prospensity score is estimated by Series Logit as in Hirano, Imbens, and Ridder (2003).

6 Empirical Exercise: Quantile Treatment Effects of a Job Training Program on Subsequent Wages

In this section, we use a well known dataset from LaLonde (1986) that consists of (i) data from randomly assigning job training program applicants to a job training program and (ii) a second dataset consisting of observational data consisting of some individuals who are treated and some who are not treated. This dataset has been widely used in the program evaluation literature. Having access to both a randomized control and an observational control group is a powerful tool for evaluating the performance of observational methods in estimating the effect of treatment. The original contribution of LaLonde (1986) is that many typically used methods (least squares regression, Difference in Differences, and the Heckman selection model) did not perform very well in estimating the average effect of participation in the job training program. An important subsequent literature argued that observational methods can effectively estimate the effect of a job training

program, but the results are sensitive to the implementation (Heckman and Hotz 1989; Heckman, Ichimura, and Todd 1997; Heckman, Ichimura, Smith, and Todd 1998; Dehejia and Wahba 1999; Smith and Todd 2005). Finally, Firpo (2007) has used this dataset to study the quantile treatment effects of participating in the job training program under the selection on observables assumption.

One limitation of the dataset for estimating quantile treatment effects is that the 185 treated observations form only a moderately sized dataset. A second well known issue is that properly evaluating the training program, even with appropriate methods, may not be possible using the Lalonde dataset because control observations do not come from the same local labor markets and surveys for the control group do not use the same questionnaire (Heckman, Ichimura, and Todd 1997) though some of these issues may be alleviated using Difference in Differences methods.

In the rest of this section, we implement the procedure outlined in this paper, and compare the resulting QTET estimates to those from the randomized experiment and the various other procedures available to estimate quantile treatment effects.

6.1   Data

The job training data is from the National Supported Work (NSW) Demonstration. The program consisted of providing extensive training to individuals who were unemployed (or working very few hours) immediately prior to participating in the program. Detailed descriptions of the program are available in Hollister, Kemper, and Maynard (1984), LaLonde (1986), and Smith and Todd (2005). Our analysis focuses on the all-male subset used in Dehejia and Wahba (1999). This subset has been the most frequently studied. In particular, Firpo (2007) uses this subset. Importantly for applying the method presented in this paper, this subset contains data on participant earnings in 1974, 1975, and 1978.[9]

The experimental portion of the dataset contains 445 observations. Of these, 185 individuals are randomly assigned to participate in the job training program. The observational control group comes from the Panel Study of Income Dynamics (PSID). There are 2490 observations in the PSID sample. Estimates using the observational data combine the 185 treated observations for the job training program with the 2490 untreated observations from the PSID sample. The PSID sample is a random sample from the U.S. population that is likely to be dissimilar to the treated group in many observed and unobserved ways. For this reason, conditioning on observed factors that affect whether or not an individual participates in the job training program *and* using a method that adjusts for unobserved differences between the treated and control groups are likely to be important steps to take to correctly understand the effects of the job training program.

Summary statistics for earnings by treatment status (treated, randomized controls, observational controls) are presented in Table 1. Average earnings are very similar between the treated group and the randomized control group in the two years prior to treatment. After treatment, average earnings are about $1700 higher for the treated group than the control group indicating

---

[9]Dehejia and Wahba (1999) showed that conditioning on two periods of lagged earning was important for correctly estimating the average treatment effect on the treated using propensity score matching techniques.

that treatment has, on average, a positive effect on earnings. Average earnings for the observational control group are well above the earnings of the treated group in all periods (including the after treatment period).

For the available covariates, no large differences exist between the treated group and the randomized control group. The largest normalized difference is for high school degree status. The treated group is about 13% more likely to have a high school degree. There are large differences between the treated group and the observational control group. The observational control group is much less likely to have been unemployed in either of the past two years. They are older, more educated, more likely to be married, and less likely to be a minority. These large differences between the two groups are likely to explain much of the large differences in earnings outcomes.

6.2 Results

The PanelQTET identification results require the underlying distributions to be continuous. However, because participants in the job training program were very likely to have no earnings during the period of study due to high rates of unemployment, we estimated the effect of job training only for $\tau = (0.7, 0.8, 0.9)$. This strategy is similar to Buchinsky (1994, Footnote 22) though we must focus on higher quantiles than in that paper. We plan future work on developing identification or partial identification strategies when the outcomes have a mixed continuous and discrete distribution.

Main Results Table 3 provides estimates of the 0.7-, 0.8-, and 0.9-QTET using the method of this paper (which we hereafter term PanelQTET), the conditional independence (CI) method (Firpo 2007), the Change in Changes method (Athey and Imbens 2006), the Quantile Difference in Differences (QDiD) method, and the Mean Difference in Differences (MDiD) method. It also compares the resulting estimates using each of these methods with the experimental results.

For each type of estimation, results are presented using three sets of covariates: (i) the first row includes age, education, black dummy variable, hispanic dummy variable, married dummy variable, and no high school degree dummy variable (call this COV below) – this represents the set of covariates that are likely to be available with cross sectional data; (ii) the second row includes the same covariates plus two dummy variables indicating whether or not the individual was unemployed in 1974 or 1975 (call this UNEM below) – this represents the set of covariates that may be available with panel data or when the dataset contains some retrospective questions; and (iii) the third row includes no covariates (call this NO COV below) – including this set of covariates allows us to judge the relative importance of adjust for both observable differences across individuals and time invariant unobserved differences across individuals.

The PanelQTET method and the CI method admit estimation based on a first step estimate of the propensity score. For both of these methods, we estimate parametric versions of the propensity score using the three specifications mentioned above. Additionally, we also include an additional set of results based on nonparametric estimate of the propensity score using a series logit method.

In practice, the PanelQTET method and the CI method use slightly different series logit estimates. For the PanelQTET method, we select the number of approximating terms using a cross-validation method. We use only covariates available from the UNEM covariate set as it would not be appropriate to condition on lags of the dependent variable. We do condition on lags of unemployment. For the CI method, we use the series logit specification used in Firpo (2007). The key difference between the two is that the CI method can condition on lags of the dependent variable real earnings in addition to all the other available covariates.

For CiC, QDiD, and MDiD, propensity score re-weighting techniques are not available. One could potentially attempt to nonparametrically implement these estimators, but the resulting estimators are likely to be quite computationally challenging. Instead, we follow the idea of Athey and Imbens (2006) and residualize the earnings outcome by regressing earnings on a dummy variable indicating whether or not the observations belongs to one of the four groups: (treated, 1978), (untreated, 1978), (treated, 1975), (untreated, 1975) and the available covariates. The residuals remove the effect of the covariates but not the group (See Athey and Imbens (2006) for more details). Then, we perform each method on the residualized outcome. We discuss the estimation results for each method in turn.

The first section of Table 2 reports estimates of the QTET using the PanelQTET method. The first row provides results where the propensity score is estimated nonparametrically using series logit. The estimated QTET is positive and statistically significant at each of the 0.7, 0.8, and 0.9-quantiles though the estimates tend to be larger than the experimental results These estimates are statistically different from the experimental results at the 0.8 and 0.9-quantiles. These results also indicate that the QTET is increasing at larger quantiles which is in line with the experimental results. The second row provides results using the COV conditioning set. In our view, this specification is likely to be what an empirical researcher would estimate given the available data and if he were to use the PanelQTET method. Out of all 16 method-covariate set estimates presented in Table 2, the QTETs come closest to matching the experimental results using the PanelQTET method and the COV conditioning set. The point estimate for each of the 0.7, 0.8, and 0.9-quantiles are somewhat smaller than the ATT indicating that the gain from the job training program was either similar across quantiles or slightly at lower income parts of the distribution than at higher income parts of the distribution. The experimental dataset gives precisely the opposite conclusion though: gains at the higher income part of the distribution were somewhat larger than average gains. The difference in conclusions results mainly from a large difference in the estimated ATT[10] and the experimental ATT. When using the UNEM conditioning set, the estimates of the QTET are very similar to the nonparametric specification. Finally, the NO COV conditioning set tends to perform the most poorly. The QTET is estimated to be close to zero at each quantile and is statistically different from the experimental results for the 0.7 and 0.9-quantile.

The second section presents results using cross sectional data. The results in the first row come

---

[10]The ATT is estimated under the same assumptions as the QTET. In this case, however, the same assumptions imply that the propensity score re-weighting technique of Abadie (2005) should be used.

from estimating the propensity score nonparametrically using series logit where the conditioning set can include lags of the dependent variable real earnings. If we had imposed linearity (and momentarily ignoring the nonparametric estimation of the propensity score), the difference between the CI and the PanelQTET model is that the CI model would include lags of the dependent variable but no fixed effect while the PanelQTET model would include a fixed effect but no lags of the dependent variable. Just as in the case of the linear model, the choice of which model to use depends on the application and the decision of the researcher. Not surprisingly then, the results that include dynamics under the CI assumption are much better than those that do not include dynamics. The results are, in fact, quite similar to the results using the PanelQTET method with the propensity score estimated nonparametrically; particularly, the estimated effect have the right sign but tend to be overestimated. The results in the second row come from conditioning on the COV conditioning set. The COV conditioning set contain only the values of the covariates that would be available in a strictly cross sectional dataset. These results are very poor. The QTET and ATT are estimated to be large and negative indicating that participating in the job training program tended to strongly decrease wages. In fact, the CI procedure using purely cross sectional data performs much worse than any of the other methods that take into account having multiple periods of data (notably, this includes specifications that include no covariates at all). The third specifications uses the UNEM conditioning set, and the performance is similar to the nonparametric estimation of the propensity score. Finally, the fourth row considers estimates that invoke CI without the need to condition on covariates. This assumption is highly unlikely to be true as individuals in the treated group differ in many observed ways from untreated individuals. This method would attribute higher earnings among untreated individuals to not being in the job training program despite the fact that they tended to have much larger earnings before anyone entered job training as well as more education and more experience.

The final three sections of Table 3 provide results using CiC, QDiD, and MDiD. We briefly summarize these results. Broadly speaking, each of these three methods, regardless of conditioning set, performs better than invoking the CI assumption using covariates that are available only in the same period as the outcome (CI-COV results). Between the three methods, the QDiD method performs slightly better than the CiC and MDiD model. Comparing the results of these three models to the results from the PanelQTET method, the PanelQTET method performs slightly better than the CiC and MDiD model. With the COV specification, it performs evenly with the QDiD method. With the UNEM specification, it performs slightly worse than the QDiD method.

7 Conclusion

Previous work on estimating quantile treatment effects in Difference in Differences models has relied on either (i) strong distributional assumptions or (ii) partial identification. This paper has shown how to combine the most straightforward Distributional Difference in Differences assumption with an additional assumption on the dependence between the change in untreated potential

outcomes and the initial level of the untreated potential outcome and access to three periods of panel data (rather than the more typical two periods of repeated cross sections) to estimate the Quantile Treatment Effect on the Treated. This procedure may be useful to empirical researchers who want to invoke what may be a more plausible set of assumptions to estimate quantile treatment effects in some applications.

The second contribution of the paper is to show how to easily accomodate covariates that may be needed to justify the Distributional Difference in Differences assumption using a propensity score re-weighting technique. This is an important contribution because in many applications, a Difference in Differences assumption may only be valid after conditioning on some covariates. Other existing methods for estimating the Quantile Treatment Effect on the Treated under an assumption holding conditional on covariates are either unavailable or require strong parametric functional form assumptions to implement.

References

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

Abadie, Alberto, Joshua Angrist, and Guido Imbens (2002). "Instrumental variables estimates of the effect of subsidized training on the quantiles of trainee earnings". *Econometrica* 70.1, pp. 91–117.

Abbring, Jaap and James Heckman (2007). "Econometric evaluation of social programs, part III: Distributional treatment effects, dynamic treatment effects, dynamic discrete choice, and general equilibrium policy evaluation". *Handbook of econometrics* 6, pp. 5145–5303.

Ashenfelter, Orley (1978). "Estimating the effect of training programs on earnings". *The Review of Economics and Statistics*, pp. 47–57.

Athey, Susan and Guido Imbens (2006). "Identification and Inference in Nonlinear Difference-in-Differences Models". *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". *Review of Economics and Statistics* 93.2, pp. 479–494.

Buchinsky, Moshe (1994). "Changes in the US wage structure 1963-1987: Application of quantile regression". *Econometrica: Journal of the Econometric Society*, pp. 405–458.

Carneiro, Pedro, Karsten Hansen, and James Heckman (2001). "Removing the Veil of Ignorance in assessing the distributional impacts of social policies". *Swedish Economic Policy Review* 8, pp. 273–301.

— (2003). "Estimating distributions of treatment effects with an application to the returns to schooling and measurement of the effects of uncertainty on college choice". *International Economic Review* 44.2, pp. 361–422.

Chernozhukov, Victor and Christian Hansen (2005). "An IV model of quantile treatment effects". *Econometrica* 73.1, pp. 245–261.

Dehejia, Rajeev and Sadek Wahba (1999). "Causal effects in nonexperimental studies: Reevaluating the evaluation of training programs". *Journal of the American Statistical Association* 94.448, pp. 1053–1062.

Fan, Yanqin and Sang Soo Park (2009). "Partial identification of the distribution of treatment effects and its confidence sets". *Advances in Econometrics* 25, pp. 3–70.

Fan, Yanqin and Zhengfei Yu (2012). "Partial identification of distributional and quantile treatment effects in difference-in-differences models". *Economics Letters* 115.3, pp. 511–515.

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

Froelich, Markus and Blaise Melly (2013). "Unconditional quantile treatment effects under endogeneity". *Journal of Business & Economic Statistics* 31.3, pp. 346–357.

Heckman, James and V Joseph Hotz (1989). "Choosing among alternative nonexperimental methods for estimating the impact of social programs: The case of manpower training". *Journal of the American statistical Association* 84.408, pp. 862–874.

Heckman, James, Hidehiko Ichimura, Jeffrey Smith, and Petra Todd (1998). "Characterizing Selection Bias Using Experimental Data". *Econometrica* 66.5, pp. 1017–1098.

Heckman, James, Hidehiko Ichimura, and Petra Todd (1997). "Matching as an econometric evaluation estimator: Evidence from evaluating a job training programme". *The review of economic studies* 64.4, pp. 605–654.

Heckman, James, Robert LaLonde, and Jeffrey Smith (1999). "The economics and econometrics of active labor market programs". *Handbook of labor economics* 3, pp. 1865–2097.

Heckman, James and Richard Robb (1985). "Alternative methods for evaluating the impact of interventions". *Longitudinal Analysis of Labor Market Data.* Ed. by James Heckman and Burton Singer. Cambridge: Cambridge University Press, pp. 156–246.

Heckman, James and Jeffrey Smith (1999). "The Pre-programme Earnings Dip and the Determinants of Participation in a Social Programme. Implications for Simple Programme Evaluation Strategies". *The Economic Journal* 109.457, pp. 313–348.

Heckman, James, Jeffrey Smith, and Nancy Clements (1997). "Making the most out of programme evaluations and social experiments: Accounting for heterogeneity in programme impacts". *The Review of Economic Studies* 64.4, pp. 487–535.

Heckman, James and Edward Vytlacil (2005). "Structural equations, treatment effects, and econometric policy evaluation1". *Econometrica* 73.3, pp. 669–738.

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

Hollister, Robinson, Peter Kemper, and Rebecca Maynard (1984). *The national supported work demonstration.* Univ of Wisconsin Pr.

Ichimura, Hidehiko (1993). "Semiparametric least squares (SLS) and weighted SLS estimation of single-index models". *Journal of Econometrics* 58.1, pp. 71–120.

Imbens, Guido and Joshua Angrist (1994). "Identification and estimation of local average treatment effects". *Econometrica: Journal of the Econometric Society*, pp. 467–475.

Imbens, Guido and Jeffrey Wooldridge (2009). "Recent Developments in the Econometrics of Program Evaluation". *Journal of Economic Literature* 47.1, pp. 5–86.

Joe, Harry (1997). *Multivariate models and multivariate dependence concepts.* CRC Press.

— (2015). *Dependence modeling with copulas. No. 134 in monographs on statistics and applied probability.* CRC Press, Boca Raton, FL.

Klein, Roger and Richard Spady (1993). "An efficient semiparametric estimator for binary response models". *Econometrica*, pp. 387–421.

Koenker, Roger (2005). *Quantile regression.* 38. Cambridge university press.

LaLonde, Robert (1986). "Evaluating the econometric evaluations of training programs with experimental data". *The American Economic Review*, pp. 604–620.

Lee, Alan (1990). *U-statistics. Theory and Practice.* Marcel Dekker, Inc., New York.

Meyer, Bruce, Kip Viscusi, and David Durbin (1995). "Workers' compensation and injury duration: evidence from a natural experiment". *The American economic review*, pp. 322–340.

Nelsen, Roger (2007). *An introduction to copulas.* Springer.

Newey, Whitney and Daniel McFadden (1994). "Large sample estimation and hypothesis testing". *Handbook of econometrics* 4, pp. 2111–2245.

Sen, Amartya (1997). "On economic inequality". *On economic inequality.* Clarendon Press.

Sklar, Abe (1959). *Fonctions de répartition à n dimensions et leurs marges.* Publications de L Institut de Statistique de L Universite de Paris.

Smith, Jeffrey and Petra Todd (2005). "Does matching overcome LaLonde's critique of nonexperimental estimators?" *Journal of econometrics* 125.1, pp. 305–353.

Thuysbaert, Bram (2007). "Distributional comparisons in difference *in* differences models".

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

A   Proofs

A.1   Identification

A.1.1   Identification without covariates

In this section, we prove Theorem 1. Namely, we show that the counterfactual distribution of untreated outcome $F_{Y_{0t}|D_t=1}(y)$ is identified. First, we state two well known results without proof used below that come directly from Sklar's Theorem.

Lemma 1. *The joint density in terms of the copula pdf*

$$f(x,y) = c(F_X(x), F_Y(y))f_X(x)f_Y(y)$$

Lemma 2. *The copula pdf in terms of the joint density*

$$c(u,v) = f(F_X^{-1}(u), F_Y^{-1}(u))\frac{1}{f_X(F_X^{-1}(u))}\frac{1}{f_Y(F_Y^{-1}(u))}$$

Proof of Theorem 1. To minimize notation, let $\varphi_t(\cdot,\cdot) = \varphi_{\Delta Y_{0t}, Y_{0t-1}|D_t=1}(\cdot,\cdot)$ be the joint pdf of the change in untreated potential outcome and the initial untreated potential outcome for the treated group, and let $\varphi_{t-1}(\cdot,\cdot) = \varphi_{\Delta Y_{0t-1}, Y_{0t-2}|D_t=1}(\cdot,\cdot)$ be the joint pdf in the previous period. Similarly, let $c_t(\cdot,\cdot) = c_{\Delta Y_{0t}, Y_{0t-1}|D_t=0}(\cdot,\cdot)$ and $c_{t-1}(\cdot,\cdot) = c_{\Delta Y_{0t-1}, Y_{0t-2}}(\cdot,\cdot)$ be the copula pdfs for the change in untreated potential outcomes and initial level of untreated outcomes for the treated group at period t and $t-1$, respectively. Then,

$$P(Y_{0t} \le y|D_t=1) = P(\Delta Y_{0t} + Y_{0t-1} \le y|D_t=1)$$

$$= \mathrm{E}[\mathbb{1}\{\Delta Y_{0t} \le y - Y_{0t-1}\}|D_t=1)$$

$$= \int_{\mathcal{Y}_{t-1|D_t=1}} \int_{\Delta \mathcal{Y}_{t|D_t=1}} \mathbb{1}\{\Delta y_{0t} \le y - y_{0t-1}\}\varphi_t(\Delta y_{0t}, y_{0t-1}|D_t=1)\mathrm{d}\Delta y_{0t}\mathrm{d}y_{0t-1}$$

$$
\begin{aligned}
= &\int_{\mathcal{Y}_{t-1|D_t=1}} \int_{\Delta \mathcal{Y}_{t|D_t=1}} \mathbb{1}\{\Delta y_{0t} \le y - y_{0t-1}\} \\
&\times \quad c_t(F_{\Delta Y_{0t}|D_t=1}(\Delta y_{0t}), F_{Y_{0t-1}|D_t=1}(y_{0t-1})) \\
&\times \quad f_{\Delta Y_{0t}|D_t=1}(\Delta y_{0t})f_{Y_{0t-1}|D_t=1}(y_{0t-1})\mathrm{d}\Delta y_{0t}\mathrm{d}y_{0t-1}
\end{aligned}
\tag{16}
$$

$$
\begin{aligned}
= &\int_{\mathcal{Y}_{t-1|D_t=1}} \int_{\Delta \mathcal{Y}_{t|D_t=1}} \mathbb{1}\{\Delta y_{0t} \le y - y_{0t-1}\} \\
&\times \quad c_{t-1}(F_{\Delta Y_{0t}|D_t=1}(\Delta y_{0t}), F_{Y_{0t-1}|D_t=1}(y_{0t-1})) \\
&\times \quad f_{\Delta Y_{0t}|D_t=1}(\Delta y_{0t})f_{Y_{0t-1}|D_t=1}(y_{0t-1})\mathrm{d}\Delta y_{0t}\mathrm{d}y_{0t-1}
\end{aligned}
\tag{17}
$$

$$= \int_{\mathcal{Y}_{t-1|D_t=1}} \int_{\Delta \mathcal{Y}_{t|D_t=1}} \mathbb{1}\{\Delta y_{0t} \leq y - y_{0t-1}\} \tag{18}$$

$$\times \quad \varphi_{t-1}\left\{F^{-1}_{\Delta Y_{0t-1}|D_t=1}(F_{\Delta Y_{0t}|D_t=1}(\Delta y_{0t})), F^{-1}_{Y_{0t-2}|D_t=1}(F_{Y_{0t-1}|D_t=1}(y_{0t-1}))\right\}$$

$$\times \quad \frac{f_{\Delta Y_{0t}|D_t=1}(\Delta y_{0t})}{f_{\Delta Y_{0t-1}|D_t=1}(F^{-1}_{\Delta Y_{0t-1}|D_t=1}(F_{\Delta Y_{0t}|D_t=1}(\Delta y_{0t})))}$$

$$\times \quad \frac{f_{Y_{0t-1}|D_t=1}(y_{0t-1})}{f_{Y_{0t-2}|D_t=1}(F^{-1}_{Y_{0t-2}|D_t=1}(F_{Y_{0t-1}|D_t=1}(y_{0t-1})))} \mathrm{d}\Delta y_{0t}\mathrm{d}y_{0t-1}$$

Equation 16 rewrites the joint distribution in terms of the copula pdf using Lemma 1; Equation 17 uses the copula stability assumption; Equation 18 rewrites the copula pdf as the joint distribution (now in period $t - 1$) using Lemma 2.

Now, make a change of variables: $u = F^{-1}_{\Delta Y_{0t-1}|D_t=1}(F_{\Delta Y_{0t}|D_t=1}(\Delta y_{0t}))$ and $v = F^{-1}_{Y_{0t-2}|D_t=1}(F_{Y_{0t-1}|D_t=1}(y_{0t-1}))$. This implies the following:

1. $\Delta y_{0t} = F^{-1}_{\Delta Y_{0t}|D_t=1}(F_{\Delta Y_{0t-1}|D_t=1}(u))$

2. $y_{0t-1} = F^{-1}_{Y_{0t-1}|D_t=1}(F_{Y_{0t-2}|D_t=1}(v))$

3. $\mathrm{d}\Delta y_{0t} = \dfrac{f_{\Delta Y_{0t-1}|D_t=1}(u)}{f_{\Delta Y_{0t}|D_t=1}(F^{-1}_{\Delta Y_{0t}|D_t=1}(F_{\Delta Y_{0t-1}|D_t=1}(u)))}\mathrm{d}u$

4. $\mathrm{d}y_{0t-1} = \dfrac{f_{Y_{0t-2}|D_t=1}(v)}{f_{Y_{0t-1}|D_t=1}(F^{-1}_{Y_{0t-1}|D_t=1}(F_{Y_{0t-2}|D_t=1}(v)))}\mathrm{d}v$

Plugging in (1)-(4) in Equation 18 and noticing that the substitutions for $\mathrm{d}\Delta y_{0t}$ and $\mathrm{d}y_{0t-1}$ cancel out the fractional terms in the third and fourth lines of Equation 18 implies

$$\text{Equation 18} = \int_{\mathcal{Y}_{t-2|D_t=1}} \int_{\Delta \mathcal{Y}_{t-1|D_t=1}} \mathbb{1}\{F^{-1}_{\Delta Y_{0t}|D_t=1}(F_{\Delta Y_{0t-1}|D_t=1}(u)) \leq y - F^{-1}_{Y_{0t-1}|D_t=1}(F_{Y_{0t-2}|D_t=1}(v))\}$$

$$\tag{19}$$

$$\times \quad \varphi_{t-1}\left(u, v\right)\mathrm{d}u\mathrm{d}v$$

$$= \mathrm{E}\left[\mathbb{1}\{F^{-1}_{\Delta Y_{0t}|D_t=1}(F_{\Delta Y_{0t-1}|D_t=1}(\Delta Y_{0t-1})) \leq y - F^{-1}_{Y_{0t-1}|D_t=1}(F_{Y_{0t-2}|D_t=1}(Y_{0t-2}))\}|D_t=1\right]$$

$$\tag{20}$$

$$= \mathrm{E}\left[\mathbb{1}\{F^{-1}_{\Delta Y_{0t}|D_t=0}(F_{\Delta Y_{0t-1}|D_t=1}(\Delta Y_{0t-1})) \leq y - F^{-1}_{Y_{0t-1}|D_t=1}(F_{Y_{0t-2}|D_t=1}(Y_{0t-2}))\}|D_t=1\right]$$

$$\tag{21}$$

where Equation 19 follows from the discussion above, Equation 20 follows by the definition of expectation, and Equation 21 follows from the Distributional Difference in Differences Assumption.

$\square$

A.1.2 Identification with covariates

In this section, we prove Theorem 2.

Proof. All of the results from the proof of Theorem 1 are still valid. Therefore, all that needs to be shown is that Equation 8 holds. Notice,

$$P(\Delta Y_{0t} \leq \Delta y | D_t = 1) = \frac{P(\Delta Y_{0t} \leq \Delta y, D_t = 1)}{p}$$

$$= E\left[\frac{P(\Delta Y_{0t} \leq \Delta y, D_t = 1 | X)}{p}\right]$$

$$= E\left[\frac{p(X)}{p} P(\Delta Y_{0t} \leq \Delta y | X, D_t = 1)\right]$$

$$= E\left[\frac{p(X)}{p} P(\Delta Y_{0t} \leq \Delta y | X, D_t = 0)\right] \tag{22}$$

$$= E\left[\frac{p(X)}{p} E[(1 - D_t) \mathbb{1}\{\Delta Y_t \leq \Delta y)\} | X, D_t = 0]\right] \tag{23}$$

$$= E\left[\frac{p(X)}{p(1 - p(X))} E[(1 - D_t) \mathbb{1}\{\Delta Y_t \leq \Delta y)\} | X]\right]$$

$$= E\left[\frac{1 - D_t}{1 - p(X)} \frac{p(X)}{p} \mathbb{1}\{\Delta Y_t \leq \Delta y\}\right] \tag{24}$$

where Equation 22 holds by Conditional Distributional Difference in Differences Assumption. Equation 23 holds by replacing $P(\cdot)$ with $E(\mathbb{1}\{\cdot\})$ and then multiplying by $(1 - D_t)$ which is permitted because the expectation conditions on $D_t = 0$. Additionally, conditioning on $D_t = 0$ allows us to replace the potential outcome ΔY_{0t} with the actual outcome ΔY_t because ΔY_t is the observed change in potential untreated outcomes for the untreated group. Finally, Equation 24 simply applies the Law of Iterated Expectations to conclude the proof. □

A.2 Proof of Proposition 1

Proof. We are interested in showing that the Copula Stability Assumption holds in the case of the model in Equation 7 in Proposition 1. First, recall the definition of the copula for the change in untreated potential outcomes for the treated group and the initial level of untreated potential outcomes for the treated group.

$$C_{\Delta Y_{0t}, Y_{0t-1}}(v, w) = P\left(F_{\Delta Y_{0t}}(\Delta Y_{0t}) \leq v, F_{Y_{0t-1}}(Y_{0t-1}) \leq w\right) \tag{25}$$

The model that we consider is the following

$$Y_{0it} = h(X_{it}, U_{it}) + \tau_t$$
$$Y_{0it-1} = h(X_{it-1}, U_{it}) + \tau_{t-1}$$

where for all time periods s, $U_{is} = \varsigma_i + \epsilon_{is}$ and $\epsilon_{is} \perp\!\!\!\perp (X_i, \varsigma_i)$ where X_i is the vector of all covariates for individual i over time. We consider several additional properties for ϵ_{is} below. We have also assumed that $h(\cdot, u)$ is strictly increasing in u. This implies

$$\begin{aligned} F_{Y_{0t-1}}(y) &= P(Y_{0t-1} \leq y) \\ &= P(h(X_{t-1}, \varsigma + \epsilon_{0t}) + \tau_{t-1} \leq y) \\ &= E_{X_{t-1}, \varsigma}[F_\epsilon(h^{-1}(y - \tau_{t-1}; X_{t-1}) - \varsigma)] \end{aligned}$$

This also implies

$$F_{Y_{0t-1}}(\tilde{Y}_{0t-1}) = E_{X_{t-1}, \varsigma}[F_\epsilon(h^{-1}(h(\tilde{X}_{t-1}, \tilde\varsigma + \tilde\epsilon_{0t-1}); X_{t-1}) - \varsigma)] \tag{26}$$

Similarly,

$$\begin{aligned} F_{\Delta Y_{0t}}(\Delta) &= P(\Delta Y_{0t} \leq \Delta) \\ &= P(h(X_t, \varsigma + \epsilon_{0t}) - h(X_{t-1}, \varsigma + \epsilon_{0t} + \tau_t - \tau_{t-1} \leq \Delta) \\ &= E_{X_t, X_{t-1}, \varsigma, \epsilon_{0t}, \epsilon_{0t-1}}[\mathbb{1}\{h(X_t, \varsigma + \epsilon_{0t}) - h(X_{t-1}, \varsigma + \epsilon_{0t-1}) + \tau_t - \tau_{t-1} \leq \Delta\}] \end{aligned}$$

This implies

$$F_{\Delta Y_{0t}}(\Delta \tilde{Y}_{0t}) = E_{X_t, X_{t-1}, \varsigma, \epsilon_{0t}, \epsilon_{0t-1}}[\mathbb{1}\{h(X_t, \varsigma + \epsilon_{0t}) - h(X_{t-1}, \varsigma + \epsilon_{0t-1}) \leq h(\tilde{X}_t, \tilde\varsigma + \tilde\epsilon_{0t}) - h(\tilde{X}_{t-1}, \tilde\varsigma + \tilde\epsilon_{0t-1})\}] \tag{27}$$

Since neither Equation 26 nor Equation 27 change over time, the copula stability assumption holds.

$\square$

A.3 Consistency

Before proving consistency, we state several well known results that are used in the proof.

Lemma 3. *Pointwise convergence of empirical distribution function*

$$\hat{F}_X(x) \xrightarrow{p} F_X(x)$$

Lemma 4. *Pointwise convergence of empirical quantiles*

$$\hat{F}_X^{-1}(x) \xrightarrow{p} F_X^{-1}(x)$$

Lemma 5. *Uniform convergence of empirical distribution function*

$$\sup_{x \in \mathcal{X}} |\hat{F}_X(x) - F_X(x)| \xrightarrow{p} 0$$

Lemma 6. *Uniform convergence of empirical quantiles*

$$\sup_{\tau \in (0,1)} |\hat{F}_X^{-1}(\tau) - F_X^{-1}(\tau)| \xrightarrow{p} 0$$

Proof of Theorem 1. First, $\hat{F}_{Y_{1t}|D_t=1}^{-1}(\tau) \xrightarrow{p} F_{Y_{1t}|D_t=1}^{-1}(\tau)$ which follows immediately from Lemma 4.

Second, we show that $\hat{F}_{Y_{0t}|D_t=1}^{-1}(\tau) \xrightarrow{p} F_{Y_{0t}|D_t=1}^{-1}(\tau)$ separately for the cases when there are covariates and no covariates.

Case 1: No covariates   Recall that the estimator of $F_{Y_{0t}|D_t=1}(y)$ is

$$\frac{1}{n_T} \sum_{i \in \mathcal{T}} \left[\mathbb{1}\{\hat{F}_{\Delta Y_{0t}|D_t=0}^{-1}(\hat{F}_{\Delta Y_{0t-1}|D_t=1}(\Delta Y_{0t-1,i})) \le y - \hat{F}_{Y_{0t-1}|D_t=1}^{-1}(\hat{F}_{Y_{0t-2}|D_t=1}(Y_{0t-2,i}))\} \right] \qquad (28)$$

Based on Lemma 5 and Lemma 6, each of the distributions and quantiles in Equation 28 converges uniformly to their population counterpart. This implies that

$$\frac{1}{n_T} \sum_{i \in \mathcal{T}} \left[\mathbb{1}\{\hat{F}_{\Delta Y_{0t}|D_t=0}^{-1}(\hat{F}_{\Delta Y_{0t-1}|D_t=1}(\Delta Y_{0t-1,i})) \le y - \hat{F}_{Y_{0t-1}|D_t=1}^{-1}(\hat{F}_{Y_{0t-2}|D_t=1}(Y_{0t-2,i}))\} \right]$$

converges uniformly to

$$\frac{1}{n_T} \sum_{i \in \mathcal{T}} \left[\mathbb{1}\{F_{\Delta Y_{0t}|D_t=0}^{-1}(F_{\Delta Y_{0t-1}|D_t=1}(\Delta Y_{0t-1,i})) \le y - F_{Y_{0t-1}|D_t=1}^{-1}(F_{Y_{0t-2}|D_t=1}(Y_{0t-2,i}))\} \right]$$

which by the law of large numbers converges in probability to

$$\mathrm{E}\left[\mathbb{1}\{F_{\Delta Y_{0t}|D_t=0}^{-1}(F_{\Delta Y_{0t-1}|D_t=1}(\Delta Y_{0t-1,i})) \le y - F_{Y_{0t-1}|D_t=1}^{-1}(F_{Y_{0t-2}|D_t=1}(Y_{0t-2,i}))\} \right]$$

The result then follows by the convergence of $\hat{F}_{Y_{1t}|D_t=1}^{-1}(\tau)$ and Slutsky's Lemma.

Case 2: Covariates   The result will hold as long as

$$\frac{1}{n} \sum_{i=1}^{n} \left[\frac{(1 - D_{it})}{p} \frac{\hat{p}(X_i)}{1 - \hat{p}(X_i)} \mathbb{1}\{\Delta Y_{t,i} \le \Delta y\} \right]$$

converges uniformly to

$$\mathrm{E}\left[\frac{(1 - D_t)}{p} \frac{p(X)}{1 - p(X)} \mathbb{1}\{\Delta Y_t \leq \Delta y\}\right]$$

Under Assumption 8 and Assumption 9 when the propensity score is estimated nonparametrically and under Assumption 10 when the propensity score is estimated parametrically, $\hat{p}(x)$ converges uniformly to $p(x)$ which implies that

$$\frac{1}{n}\sum_{i=1}^{n}\left[\frac{(1 - D_{it})}{p} \frac{\hat{p}(X_i)}{1 - \hat{p}(X_i)} \mathbb{1}\{\Delta Y_{t,i} \leq \Delta y\}\right]$$

converges uniformly to

$$\frac{1}{n}\sum_{i=1}^{n}\left[\frac{(1 - D_{it})}{p} \frac{p(X_i)}{1 - p(X_i)} \mathbb{1}\{\Delta Y_{t,i} \leq \Delta y\}\right]$$

Then,

$$\frac{1}{n}\sum_{i=1}^{n}\left[\frac{(1 - D_{it})}{p} \frac{p(X_i)}{1 - p(X_i)} \mathbb{1}\{\Delta Y_{t,i} \leq \Delta y\}\right]$$

converges uniformly to

$$\mathrm{E}\left[\frac{(1 - D_t)}{p} \frac{p(X)}{1 - p(X)} \mathbb{1}\{\Delta Y_t \leq \Delta y\}\right]$$

$\square$

A.4 Asymptotic Normality

In this section we derive the asymptotic distribution of $\hat{\mathrm{QTET}}$.

We make use of several lemmas in this section and state these first.

Lemma 7. *(i)* $\sqrt{n}(\hat{F}_X(x) - F_X(x)) \xrightarrow{d} N(0, p(1 - p))$ *where* $p = F_X(x)$, *and (ii)* $\sqrt{n}(\hat{F}_X(x) - F_X(x)) = \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}\{X_i \leq x\} - \mathrm{E}[\mathbb{1}\{X_i \leq x\}]\right)$.

Lemma 8. *For* $0 \leq \tau \leq 1$, *(i)* $\sqrt{n}(\hat{F}_X^{-1}(\tau) - F_X^{-1}(\tau)) \xrightarrow{d} N\left(0, \frac{\tau(1-\tau)}{f_X^2(F_X^{-1}(\tau))}\right)$, *and (ii)* $\sqrt{n}(\hat{F}_X^{-1}(\tau) - F_X^{-1}(\tau)) = \frac{1}{f_X(F_X^{-1}(\tau))}\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}\{X_i \leq F_X^{-1}(\tau)\} - \mathrm{E}[\mathbb{1}\{X_i \leq F_X^{-1}(\tau)\}]\right) + o_p(1)$

Proof. See Van der Vaart (2000)[p.307-308] $\square$

Lemma 9. *Uniform Convergence of empricial distribution For* $0 \leq \delta \leq 1/2$, $\sup_x n^\delta |\hat{F}_X(x) - F_X(x)| \xrightarrow{p} 0$

Lemma 10. *Uniform Convergence of empricial quantiles For* $0 \leq \delta \leq 1/2$, $\sup_q n^\delta |\hat{F}_X^{-1}(q) - F_X^{-1}(q)| \xrightarrow{p} 0$

Lemma 11. $\sqrt{n}(\hat{F}_X(\hat{F}_X^{-1}(\tau)) - \tau) \xrightarrow{p} 0$

Proof. From the definitions of empirical distributions and empirical quantiles, it is easy to see that $0 \leq \hat{F}_X(\hat{F}_X^{-1}(\tau)) - \tau \leq \frac{1}{n}$ which implies the result. $\qquad\square$

Lemma 12. $\sup_x \sqrt{n}|\hat{F}_X(x) - F_X(x) - f_X(x)(\hat{F}_X^{-1}(F_X(x)) - x)| \xrightarrow{p} 0$

Proof. The result holds because

$$
\begin{aligned}
\hat{F}_X^{-1}(F_X(x)) - x &= \hat{F}_X^{-1}(F_X(x)) - F_X^{-1}(F_X(x)) \\
&= \frac{1}{f_X(F_X^{-1}(F_X(x)))} \left(\hat{F}_X(F_X^{-1}(F_X(x))) - F_X(F_X^{-1}(F_X(x))) \right) + o_p(1/\sqrt{n}) \\
&= \frac{1}{f_X(x)} \left(\hat{F}_X(x) - F_X(x) \right) + o_p(1/\sqrt{n})
\end{aligned}
$$

where the second equality uses Lemma 4. $\qquad\square$

Lemma 13. *For* $\delta > 1/2$, $y \in \mathcal{Y}$, $(y + x) \in \mathcal{Y}$, $\sup_{x \leq n^{-\delta}} \sqrt{n}|\hat{F}_Y(y + x) - \hat{F}_Y(y) - (F_Y(y + x) - F_Y(y))| \xrightarrow{p} 0$.

Proof. This is a special case of Lemma A.5 in Athey and Imbens (2006) $\qquad\square$

Lemma 14 and Lemma 15 are helpful to work with empirical distributions. These lemmas are not strictly needed as the identification result could be adjusted so that there would be no need for these two lemmas, but having them makes for cleaner presentation. Their proofs are omitted.

Lemma 14.

$$
\sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \mathbb{1}\{\hat{F}_Y(X_i) \leq q\} - \frac{1}{n} \sum_{i=1}^n \mathbb{1}\{X_i \leq \hat{F}_Y^{-1}(q)\} \right) \xrightarrow{p} 0
$$

Lemma 15.

$$
\sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \mathbb{1}\{\hat{F}_Z^{-1}(\hat{F}_Y(X_i)) \leq z\} - \frac{1}{n} \sum_{i=1}^n \mathbb{1}\{X_i \leq \hat{F}_Y^{-1}(\hat{F}_Z(z))\} \right) \xrightarrow{p} 0
$$

Lemma 16. *If* $f_Z(z)$, $f_Y(y)$, *and* $\frac{\partial f_Z}{\partial z}(z)$ *are bounded, then*

$$
\sup_x \sqrt{n} \left| F_Z(\hat{F}_Y^{-1}(\hat{F}_X(x))) - F_Z(F_Y^{-1}(\hat{F}_X(x))) - \left\{ F_Z(\hat{F}_Y^{-1}(F_X(x))) - F_Z(F_Y^{-1}(F_X(x))) \right\} \right| \xrightarrow{p} 0
$$

Proof. First, note that by Taylor expansions

$$
F_Z(\hat{F}_Y^{-1}(\hat{F}_X(x))) - F_Z(F_Y^{-1}(\hat{F}_X(x))) = f_Z(F_Y^{-1}(\hat{F}_X(x)))(\hat{F}_Y^{-1}(\hat{F}_X(x)) - F_Y^{-1}(\hat{F}_X(x))) + o_p(1)
$$

$$
\tag{29}
$$

and

$$F_Z(\hat{F}_Y^{-1}(F_X(x))) - F_Z(F_Y^{-1}(F_X(x))) = f_Z(F_Y^{-1}(F_X(x)))(\hat{F}_Y^{-1}(F_X(x)) - F_Y^{-1}(F_X(x))) + o_p(1)$$
$$= \left(f_Z(F_Y^{-1}(F_X(x))) - f_Z(F_Y^{-1}(\hat{F}_X(x))) \right) (\hat{F}_Y^{-1}(F_X(x)) - F_Y^{-1}(F_X(x)))$$
$$+ f_Z(F_Y^{-1}(\hat{F}_X(x)))(\hat{F}_Y^{-1}(F_X(x)) - F_Y^{-1}(F_X(x)))) + o_p(1)$$
$$(30)$$

This implies that

$$\sup_x \sqrt{n} \left| F_Z(\hat{F}_Y^{-1}(\hat{F}_X(x))) - F_Z(F_Y^{-1}(\hat{F}_X(x))) - \left\{ F_Z(\hat{F}_Y^{-1}(F_X(x))) - F_Z(F_Y^{-1}(F_X(x))) \right\} \right|$$
$$\leq \sup_x \sqrt{n} \left| f_Z(F_Y^{-1}(\hat{F}_X(x))) \left(\hat{F}_Y^{-1}(\hat{F}_X(x)) - F_Y^{-1}(\hat{F}_X(x)) \right) - \left\{ \hat{F}_Y^{-1}(F_X(x)) - F_Y^{-1}(F_X(x)) \right\} \right) \right|$$
$$+ \sup_x \sqrt{n} \left| \left(f_Z(F_Y^{-1}(F_X(x))) - f_Z(F_Y^{-1}(\hat{F}_X(x))) \right) (\hat{F}_Y^{-1}(F_X(x)) - F_Y^{-1}(F_X(x))) \right| + o_p(1)$$
$$(31)$$

$$\xrightarrow{p} 0$$

The first term in Equation 31 converges to 0 because $f_Z(F_Y^{-1}(\hat{F}_X(x)))$ is bounded by assumption and Lemma 13 implies $\sqrt{n} \left(\hat{F}_Y^{-1}(\hat{F}_X(x)) - F_Y^{-1}(\hat{F}_X(x))) - \left\{ \hat{F}_Y^{-1}(F_X(x)) - F_Y^{-1}(F_X(x))) \right\} \right)$ converges to 0. For the second term, $(\hat{F}_Y^{-1}(F_X(x)) - F_Y^{-1}(F_X(x)))$ is clearly $O_p(1/\sqrt{n})$. The term $\left(f_Z(F_Y^{-1}(F_X(x))) - f_Z(F_Y^{-1}(\hat{F}_X(x))) \right)$ is also $O_p(1/\sqrt{n})$ which can be seen by taking a Taylor approximation and using the assumptions that $f_Y(y)$ and $\frac{\partial f_Z}{\partial z}(z)$ are bounded. This implies the result.

$$\sqrt{n} \left(\hat{F}_Y^{-1}(\hat{F}_X(x)) - F_Y^{-1}(\hat{F}_X(x))) - \left\{ \hat{F}_Y^{-1}(F_X(x)) - F_Y^{-1}(F_X(x))) \right\} \right)$$
$$= \frac{1}{f_Y(F_Y^{-1}(\hat{F}_X(x)))} \left\{ \hat{F}_Y(F_Y^{-1}(\hat{F}_X(x))) - F_Y(F_Y^{-1}(\hat{F}_X(x))) \right\}$$
$$- \frac{1}{f_Y(F_Y^{-1}(F_X(x)))} \left\{ \hat{F}_Y(F_Y^{-1}(F_X(x))) - F_Y(F_Y^{-1}(F_X(x))) \right\}$$

$\square$

Lemma 17. *One Sample V-Statistic*

$$\frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n g(X_i, Y_j) = \frac{1}{n} \sum_{i=1}^n g_1(X_i) + \frac{1}{n} \sum_{i=1}^n g_2(Y_i) - \mu + o_p(1)$$

where $g_1(x) = \mathrm{E}[g(x, Y)]$, $g_2(y) = \mathrm{E}[g(X, y)]$, and $\mu = \mathrm{E}[g(X, Y)]$.

Lemma 18. *Two Sample V-Statistic*

$$\frac{1}{n_1 n_2} \sum_{i \in G_1} \sum_{j \in G_2} g(X_i, Y_j) = \frac{1}{n_1} \sum_{i=1}^{n_1} g_1(X_i) + \frac{1}{n_2} \sum_{i=1}^{n_2} g_2(Y_i) - \mu + o_p(1)$$

where $g_1(x) = \mathrm{E}[g(x, Y)]$, $g_2(y) = \mathrm{E}[g(X, y)]$, and $\mu = \mathrm{E}[g(X, Y)]$.

Proof. The proofs of Lemma 17 and Lemma 18 are omitted as these are well known results. Useful references are Newey and McFadden (1994, p. 2200), Lee (1990), and Van der Vaart (2000). $\quad\square$

Lemma 19. *Asymptotic Representation of $\sqrt{n}\left(\hat{P}(Y_{0t} \le y | D_t = 1) - P(Y_{0t} \le y | D_t = 1)\right)$*
   Let $\hat{\mu}_1$, $\hat{\mu}_2$, $\hat{\mu}_3$, $\hat{\mu}_4$, and $\hat{\mu}_5$ be defined as in the main text and restated here:[11]

$$\hat{\mu}_1 = \frac{1}{n_T n_C} \sum_{i \in \mathcal{T}} \sum_{j \in \mathcal{C}} \mathbb{1}\{\Delta Y_{0t,j} \le (y - F_{Y_{0t-1}|D_t=1}^{-1}(F_{Y_{0t-2}|D_t=1}(Y_{0t-2,i})))\}$$
$$- F_{\Delta Y_{0t}|D_t=0}(y - F_{Y_{0t-1}|D_t=1}^{-1}(F_{Y_{0t-2}|D_t=1}(Y_{0t-2,i})))$$
$$\equiv \frac{1}{n_T n_C} \sum_{i \in \mathcal{T}} \sum_{j \in \mathcal{C}} \phi_1(Y_{0t-2,i}, \Delta Y_{0t,j})$$

$$\hat{\mu}_2 = \frac{1}{n_T^2} \sum_{i \in \mathcal{T}} \sum_{j \in \mathcal{T}} \mathbb{1}\{\Delta Y_{0t-1,j} \le F_{\Delta Y_{0t-1}|D_t=1}^{-1}(F_{\Delta Y_{0t}|D_t=0}(y - F_{Y_{0t-1}|D_t=1}^{-1}(F_{Y_{0t-2}|D_t=1}(Y_{0t-2,i}))))\}$$
$$- F_{\Delta Y_{0t-1}|D_t=1}(F_{\Delta Y_{0t-1}|D_t=1}^{-1}(F_{\Delta Y_{0t}|D_t=0}(y - F_{Y_{0t-1}|D_t=1}^{-1}(F_{Y_{0t-2}|D_t=1}(Y_{0t-2,i})))))]$$
$$\equiv \frac{1}{n_T^2} \sum_{i \in \mathcal{T}} \sum_{j \in \mathcal{T}} \phi_2(Y_{t-2,i}, \Delta Y_{t-1,j})$$

$$\hat{\mu}_3 = \frac{1}{n_T^2} \sum_{i \in \mathcal{T}} \sum_{j \in \mathcal{T}} \frac{f_{\Delta Y_{0t}|D_t=0}(y - F_{Y_{0t-1}|D_t=1}^{-1}(F_{Y_{0t-2}|D_t=1}(Y_{0t-2,i})))}{f_{Y_{0t-1}|D_t=1}(F_{Y_{0t-2}|D_t=1}(Y_{0t-2,i}))}$$
$$\times \left(\mathbb{1}\{Y_{0t-1,j} \le F_{Y_{0t-1}|D_t=1}^{-1}(F_{Y_{0t-2}|D_t=1}(Y_{0t-2,i}))\}\right.$$
$$\left. - F_{Y_{0t-1}|D_t=1}(F_{Y_{0t-1}|D_t=1}^{-1}(F_{Y_{0t-2}|D_t=1}(Y_{0t-2,i})))\right)$$
$$\equiv \frac{1}{n_T^2} \sum_{i \in \mathcal{T}} \sum_{j \in \mathcal{T}} \phi_3(Y_{t-1,i}, Y_{t-2,j})$$

$$\hat{\mu}_4 = \frac{1}{n_T^2} \sum_{i \in \mathcal{T}} \sum_{j \in \mathcal{T}} \frac{f_{\Delta Y_{0t}|D_t=0}(y - F_{Y_{0t-1}|D_t=1}^{-1}(F_{Y_{0t-2}|D_t=1}(Y_{0t-2,i})))}{f_{Y_{0t-1}|D_t=1}(F_{Y_{0t-1}|D_t=1}^{-1}(F_{Y_{0t-2}|D_t=1}(Y_{0t-2,i})))}$$
$$\times \left(\mathbb{1}\{Y_{0t-2,j} \le Y_{0t-2,i}\} - F_{Y_{0t-2}|D_t=1}(Y_{0t-2,i})\right)$$

---

[11]It should be noted that each $\hat{\mu}$ and $\mu_j(\cdot)$ depends on the value of y. We suppress this dependence throughout each of the Lemmas

$$\equiv \frac{1}{n_T^2} \sum_{i \in \mathcal{T}} \sum_{j \in \mathcal{T}} \phi_4(Y_{t-2,i}, Y_{t-2,j})$$

$$\hat{\mu}_5 = \frac{1}{n_T} \sum_{i \in \mathcal{T}} \left[\mathbb{1}\{F_{\Delta Y_{0t}|D_t=0}^{-1}(F_{\Delta Y_{0t-1}|D_t=1}(\Delta Y_{0t-1,i})) \leq y - F_{Y_{0t-1}|D_t=1}^{-1}(F_{Y_{0t-2}|D_t=1}(Y_{0t-2,i)))\} \right]$$

$$- \mathrm{E}\left[\mathbb{1}\{F_{\Delta Y_{0t}|D_t=0}^{-1}(F_{\Delta Y_{0t-1}|D_t=1}(\Delta Y_{0t-1,i})) \leq y - F_{Y_{0t-1}|D_t=1}^{-1}(F_{Y_{0t-2}|D_t=1}(Y_{0t-2,i)))\}|D_t = 1 \right]$$

$$\equiv \frac{1}{n_T} \sum_{i \in \mathcal{T}} \phi_5(\Delta Y_{t-1,i}, Y_{0t-2,i})$$

Then,

$$\sqrt{n}\left(\hat{P}(Y_{0t} \leq y|D_t = 1) - P(Y_{0t} \leq y|D_t = 1) - \hat{\mu}_1 - \hat{\mu}_2 - \hat{\mu}_3 - \hat{\mu}_4 - \hat{\mu}_5 \right) \xrightarrow{p} 0$$

Proof. To prove the lemma, we add and subtract a number of terms and then show that each term converges in probability to 0.

$$\sqrt{n}\left(\hat{P}(Y_{0t} \leq y|D_t = 1) - P(Y_{0t} \leq y|D_t = 1) - \hat{\mu}_1 - \hat{\mu}_2 - \hat{\mu}_3 - \hat{\mu}_4 - \hat{\mu}_5 \right)$$

$$= \sqrt{n}\left(\frac{1}{n_T} \sum_{i \in \mathcal{T}} \left[\mathbb{1}\{\hat{F}_{\Delta Y_{0t}|D_t=0}^{-1}(\hat{F}_{\Delta Y_{0t-1}|D_t=1}(\Delta Y_{0t-1,i})) \leq y - \hat{F}_{Y_{0t-1}|D_t=1}^{-1}(\hat{F}_{Y_{0t-2}|D_t=1}(Y_{0t-2,i)))\} \right] \right.$$

$$- \mathrm{E}\left[\mathbb{1}\{F_{\Delta Y_{0t}|D_t=0}^{-1}(F_{\Delta Y_{0t-1}|D_t=1}(\Delta Y_{0t-1,i})) \leq y - F_{Y_{0t-1}|D_t=1}^{-1}(F_{Y_{0t-2}|D_t=1}(Y_{0t-2,i)))\}|D_t = 1 \right]$$

$$\left. -\hat{\mu}_1 - \hat{\mu}_2 - \hat{\mu}_3 - \hat{\mu}_4 - \hat{\mu}_5 \right) \tag{32}$$

$$= \sqrt{n}\left(\frac{1}{n_T} \sum_{i \in \mathcal{T}} \left[\mathbb{1}\{\hat{F}_{\Delta Y_{0t}|D_t=0}^{-1}(\hat{F}_{\Delta Y_{0t-1}|D_t=1}(\Delta Y_{0t-1,i})) \leq y - \hat{F}_{Y_{0t-1}|D_t=1}^{-1}(\hat{F}_{Y_{0t-2}|D_t=1}(Y_{0t-2,i)))\} \right] \right.$$

$$\left. - \frac{1}{n_T} \sum_{i \in \mathcal{T}} \left[\mathbb{1}\{F_{\Delta Y_{0t}|D_t=0}^{-1}(\hat{F}_{\Delta Y_{0t-1}|D_t=1}(\Delta Y_{0t-1,i})) \leq y - \hat{F}_{Y_{0t-1}|D_t=1}^{-1}(\hat{F}_{Y_{0t-2}|D_t=1}(Y_{0t-2,i)))\} \right] - \hat{\mu}_1 \right) \tag{33}$$

$$+ \sqrt{n}\left(\frac{1}{n_T} \sum_{i \in \mathcal{T}} \left[\mathbb{1}\{F_{\Delta Y_{0t}|D_t=0}^{-1}(\hat{F}_{\Delta Y_{0t-1}|D_t=1}(\Delta Y_{0t-1,i})) \leq y - \hat{F}_{Y_{0t-1}|D_t=1}^{-1}(\hat{F}_{Y_{0t-2}|D_t=1}(Y_{0t-2,i)))\} \right] \right.$$

$$\left. - \frac{1}{n_T} \sum_{i \in \mathcal{T}} \left[\mathbb{1}\{F_{\Delta Y_{0t}|D_t=0}^{-1}(F_{\Delta Y_{0t-1}|D_t=1}(\Delta Y_{0t-1,i})) \leq y - \hat{F}_{Y_{0t-1}|D_t=1}^{-1}(\hat{F}_{Y_{0t-2}|D_t=1}(Y_{0t-2,i)))\} \right] - \hat{\mu}_2 \right) \tag{34}$$

$$+ \sqrt{n}\left(\frac{1}{n_T} \sum_{i \in \mathcal{T}} \left[\mathbb{1}\{F_{\Delta Y_{0t}|D_t=0}^{-1}(F_{\Delta Y_{0t-1}|D_t=1}(\Delta Y_{0t-1,i})) \leq y - \hat{F}_{Y_{0t-1}|D_t=1}^{-1}(\hat{F}_{Y_{0t-2}|D_t=1}(Y_{0t-2,i)))\} \right] \right.$$

$$\left. - \frac{1}{n_T} \sum_{i \in \mathcal{T}} \left[\mathbb{1}\{F_{\Delta Y_{0t}|D_t=0}^{-1}(F_{\Delta Y_{0t-1}|D_t=1}(\Delta Y_{0t-1,i})) \leq y - F_{Y_{0t-1}|D_t=1}^{-1}(\hat{F}_{Y_{0t-2}|D_t=1}(Y_{0t-2,i)))\} \right] - \hat{\mu}_3 \right) \tag{35}$$

$$+ \sqrt{n} \left(\frac{1}{n_T} \sum_{i \in \mathcal{T}} \left[\mathbb{1}\{F^{-1}_{\Delta Y_{0t}|D_t=0}(F_{\Delta Y_{0t-1}|D_t=1}(\Delta Y_{0t-1,i})) \leq y - F^{-1}_{Y_{0t-1}|D_t=1}(\hat{F}_{Y_{0t-2}|D_t=1}(Y_{0t-2,i}))\} \right] \right.$$

$$\left. - \frac{1}{n_T} \sum_{i \in \mathcal{T}} \left[\mathbb{1}\{F^{-1}_{\Delta Y_{0t}|D_t=0}(F_{\Delta Y_{0t-1}|D_t=1}(\Delta Y_{0t-1,i})) \leq y - F^{-1}_{Y_{0t-1}|D_t=1}(F_{Y_{0t-2}|D_t=1}(Y_{0t-2,i}))\} \right] - \hat{\mu}_4 \right)$$

$$(36)$$

$$+ \sqrt{n} \left(\frac{1}{n_T} \sum_{i \in \mathcal{T}} \left[\mathbb{1}\{F^{-1}_{\Delta Y_{0t}|D_t=0}(F_{\Delta Y_{0t-1}|D_t=1}(\Delta Y_{0t-1,i})) \leq y - F^{-1}_{Y_{0t-1}|D_t=1}(F_{Y_{0t-2}|D_t=1}(Y_{0t-2,i}))\} \right] \right.$$

$$\left. - \mathrm{E} \left[\mathbb{1}\{F^{-1}_{\Delta Y_{0t}|D_t=0}(F_{\Delta Y_{0t-1}|D_t=1}(\Delta Y_{0t-1,i})) \leq y - F^{-1}_{Y_{0t-1}|D_t=1}(F_{Y_{0t-2}|D_t=1}(Y_{0t-2,i}))\}|D_t = 1 \right] - \hat{\mu}_5 \right)$$

$$(37)$$

Next, we show that Equation 33, Equation 34, Equation 35, and Equation 36 each converge to 0. We analyze each equation in turn.

Equation 33:

Recall

$$\hat{\mu}_1 = \frac{1}{n_T n_C} \sum_{i \in \mathcal{T}} \sum_{j \in \mathcal{C}} \mathbb{1}\{\Delta Y_{0t,j} \leq (y - F^{-1}_{Y_{0t-1}|D_t=1}(F_{Y_{0t-2}|D_t=1}(Y_{0t-2,i})))\}$$

$$- F_{\Delta Y_{0t}|D_t=0}(y - F^{-1}_{Y_{0t-1}|D_t=1}(F_{Y_{0t-2}|D_t=1}(Y_{0t-2,i})))$$

Next, notice that

$$= \sqrt{n} \left| \frac{1}{n_T} \sum_{i \in \mathcal{T}} \left[\mathbb{1}\{\hat{F}^{-1}_{\Delta Y_{0t}|D_t=0}(\hat{F}_{\Delta Y_{0t-1}|D_t=1}(\Delta Y_{0t-1,i})) \leq y - \hat{F}^{-1}_{Y_{0t-1}|D_t=1}(\hat{F}_{Y_{0t-2}|D_t=1}(Y_{0t-2,i}))\} \right] \right.$$

$$\left. - \frac{1}{n_T} \sum_{i \in \mathcal{T}} \left[\mathbb{1}\{F^{-1}_{\Delta Y_{0t}|D_t=0}(\hat{F}_{\Delta Y_{0t-1}|D_t=1}(\Delta Y_{0t-1,i})) \leq y - \hat{F}^{-1}_{Y_{0t-1}|D_t=1}(\hat{F}_{Y_{0t-2}|D_t=1}(Y_{0t-2,i}))\} \right] - \hat{\mu}_1 \right|$$

$$(38)$$

$$= \sqrt{n} \left| \frac{1}{n_T} \sum_{i \in \mathcal{T}} \left[\mathbb{1}\{\hat{F}^{-1}_{\Delta Y_{0t}|D_t=0}(\hat{F}_{\Delta Y_{0t-1}|D_t=1}(\Delta Y_{0t-1,i})) \leq g(Y_{0t-2,i})\} \right] \right.$$

$$\left. - \frac{1}{n_T} \sum_{i \in \mathcal{T}} \left[\mathbb{1}\{F^{-1}_{\Delta Y_{0t}|D_t=0}(\hat{F}_{\Delta Y_{0t-1}|D_t=1}(\Delta Y_{0t-1,i})) \leq g(Y_{0t-2,i})\} \right] - \hat{\mu}_1 \right|$$

$$(39)$$

$$\leq \sqrt{n} \left| \frac{1}{n_T} \sum_{i \in \mathcal{T}} \left[\mathbb{1}\{\hat{F}^{-1}_{\Delta Y_{0t}|D_t=0}(\hat{F}_{\Delta Y_{0t-1}|D_t=1}(\Delta Y_{0t-1,i})) \leq g(Y_{0t-2,i})\} \right] \right.$$

$$\left. - \frac{1}{n_T} \sum_{i \in \mathcal{T}} \left[\mathbb{1}\{\Delta Y_{0t-1,i} \leq \hat{F}^{-1}_{\Delta Y_{0t-1}|D_t=1}(\hat{F}_{\Delta Y_{0t}|D_t=0}(g(Y_{0t-2,i})))\} \right] \right|$$

$$(40)$$

$$+ \sqrt{n} \left| \frac{1}{n_T} \sum_{i \in \mathcal{T}} \left[\mathbb{1}\{\Delta Y_{0t-1,i} \leq \hat{F}^{-1}_{\Delta Y_{0t-1}|D_t=1}(F_{\Delta Y_{0t}|D_t=0}(g(Y_{0t-2,i})))\} \right] \right.$$

$$- \frac{1}{n_T} \sum_{i \in \mathcal{T}} \left[\mathbb{1}\{F_{\Delta Y_{0t}|D_t=0}^{-1}(\hat{F}_{\Delta Y_{0t-1}|D_t=1}(\Delta Y_{0t-1,i})) \leq g(Y_{0t-2,i})\} \right] \Bigg| \tag{41}$$

$$+ \sqrt{n} \left| \frac{1}{n_T} \sum_{i \in \mathcal{T}} \left[\mathbb{1}\{\Delta Y_{0t-1,i} \leq \hat{F}_{\Delta Y_{0t-1}|D_t=1}^{-1}(\hat{F}_{\Delta Y_{0t}|D_t=0}(g(Y_{0t-2,i})))\} \right] \right.$$

$$\left. - \frac{1}{n_T} \sum_{i \in \mathcal{T}} \left[\mathbb{1}\{\Delta Y_{0t-1,i} \leq \hat{F}_{\Delta Y_{0t-1}|D_t=1}^{-1}(F_{\Delta Y_{0t}|D_t=0}(g(Y_{0t-2,i})))\} \right] - \hat{\mu}_1 \right| \tag{42}$$

$$\leq \sup_z \sqrt{n} \left| \frac{1}{n_T} \sum_{i \in \mathcal{T}} \left[\mathbb{1}\{\hat{F}_{\Delta Y_{0t}|D_t=0}^{-1}(\hat{F}_{\Delta Y_{0t-1}|D_t=1}(\Delta Y_{0t-1,i})) \leq g(z)\} \right] \right.$$

$$\left. - \frac{1}{n_T} \sum_{i \in \mathcal{T}} \left[\mathbb{1}\{\Delta Y_{0t-1,i} \leq \hat{F}_{\Delta Y_{0t-1}|D_t=1}^{-1}(\hat{F}_{\Delta Y_{0t}|D_t=0}(g(z)))\} \right] \right| \tag{43}$$

$$+ \sup_z \sqrt{n} \left| \frac{1}{n_T} \sum_{i \in \mathcal{T}} \left[\mathbb{1}\{\Delta Y_{0t-1,i} \leq \hat{F}_{\Delta Y_{0t-1}|D_t=1}^{-1}(F_{\Delta Y_{0t}|D_t=0}(g(z)))\} \right] \right.$$

$$\left. - \frac{1}{n_T} \sum_{i \in \mathcal{T}} \left[\mathbb{1}\{F_{\Delta Y_{0t}|D_t=0}^{-1}(\hat{F}_{\Delta Y_{0t-1}|D_t=1}(\Delta Y_{0t-1,i})) \leq g(z)\} \right] \right| \tag{44}$$

$$+ \sqrt{n} \left| \frac{1}{n_T} \sum_{i \in \mathcal{T}} \left[\mathbb{1}\{\Delta Y_{0t-1,i} \leq \hat{F}_{\Delta Y_{0t-1}|D_t=1}^{-1}(\hat{F}_{\Delta Y_{0t}|D_t=0}(g(Y_{0t-2,i})))\} \right] \right.$$

$$\left. - \frac{1}{n_T} \sum_{i \in \mathcal{T}} \left[\mathbb{1}\{\Delta Y_{0t-1,i} \leq \hat{F}_{\Delta Y_{0t-1}|D_t=1}^{-1}(F_{\Delta Y_{0t}|D_t=0}(g(Y_{0t-2,i})))\} \right] - \hat{\mu}_1 \right| \tag{45}$$

$$\leq \sqrt{n} \left| \frac{1}{n_T} \sum_{i \in \mathcal{T}} \left[\mathbb{1}\{\Delta Y_{0t-1,i} \leq \hat{F}_{\Delta Y_{0t-1}|D_t=1}^{-1}(\hat{F}_{\Delta Y_{0t}|D_t=0}(g(Y_{0t-2,i})))\} \right] \right.$$

$$- \frac{1}{n_T} \sum_{i \in \mathcal{T}} \left[\mathbb{1}\{\Delta Y_{0t-1,i} \leq \hat{F}_{\Delta Y_{0t-1}|D_t=1}^{-1}(F_{\Delta Y_{0t}|D_t=0}(g(Y_{0t-2,i})))\} \right]$$

$$\left. - \frac{1}{n_T n_C} \sum_{i \in \mathcal{T}} \sum_{j \in \mathcal{C}} \mathbb{1}\{\Delta Y_{0t,j} \leq g(Y_{0t-2,i})\} - F_{\Delta Y_{0t}|D_t=0}(g(Y_{0t-2,i})) \right| \tag{46}$$

$$+ \sqrt{n} \left| \frac{1}{n_T n_C} \sum_{i \in \mathcal{T}} \sum_{j \in \mathcal{C}} \left[\mathbb{1}\{\Delta Y_{0t,j} \leq y - \hat{F}_{Y_{0t-1}|D_t=1}^{-1}(\hat{F}_{Y_{0t-2}|D_t=1}(Y_{0t-1,i}))\} \right. \right.$$

$$\left. - F_{\Delta Y_{0t}|D_t=0}(y - \hat{F}_{Y_{0t-1}|D_t=1}^{-1}(\hat{F}_{Y_{0t-2}|D_t=1}(Y_{0t-1,i}))) \right]$$

$$- \frac{1}{n_T n_C} \sum_{i \in \mathcal{T}} \sum_{j \in \mathcal{C}} \left[\mathbb{1}\{\Delta Y_{0t,j} \leq y - F_{Y_{0t-1}|D_t=1}^{-1}(F_{Y_{0t-2}|D_t=1}(Y_{0t-1,i}))\} \right.$$

$$\left. \left. - F_{\Delta Y_{0t}|D_t=0}(y - F_{Y_{0t-1}|D_t=1}^{-1}(F_{Y_{0t-2}|D_t=1}(Y_{0t-1,i}))) \right] \right| + o_p(1) \tag{47}$$

$$\leq \sup_z \sqrt{n} \left| \hat{F}_{\Delta Y_{0t-1}|D_t=1}(\hat{F}_{\Delta Y_{0t-1}|D_t=1}^{-1}(\hat{F}_{\Delta Y_{0t}|D_t=0}(g(z)))) \right.$$

$$- \hat{F}_{\Delta Y_{0t-1}|D_t=1}(\hat{F}_{\Delta Y_{0t-1}|D_t=1}^{-1}(F_{\Delta Y_{0t}|D_t=0}(g(z))))$$

$$\left. - \left\{ \hat{F}_{\Delta Y_{0t}|D_t=0}(g(z)) - F_{\Delta Y_{0t}|D_t=0}(g(z)) \right\} \right| \tag{48}$$

$$+ \sup_z \sqrt{n} \left| \hat{F}_{\Delta Y_{0t}|D_t=0}(y - \hat{F}^{-1}_{Y_{0t-1}|D_t=1}(\hat{F}_{Y_{0t-2}|D_t=1}(z))) \right.$$

$$- F_{\Delta Y_{0t}|D_t=0}(y - \hat{F}^{-1}_{Y_{0t-1}|D_t=1}(\hat{F}_{Y_{0t-2}|D_t=1}(z)))$$

$$- \hat{F}_{\Delta Y_{0t}|D_t=0}(y - F^{-1}_{Y_{0t-1}|D_t=1}(F_{Y_{0t-2}|D_t=1}(z)))$$

$$\left. - F_{\Delta Y_{0t}|D_t=0}(y - F^{-1}_{Y_{0t-1}|D_t=1}(F_{Y_{0t-2}|D_t=1}(z))) \right| + o_p(1) \qquad (49)$$

$$\xrightarrow{p} 0$$

Equation 43 and Equation 44 converge to 0 by Lemma 14 and Lemma 15, respectively. Equation 48 converges to 0 by an application of Lemma 11 followed by some simple cancellations. Equation 49 converges to 0 by Lemma 13.

Equation 34: First, recall that

$$\hat{\mu}_2 = \frac{1}{n_T^2} \sum_{i \in \mathcal{T}} \sum_{j \in \mathcal{T}} \mathbb{1}\{\Delta Y_{0t-1,j} \leq F^{-1}_{\Delta Y_{0t-1}|D_t=1}(F_{\Delta Y_{0t}|D_t=0}(y - F^{-1}_{Y_{0t-1}|D_t=1}(F_{Y_{0t-2}|D_t=1}(Y_{0t-2,i}))))\}$$

$$- F_{\Delta Y_{0t-1}|D_t=1}(F^{-1}_{\Delta Y_{0t-1}|D_t=1}(F_{\Delta Y_{0t}|D_t=0}(y - F^{-1}_{Y_{0t-1}|D_t=1}(F_{Y_{0t-2}|D_t=1}(Y_{0t-2,i})))))]$$

Then,

$$\sqrt{n} \left(\frac{1}{n_T} \sum_{i \in \mathcal{T}} \left[\mathbb{1}\{F^{-1}_{\Delta Y_{0t}|D_t=0}(\hat{F}_{\Delta Y_{0t-1}|D_t=1}(\Delta Y_{0t-1,i})) \leq y - \hat{F}^{-1}_{Y_{0t-1}|D_t=1}(\hat{F}_{Y_{0t-2}|D_t=1}(Y_{0t-2,i}))\} \right] \right.$$

$$\left. - \frac{1}{n_T} \sum_{i \in \mathcal{T}} \left[\mathbb{1}\{F^{-1}_{\Delta Y_{0t}|D_t=0}(F_{\Delta Y_{0t-1}|D_t=1}(\Delta Y_{0t-1,i})) \leq y - \hat{F}^{-1}_{Y_{0t-1}|D_t=1}(\hat{F}_{Y_{0t-2}|D_t=1}(Y_{0t-2,i}))\} \right] - \hat{\mu}_2 \right)$$

$$\leq \sqrt{n} \left| \frac{1}{n_T} \sum_{i \in \mathcal{T}} \left[\mathbb{1}\{F^{-1}_{\Delta Y_{0t}|D_t=0}(\hat{F}_{\Delta Y_{0t-1}|D_t=1}(\Delta Y_{0t-1,i})) \leq y - \hat{F}^{-1}_{Y_{0t-1}|D_t=1}(\hat{F}_{Y_{0t-2}|D_t=1}(Y_{0t-2,i}))\} \right] \right.$$

$$\left. - \frac{1}{n_T} \sum_{i \in \mathcal{T}} \left[\mathbb{1}\{\Delta Y_{0t-1,i} \leq \hat{F}^{-1}_{\Delta Y_{0t-1}|D_t=1}(F_{\Delta Y_{0t}|D_t=0}(y - \hat{F}^{-1}_{Y_{0t-1}|D_t=1}(\hat{F}_{Y_{0t-2}|D_t=1}(Y_{0t-2,i}))))\} \right] \right|$$

$$- \left| \frac{1}{n_T} \sum_{i \in \mathcal{T}} \left[\mathbb{1}\{\Delta Y_{0t-1,i} \leq \hat{F}^{-1}_{\Delta Y_{0t-1}|D_t=1}(F_{\Delta Y_{0t}|D_t=0}(y - \hat{F}^{-1}_{Y_{0t-1}|D_t=1}(\hat{F}_{Y_{0t-2}|D_t=1}(Y_{0t-2,i}))))\} \right] \right.$$

$$\left. - \frac{1}{n_T} \sum_{i \in \mathcal{T}} \left[\mathbb{1}\{\Delta Y_{0t-1,i} \leq F^{-1}_{\Delta Y_{0t-1}|D_t=1}(F_{\Delta Y_{0t}|D_t=0}(y - \hat{F}^{-1}_{Y_{0t-1}|D_t=1}(\hat{F}_{Y_{0t-2}|D_t=1}(Y_{0t-2,i}))))\} \right] - \hat{\mu}_2 \right|$$

$$\qquad (50)$$

$$\leq \sup_z \sqrt{n} \left| \frac{1}{n_T} \sum_{i \in \mathcal{T}} \left[\Delta Y_{0t-1,i} \leq \hat{F}^{-1}_{\Delta Y_{0t-1}|D_t=1}(F_{\Delta Y_{0t}|D_t=0}(g(z))) \right] \right.$$

$$- \frac{1}{n_T} \sum_{i \in \mathcal{T}} \left[\Delta Y_{0t-1,i} \leq F^{-1}_{\Delta Y_{0t-1}|D_t=1}(F_{\Delta Y_{0t}|D_t=0}(g(z))) \right]$$

$$- \frac{1}{n_T^2} \sum_{i \in \mathcal{T}} \sum_{j \in \mathcal{T}} \mathbb{1}\{\Delta Y_{0t-1,j} \leq F^{-1}_{\Delta Y_{0t-1}|D_t=1}(F_{\Delta Y_{0t}|D_t=0}(y - F^{-1}_{Y_{0t-1}|D_t=1}(F_{Y_{0t-2}|D_t=1}(z))))\}$$

$$-F_{\Delta Y_{0t-1}|D_t=1}(F^{-1}_{\Delta Y_{0t-1}|D_t=1}(F_{\Delta Y_{0t}|D_t=0}(y - F^{-1}_{Y_{0t-1}|D_t=1}(F_{Y_{0t-2}|D_t=1}(z)))))] + o_p(1)$$

(51)

$$\leq \sup_z \sqrt{n} \left| \hat{F}_{\Delta Y_{0t-1}|D_t=1}(\hat{F}^{-1}_{\Delta Y_{0t-1}|D_t=1}(F_{\Delta Y_{0t}|D_t=0}(g(z)))) \right.$$

$$- \hat{F}_{\Delta Y_{0t-1}|D_t=1}(F^{-1}_{\Delta Y_{0t-1}|D_t=1}(F_{\Delta Y_{0t}|D_t=0}(g(z))))$$

$$- \left(F_{\Delta Y_{0t-1}|D_t=1}(\hat{F}^{-1}_{\Delta Y_{0t-1}|D_t=1}(F_{\Delta Y_{0t}|D_t=0}(g(z)))) \right.$$

$$\left. \left. - F_{\Delta Y_{0t-1}|D_t=1}(F^{-1}_{\Delta Y_{0t-1}|D_t=1}(F_{\Delta Y_{0t}|D_t=0}(g(z)))) \right) \right|$$

(52)

$$+ \sup_z \sqrt{n} \left| F_{\Delta Y_{0t-1}|D_t=1}(\hat{F}^{-1}_{\Delta Y_{0t-1}|D_t=1}(F_{\Delta Y_{0t}|D_t=0}(g(z)))) \right.$$

$$- F_{\Delta Y_{0t-1}|D_t=1}(F^{-1}_{\Delta Y_{0t-1}|D_t=1}(F_{\Delta Y_{0t}|D_t=0}(g(z)))) \right)$$

$$- \left(F_{\Delta Y_{0t-1}|D_t=1}(\hat{F}^{-1}_{\Delta Y_{0t-1}|D_t=1}(F_{\Delta Y_{0t}|D_t=0}(y - F^{-1}_{Y_{0t-1}|D_t=1}(F_{Y_{0t-2}|D_t=1}(z)))))) \right.$$

$$\left. \left. - F_{\Delta Y_{0t-1}|D_t=1}(F^{-1}_{\Delta Y_{0t-1}|D_t=1}(F_{\Delta Y_{0t}|D_t=0}(y - F^{-1}_{Y_{0t-1}|D_t=1}(F_{Y_{0t-2}|D_t=1}(z)))))) \right) \right|$$

(53)

$$+ \sup_z \sqrt{n} \left| F_{\Delta Y_{0t-1}|D_t=1}(\hat{F}^{-1}_{\Delta Y_{0t-1}|D_t=1}(F_{\Delta Y_{0t}|D_t=0}(y - F^{-1}_{Y_{0t-1}|D_t=1}(F_{Y_{0t-2}|D_t=1}(z)))))) \right.$$

$$- F_{\Delta Y_{0t-1}|D_t=1}(F^{-1}_{\Delta Y_{0t-1}|D_t=1}(F_{\Delta Y_{0t}|D_t=0}(y - F^{-1}_{Y_{0t-1}|D_t=1}(F_{Y_{0t-2}|D_t=1}(z)))))) \right)$$

$$- F_{\Delta Y_{0t-1}|D_t=1}(\hat{F}^{-1}_{\Delta Y_{0t-1}|D_t=1}(F_{\Delta Y_{0t}|D_t=0}(y - F^{-1}_{Y_{0t-1}|D_t=1}(F_{Y_{0t-2}|D_t=1}(z))))$$

$$\left. - F_{\Delta Y_{0t-1}|D_t=1}(F^{-1}_{\Delta Y_{0t-1}|D_t=1}(F_{\Delta Y_{0t}|D_t=0}(y - F^{-1}_{Y_{0t-1}|D_t=1}(F_{Y_{0t-2}|D_t=1}(z)))))] \right| + o_p(1)$$

(54)

$$\xrightarrow{p} 0$$

where Equation 52 converges to 0 by Lemma 13, Equation 53 converges to 0 by several Taylor expansions (the result is similar to the proof of Lemma 16), and Equation 54 converges to 0 by first noticing the following Taylor expansion

$$\sqrt{n} \left(F_{\Delta Y_{0t-1}|D_t=1}(\hat{F}^{-1}_{\Delta Y_{0t-1}|D_t=1}(F_{\Delta Y_{0t}|D_t=0}(y - F^{-1}_{Y_{0t-1}|D_t=1}(F_{Y_{0t-2}|D_t=1}(z))))) \right.$$

$$\left. - F_{\Delta Y_{0t-1}|D_t=1}(F^{-1}_{\Delta Y_{0t-1}|D_t=1}(F_{\Delta Y_{0t}|D_t=0}(y - F^{-1}_{Y_{0t-1}|D_t=1}(F_{Y_{0t-2}|D_t=1}(z))))) \right)$$

$$= f_{\Delta Y_{0t-1}|D_t=1}(F^{-1}_{\Delta Y_{0t-1}|D_t=1}(F_{\Delta Y_{0t}|D_t=0}(y - F^{-1}_{Y_{0t-1}|D_t=1}(F_{Y_{0t-2}|D_t=1}(z)))))$$

$$\times \sqrt{n} \left(\hat{F}^{-1}_{\Delta Y_{0t-1}|D_t=1}(F_{\Delta Y_{0t}|D_t=0}(y - F^{-1}_{Y_{0t-1}|D_t=1}(F_{Y_{0t-2}|D_t=1}(z)))) \right.$$

$$\left. - F^{-1}_{\Delta Y_{0t-1}|D_t=1}(F_{\Delta Y_{0t}|D_t=0}(y - F^{-1}_{Y_{0t-1}|D_t=1}(F_{Y_{0t-2}|D_t=1}(z)))) \right) + o_p(1)$$

(55)

and then noting that

$$\sqrt{n} \left(\hat{F}^{-1}_{\Delta Y_{0t-1}|D_t=1}(F_{\Delta Y_{0t}|D_t=0}(y - F^{-1}_{Y_{0t-1}|D_t=1}(F_{Y_{0t-2}|D_t=1}(z)))) \right.$$

$$-F_{\Delta Y_{0t-1}|D_t=1}^{-1}(F_{\Delta Y_{0t}|D_t=0}(y-F_{Y_{0t-1}|D_t=1}^{-1}(F_{Y_{0t-2}|D_t=1}(z))))\Big)$$

$$=\frac{1}{f_{\Delta Y_{0t-1}|D_t=1}(F_{\Delta Y_{0t-1}|D_t=1}^{-1}(F_{\Delta Y_{0t}|D_t=0}(y-F_{Y_{0t-1}|D_t=1}^{-1}(F_{Y_{0t-2}|D_t=1}(z)))))}$$

$$\times\sqrt{n}\Big(\hat{F}_{\Delta Y_{0t-1}|D_t=1}(F_{\Delta Y_{0t-1}|D_t=1}^{-1}(F_{\Delta Y_{0t}|D_t=0}(y-F_{Y_{0t-1}|D_t=1}^{-1}(F_{Y_{0t-2}|D_t=1}(z)))))$$

$$-F_{\Delta Y_{0t}|D_t=0}(y-F_{Y_{0t-1}|D_t=1}^{-1}(F_{Y_{0t-2}|D_t=1}(z)))\Big)\tag{56}$$

which holds by Lemma 8. Combining Equation 55 and Equation 56 completes the result.

Equation 35: Recall that

$$\hat{\mu}_3=\frac{1}{n_T^2}\sum_{i\in\mathcal{T}}\sum_{j\in\mathcal{T}}\frac{f_{\Delta Y_{0t}|D_t=0}(y-F_{Y_{0t-1}|D_t=1}^{-1}(F_{Y_{0t-2}|D_t=1}(Y_{0t-2,i})))}{f_{Y_{0t-1}|D_t=1}(F_{Y_{0t-2}|D_t=1}(Y_{0t-2,i}))}$$

$$\times\Big(\mathbb{1}\{Y_{0t-1,j}\le F_{Y_{0t-1}|D_t=1}^{-1}(F_{Y_{0t-2}|D_t=1}(Y_{0t-2,i}))\}$$

$$-F_{Y_{0t-1}|D_t=1}(F_{Y_{0t-1}|D_t=1}^{-1}(F_{Y_{0t-2}|D_t=1}(Y_{0t-2,i})))\Big)$$

Then,

$$\sqrt{n}\Bigg(\frac{1}{n_T}\sum_{i\in\mathcal{T}}\Big[\mathbb{1}\{F_{\Delta Y_{0t}|D_t=0}^{-1}(F_{\Delta Y_{0t-1}|D_t=1}(\Delta Y_{0t-1,i}))\le y-\hat{F}_{Y_{0t-1}|D_t=1}^{-1}(\hat{F}_{Y_{0t-2}|D_t=1}(Y_{0t-2,i}))\}\Big]$$

$$-\frac{1}{n_T}\sum_{i\in\mathcal{T}}\Big[\mathbb{1}\{F_{\Delta Y_{0t}|D_t=0}^{-1}(F_{\Delta Y_{0t-1}|D_t=1}(\Delta Y_{0t-1,i}))\le y-F_{Y_{0t-1}|D_t=1}^{-1}(\hat{F}_{Y_{0t-2}|D_t=1}(Y_{0t-2,i}))\}\Big]-\hat{\mu}_3\Bigg)$$

$$\le\sqrt{n}\Bigg|\frac{1}{n_T}\sum_{i\in\mathcal{T}}\Big[\mathbb{1}\{\Delta Y_{0t-1,i}\le F_{\Delta Y_{0t-1}|D_t=1}^{-1}(F_{\Delta Y_{0t}|D_t=0}(y-\hat{F}_{Y_{0t-1}|D_t=1}^{-1}(\hat{F}_{Y_{0t-2}|D_t=1}(Y_{0t-2,i}))))\}\Big]$$

$$-\frac{1}{n_T}\sum_{i\in\mathcal{T}}\Big[\mathbb{1}\{\Delta Y_{0t-1,i}\le F_{\Delta Y_{0t-1}|D_t=1}^{-1}(F_{\Delta Y_{0t}|D_t=0}(y-F_{Y_{0t-1}|D_t=1}^{-1}(\hat{F}_{Y_{0t-2}|D_t=1}(Y_{0t-2,i}))))\}\Big]-\hat{\mu}_3\Bigg|$$

$$\le\sup_z\sqrt{n}\Big|\hat{F}_{\Delta Y_{0t-1}|D_t=1}(F_{\Delta Y_{0t-1}|D_t=1}^{-1}(F_{\Delta Y_{0t}|D_t=0}(y-\hat{F}_{Y_{0t-1}|D_t=1}^{-1}(\hat{F}_{Y_{0t-2}|D_t=1}(z)))))$$

$$-\hat{F}_{\Delta Y_{0t-1}|D_t=1}(F_{\Delta Y_{0t-1}|D_t=1}^{-1}(F_{\Delta Y_{0t}|D_t=0}(y-F_{Y_{0t-1}|D_t=1}^{-1}(\hat{F}_{Y_{0t-2}|D_t=1}(z)))))$$

$$-\frac{f_{\Delta Y_{0t}|D_t=0}(y-F_{Y_{0t-1}|D_t=1}^{-1}(F_{Y_{0t-2}|D_t=1}(z)))}{f_{Y_{0t-1}|D_t=1}(F_{Y_{0t-2}|D_t=1}(z))}$$

$$\times\Big(\hat{F}_{Y_{0t-1}|D_t=1}(F_{Y_{0t-1}|D_t=1}^{-1}(F_{Y_{0t-2}|D_t=1}(z)))$$

$$-F_{Y_{0t-1}|D_t=1}(F_{Y_{0t-1}|D_t=1}^{-1}(F_{Y_{0t-2}|D_t=1}(z)))\Big)\Big|\tag{57}$$

$$\le\sup_z\sqrt{n}\Big|\hat{F}_{\Delta Y_{0t-1}|D_t=1}(F_{\Delta Y_{0t-1}|D_t=1}^{-1}(F_{\Delta Y_{0t}|D_t=0}(y-\hat{F}_{Y_{0t-1}|D_t=1}^{-1}(\hat{F}_{Y_{0t-2}|D_t=1}(z)))))$$

$$-\hat{F}_{\Delta Y_{0t-1}|D_t=1}(F_{\Delta Y_{0t-1}|D_t=1}^{-1}(F_{\Delta Y_{0t}|D_t=0}(y-F_{Y_{0t-1}|D_t=1}^{-1}(\hat{F}_{Y_{0t-2}|D_t=1}(z)))))$$

$$-\Big(F_{\Delta Y_{0t-1}|D_t=1}(F_{\Delta Y_{0t-1}|D_t=1}^{-1}(F_{\Delta Y_{0t}|D_t=0}(y-\hat{F}_{Y_{0t-1}|D_t=1}^{-1}(\hat{F}_{Y_{0t-2}|D_t=1}(z)))))$$

$$
\left. - F_{\Delta Y_{0t-1}|D_t=1}(F^{-1}_{\Delta Y_{0t-1}|D_t=1}(F_{\Delta Y_{0t}|D_t=0}(y - F^{-1}_{Y_{0t-1}|D_t=1}(\hat{F}_{Y_{0t-2}|D_t=1}(z))))) \right) \right|
$$

$$
(58)
$$

$$
+ \sup_z \sqrt{n} \left| F_{\Delta Y_{0t}|D_t=0}(y - \hat{F}^{-1}_{Y_{0t-1}|D_t=1}(\hat{F}_{Y_{0t-2}|D_t=1}(z))) \right.
$$

$$
- F_{\Delta Y_{0t}|D_t=0}(y - F^{-1}_{Y_{0t-1}|D_t=1}(\hat{F}_{Y_{0t-2}|D_t=1}(z)))
$$

$$
- \left(F_{\Delta Y_{0t}|D_t=0}(y - \hat{F}^{-1}_{Y_{0t-1}|D_t=1}(F_{Y_{0t-2}|D_t=1}(z))) \right.
$$

$$
\left. \left. - F_{\Delta Y_{0t}|D_t=0}(y - F^{-1}_{Y_{0t-1}|D_t=1}(F_{Y_{0t-2}|D_t=1}(z))) \right) \right|
$$

$$
(59)
$$

$$
+ \sup_z \sqrt{n} \left| F_{\Delta Y_{0t}|D_t=0}(y - \hat{F}^{-1}_{Y_{0t-1}|D_t=1}(F_{Y_{0t-2}|D_t=1}(z))) \right.
$$

$$
- F_{\Delta Y_{0t}|D_t=0}(y - F^{-1}_{Y_{0t-1}|D_t=1}(F_{Y_{0t-2}|D_t=1}(z)))
$$

$$
- \frac{f_{\Delta Y_{0t}|D_t=0}(y - F^{-1}_{Y_{0t-1}|D_t=1}(F_{Y_{0t-2}|D_t=1}(z)))}{f_{Y_{0t-1}|D_t=1}(F_{Y_{0t-2}|D_t=1}(z))}
$$

$$
\times \left(\hat{F}_{Y_{0t-1}|D_t=1}(F^{-1}_{Y_{0t-1}|D_t=1}(F_{Y_{0t-2}|D_t=1}(z))) \right.
$$

$$
\left. \left. - F_{Y_{0t-1}|D_t=1}(F^{-1}_{Y_{0t-1}|D_t=1}(F_{Y_{0t-2}|D_t=1}(z))) \right) \right|
$$

$$
(60)
$$

$$
\xrightarrow{p} 0
$$

Equation 60 converges to 0 by Lemma 13. Equation 52 converges to 0 by Lemma 16. For Equation 59, by a Taylor Expansion,

$$
\sqrt{n} \left(F_{\Delta Y_{0t}|D_t=0}(y - \hat{F}^{-1}_{Y_{0t-1}|D_t=1}(F_{Y_{0t-2}|D_t=1}(z))) - F_{\Delta Y_{0t}|D_t=0}(y - F^{-1}_{Y_{0t-1}|D_t=1}(F_{Y_{0t-2}|D_t=1}(z))) \right)
$$

$$
= f_{\Delta Y_{0t}|D_t=0}(y - F^{-1}_{Y_{0t-1}|D_t=1}(F_{Y_{0t-2}|D_t=1}(z))) \sqrt{n} \left(\hat{F}^{-1}_{Y_{0t-1}|D_t=1}(F_{Y_{0t-2}|D_t=1}(z)) \right.
$$

$$
\left. - F^{-1}_{Y_{0t-1}|D_t=1}(F_{Y_{0t-2}|D_t=1}(z)) \right) + o_p(1)
$$

$$
(61)
$$

The result is then obtained by using Lemma 8 on the term

$$
\sqrt{n} \left(\hat{F}^{-1}_{Y_{0t-1}|D_t=1}(F_{Y_{0t-2}|D_t=1}(z)) - F^{-1}_{Y_{0t-1}|D_t=1}(F_{Y_{0t-2}|D_t=1}(z)) \right)
$$

in Equation 61.

Equation 36 Recall that:

$$
\hat{\mu}_4 = \frac{1}{n_T^2} \sum_{i \in \mathcal{T}} \sum_{j \in \mathcal{T}} \frac{f_{\Delta Y_{0t}|D_t=0}(y - F^{-1}_{Y_{0t-1}|D_t=1}(F_{Y_{0t-2}|D_t=1}(Y_{0t-2},i)))}{f_{Y_{0t-1}|D_t=1}(F^{-1}_{Y_{0t-1}|D_t=1}(F_{Y_{0t-2}|D_t=1}(Y_{0t-2,i})))}
$$

$$
\times \left(\mathbb{1}\{Y_{0t-2}, j \le Y_{0t-2,i}\} - F_{Y_{0t-2}|D_t=1}(Y_{0t-2,i}) \right]
$$

Then,

$$\sqrt{n}\left(\frac{1}{n_T}\sum_{i\in\mathcal{T}}\left[\mathbb{1}\{F_{\Delta Y_{0t}|D_t=0}^{-1}(F_{\Delta Y_{0t-1}|D_t=1}(\Delta Y_{0t-1,i}))\le y-F_{Y_{0t-1}|D_t=1}^{-1}(\hat{F}_{Y_{0t-2}|D_t=1}(Y_{0t-2,i}))\}\right]\right.$$

$$\left.-\frac{1}{n_T}\sum_{i\in\mathcal{T}}\left[\mathbb{1}\{F_{\Delta Y_{0t}|D_t=0}^{-1}(F_{\Delta Y_{0t-1}|D_t=1}(\Delta Y_{0t-1,i}))\le y-F_{Y_{0t-1}|D_t=1}^{-1}(F_{Y_{0t-2}|D_t=1}(Y_{0t-2,i}))\}\right]-\hat{\mu}_4\right)\tag{62}$$

$$\le\sup_z\sqrt{n}\left|\hat{F}_{\Delta Y_{0t-1}|D_t=1}(F_{\Delta Y_{0t-1}|D_t=1}^{-1}(F_{\Delta Y_{0t}|D_t=1}(y-F_{Y_{0t-1}|D_t=1}^{-1}(\hat{F}_{Y_{0t-2}|D_t=1}(z)))))\right.$$

$$-\hat{F}_{\Delta Y_{0t-1}|D_t=1}(F_{\Delta Y_{0t-1}|D_t=1}^{-1}(F_{\Delta Y_{0t}|D_t=1}(y-F_{Y_{0t-1}|D_t=1}^{-1}(F_{Y_{0t-2}|D_t=1}(z)))))$$

$$-\frac{f_{\Delta Y_{0t}|D_t=0}(y-F_{Y_{0t-1}|D_t=1}^{-1}(F_{Y_{0t-2}|D_t=1}(z)))}{f_{Y_{0t-1}|D_t=1}(F_{Y_{0t-1}|D_t=1}^{-1}(F_{Y_{0t-2}|D_t=1}(z)))}$$

$$\left.\times\left(\hat{F}_{Y_{0t-2}|D_t=1}(z)-F_{Y_{0t-2}|D_t=1}(z)]\right)\right|\tag{63}$$

$$\le\sup_z\sqrt{n}\left|\hat{F}_{\Delta Y_{0t-1}|D_t=1}(F_{\Delta Y_{0t-1}|D_t=1}^{-1}(F_{\Delta Y_{0t}|D_t=1}(y-F_{Y_{0t-1}|D_t=1}^{-1}(\hat{F}_{Y_{0t-2}|D_t=1}(z)))))\right.$$

$$-\hat{F}_{\Delta Y_{0t-1}|D_t=1}(F_{\Delta Y_{0t-1}|D_t=1}^{-1}(F_{\Delta Y_{0t}|D_t=1}(y-F_{Y_{0t-1}|D_t=1}^{-1}(F_{Y_{0t-2}|D_t=1}(z)))))$$

$$-\left(F_{\Delta Y_{0t-1}|D_t=1}(F_{\Delta Y_{0t-1}|D_t=1}^{-1}(F_{\Delta Y_{0t}|D_t=1}(y-F_{Y_{0t-1}|D_t=1}^{-1}(\hat{F}_{Y_{0t-2}|D_t=1}(z)))))\right.$$

$$\left.\left.-F_{\Delta Y_{0t-1}|D_t=1}(F_{\Delta Y_{0t-1}|D_t=1}^{-1}(F_{\Delta Y_{0t}|D_t=1}(y-F_{Y_{0t-1}|D_t=1}^{-1}(F_{Y_{0t-2}|D_t=1}(z)))))\right)\right|\tag{64}$$

$$+\sup_z\sqrt{n}\left|F_{\Delta Y_{0t}|D_t=1}(y-F_{Y_{0t-1}|D_t=1}^{-1}(\hat{F}_{Y_{0t-2}|D_t=1}(z)))\right.$$

$$-F_{\Delta Y_{0t}|D_t=1}(y-F_{Y_{0t-1}|D_t=1}^{-1}(F_{Y_{0t-2}|D_t=1}(z)))$$

$$-\frac{f_{\Delta Y_{0t}|D_t=0}(y-F_{Y_{0t-1}|D_t=1}^{-1}(F_{Y_{0t-2}|D_t=1}(z)))}{f_{Y_{0t-1}|D_t=1}(F_{Y_{0t-1}|D_t=1}^{-1}(F_{Y_{0t-2}|D_t=1}(z)))}$$

$$\left.\times\left(\hat{F}_{Y_{0t-2}|D_t=1}(z)-F_{Y_{0t-2}|D_t=1}(z)]\right)\right|\tag{65}$$

$$\xrightarrow{p}0$$

Equation 64 converges to 0 by Lemma 13. For Equation 65, notice that by a Taylor expansion,

$$\sqrt{n}\left(F_{\Delta Y_{0t}|D_t=1}(y-F_{Y_{0t-1}|D_t=1}^{-1}(\hat{F}_{Y_{0t-2}|D_t=1}(z)))-F_{\Delta Y_{0t}|D_t=1}(y-F_{Y_{0t-1}|D_t=1}^{-1}(F_{Y_{0t-2}|D_t=1}(z)))\right)$$

$$=f_{\Delta Y_{0t}|D_t=1}(y-F_{Y_{0t-1}|D_t=1}^{-1}(\hat{F}_{Y_{0t-2}|D_t=1}(z)))$$

$$\times\sqrt{n}\left(F_{Y_{0t-1}|D_t=1}^{-1}(\hat{F}_{Y_{0t-2}|D_t=1}(z))-F_{Y_{0t-1}|D_t=1}^{-1}(F_{Y_{0t-2}|D_t=1}(z))\right)+o_p(1)\tag{66}$$

Then, by a second Taylor expansion,

$$\sqrt{n}\left(F_{Y_{0t-1}|D_t=1}^{-1}(\hat{F}_{Y_{0t-2}|D_t=1}(z))-F_{Y_{0t-1}|D_t=1}^{-1}(F_{Y_{0t-2}|D_t=1}(z))\right)$$

$$= \frac{1}{f_{Y_{0t-1}|D_t=1}(F^{-1}_{Y_{0t-1}|D_t=1}(F_{Y_{0t-2}|D_t=1}(z)))} \sqrt{n} \left(\hat{F}_{Y_{0t-2}|D_t=1}(z) - F_{Y_{0t-2}|D_t=1}(z) \right) + o_p(1)$$

$$(67)$$

and combining Equation 66 and Equation 67 implies the result.

Equation 44 Since

$$\hat{\mu}_5 = \left(\frac{1}{n_T} \sum_{i \in \mathcal{T}} \left[\mathbb{1}\{F^{-1}_{\Delta Y_{0t}|D_t=0}(F_{\Delta Y_{0t-1}|D_t=1}(\Delta Y_{0t-1,i})) \leq y - F^{-1}_{Y_{0t-1}|D_t=1}(F_{Y_{0t-2}|D_t=1}(Y_{0t-2,i}))\} \right] \right.$$

$$\left. - \mathrm{E}\left[\mathbb{1}\{F^{-1}_{\Delta Y_{0t}|D_t=0}(F_{\Delta Y_{0t-1}|D_t=1}(\Delta Y_{0t-1,i})) \leq y - F^{-1}_{Y_{0t-1}|D_t=1}(F_{Y_{0t-2}|D_t=1}(Y_{0t-2,i}))\}|D_t = 1 \right] \right)$$

Equation 44 is equal to 0.

$$\square$$

Based on the result of Lemma 19, we need only consider the asymptotic distribution of $\sqrt{n}(\hat{\mu}_1 + \hat{\mu}_2 + \hat{\mu}_3 + \hat{\mu}_4 + \hat{\mu}_5)$. Without needing adjustment, the Central Limit Theorem can easily be applied to $\hat{\mu}_5$. $\hat{\mu}_1$, $\hat{\mu}_2$, $\hat{\mu}_3$, and $\hat{\mu}_4$ are V-statistics. It is helpful to re-express each of these in an asymptotically equivalent form using the results of Lemma 17 and Lemma 18.

Lemma 20. *Asymptotic Representations of $\hat{\mu}_1$, $\hat{\mu}_2$, $\hat{\mu}_3$, and $\hat{\mu}_4$.*

Here we use Lemma 17 and Lemma 18 to write the V-statistics $\hat{\mu}_1$, $\hat{\mu}_2$, $\hat{\mu}_3$, and $\hat{\mu}_4$ in forms that the Central Limit Theorem can easily be applied to. Let $\mu_{j1}(x) = \mathrm{E}[\mu_j(x, Z)]$ and $\mu_{j2}(z) = \mathrm{E}[\mu_j(X, z)]$.[12] Then,

$$\hat{\mu}_1 = \frac{1}{n_C} \sum_{i \in \mathcal{C}} \mu_{12}(\Delta Y_{t,i}) + o_p(1)$$

$$\hat{\mu}_2 = \frac{1}{n_T} \sum_{i \in \mathcal{T}} \mu_{22}(Y_{t-2,i}) + o_p(1)$$

$$\hat{\mu}_3 = \frac{1}{n_T} \sum_{i \in \mathcal{T}} \mu_{31}(Y_{t-1,i}) + o_p(1)$$

$$\hat{\mu}_4 = \frac{1}{n_T} \sum_{i \in \mathcal{T}} \mu_{42}(Y_{t-2,i}) + o_p(1)$$

Proof. We show that the result holds for $\hat{\mu}_1$. The derivations of the result for $\hat{\mu}_2$, $\hat{\mu}_3$, and $\hat{\mu}_4$ proceed

---

[12]It should be noted that each of the $\mu_{jk}(\cdot)$ also depends on the value of y for which $P(Y_{0t} \leq y|D_t = 1)$ is being estimated. We suppress this notation here though.

similarly and are omitted.

$$\hat{\mu}_1 = \frac{1}{n_T n_C} \sum_{i \in \mathcal{T}} \sum_{j \in \mathcal{C}} \mu_1(Y_{0t-2,i}, \Delta Y_{0t,j}) \tag{68}$$

$$= \frac{1}{n_T} \sum_{i \in \mathcal{T}} \mu_{11}(Y_{0t-2,i}) + \frac{1}{n_C} \sum_{i \in \mathcal{C}} \mu_{12}(\Delta Y_{0t,i}) - \mathrm{E}[\mu_1(Y_{0t-2}, \Delta Y_{0t})] + o_p(1) \tag{69}$$

$$= \frac{1}{n_C} \sum_{i \in \mathcal{C}} \mu_{12}(\Delta Y_{0t,i}) + o_p(1) \tag{70}$$

Equation 69 uses Lemma 17. It is easy to show that $\mu_{11}(x) = 0$ and $\mathrm{E}[\mu_1(Y_{0t-2}, \Delta Y_{0t})] = 0$. □

Proof of Theorem 4. Let $\mu_j(\cdot; y)$ be the $\mu_j(\cdot)$ used in the previous lemmas with the dependence on the value of y in $\mathrm{P}(Y_{0t} < y|D_t = 1)$ explicit, and likewise for $\mu_{jk}(\cdot; y)$.

As a first step, notice that

$$\sqrt{n} \left(\hat{F}_{Y1t|D_t=1}^{-1}(\tau) - F_{Y1t|D_t=1}^{-1}(\tau) \right)$$

$$= \frac{1}{f_{Y_{1t}|D_t=1}(F_{Y_{1t}|D_t=1}^{-1}(\tau))} \sqrt{n} \left(\frac{1}{n_T} \sum_{i \in \mathcal{T}} \mathbb{1}\{Y_{1t} \le F_{Y_{1t}|D_t=1}^{-1}(\tau)\} - \tau \right) \tag{71}$$

$$\equiv \frac{1}{f_{Y_{1t}|D_t=1}(F_{Y_{1t}|D_t=1}^{-1}(\tau))} \sqrt{n} \left(\frac{1}{n_T} \sum_{i \in \mathcal{T}} \psi(Y_{1t,i}; F_{Y_{1t}|D_t=1}^{-1}(\tau)) \right) \tag{72}$$

Second, based on Lemma 19 and Lemma 20[13]

$$\sqrt{n} \left(\hat{F}_{Y0t|D_t=1}^{-1}(\tau) - F_{Y0t|D_t=1}^{-1}(\tau) \right)$$

$$= \frac{1}{f_{Y_{0t}|D_t=1}(F_{Y_{0t}|D_t=1}^{-1}(\tau))} \sqrt{n} \left(\hat{F}_{Y_{0t}|D_t=1}(F_{Y_{0t}|D_t=1}^{-1}(\tau)) - F_{Y_{0t}|D_t=1}(F_{Y_{0t}|D_t=1}^{-1}(\tau)) \right) + o_p(1)$$

$$= \frac{1}{f_{Y_{0t}|D_t=1}(F_{Y_{0t}|D_t=1}^{-1}(\tau))} \sqrt{n} \left(\frac{1}{n_C} \sum_{i \in \mathcal{C}} \mu_{12}(\Delta Y_{t,i}; F_{Y_{0t}|D_t=1}^{-1}(\tau)) \right.$$

$$+ \frac{1}{n_T} \sum_{i \in \mathcal{T}} \mu_{22}(\Delta Y_{t-1,i}; F_{Y_{0t}|D_t=1}^{-1}(\tau)) + \frac{1}{n_T} \sum_{i \in \mathcal{T}} \mu_{32}(Y_{t-1,i}; F_{Y_{0t}|D_t=1}^{-1}(\tau))$$

$$+ \frac{1}{n_T} \sum_{i \in \mathcal{T}} \mu_{42}(Y_{t-2,i}; F_{Y_{0t}|D_t=1}^{-1}(\tau)) + \left. \frac{1}{n_T} \sum_{i \in \mathcal{T}} \mu_5(\Delta Y_{t-1,i}, Y_{0t-2,i}; F_{Y_{0t}|D_t=1}^{-1}(\tau)) \right) + o_p(1) \tag{73}$$

---

[13]It should be noted that $f_{Y_{0t}|D_t=1}(y)$ and $F_{Y_{0t}|D_t=1}^{-1}(\tau)$ are identified because $F_{Y_{0t}|D_t=1}(y)$ is identified.

where, as defined in the text,

$$\mu_{12}(z;y) = \mathrm{E}\left[\mathbb{1}\{z \le (y - \mathrm{F}^{-1}_{Y_{0t-1}|D_t=1}(\mathrm{F}_{Y_{0t-2}|D_t=1}(Y_{0t-2})))\}\right.$$
$$\left. - \mathrm{F}_{\Delta Y_{0t}|D_t=0}(y - \mathrm{F}^{-1}_{Y_{0t-1}|D_t=1}(\mathrm{F}_{Y_{0t-2}|D_t=1}(Y_{0t-2})))|D_t = 1\right]$$

$$\mu_{22}(z;y) = \mathrm{E}\left[\mathbb{1}\{z \le \mathrm{F}^{-1}_{\Delta Y_{0t-1}|D_t=1}(\mathrm{F}_{\Delta Y_{0t}|D_t=0}(y - \mathrm{F}^{-1}_{Y_{0t-1}|D_t=1}(\mathrm{F}_{Y_{0t-2}|D_t=1}(Y_{0t-2}))))\}\right.$$
$$\left. - \mathrm{F}_{\Delta Y_{0t}|D_t=0}(y - \mathrm{F}^{-1}_{Y_{0t-1}|D_t=1}(\mathrm{F}_{Y_{0t-2}|D_t=1}(Y_{0t-2})))|D_t = 1\right]$$

$$\mu_{32}(z;y) = \mathrm{E}\left[\frac{f_{\Delta Y_{0t}|D_t=0}(y - \mathrm{F}^{-1}_{Y_{0t-1}|D_t=1}(\mathrm{F}_{Y_{0t-2}|D_t=1}(Y_{0t-2})))}{f_{Y_{0t-1}|D_t=1}(\mathrm{F}_{Y_{0t-2}|D_t=1}(Y_{0t-2}))} \times \left(\mathbb{1}\{z \le \mathrm{F}^{-1}_{Y_{0t-1}|D_t=1}(\mathrm{F}_{Y_{0t-2}|D_t=1}(Y_{0t-2}))\}\right.\right.$$
$$\left.\left. - \mathrm{F}_{Y_{0t-1}|D_t=1}(\mathrm{F}^{-1}_{Y_{0t-1}|D_t=1}(\mathrm{F}_{Y_{0t-2}|D_t=1}(Y_{0t-2})))\right)|D_t = 1\right]$$

$$\mu_{42}(z;y) = \mathrm{E}\left[\frac{f_{\Delta Y_{0t}|D_t=0}(y - \mathrm{F}^{-1}_{Y_{0t-1}|D_t=1}(\mathrm{F}_{Y_{0t-2}|D_t=1}(Y_{0t-2})))}{f_{Y_{0t-1}|D_t=1}(\mathrm{F}_{Y_{0t-2}|D_t=1}(Y_{0t-2}))} \times \left(\mathbb{1}\{z \le Y_{0t-2}\} - \mathrm{F}_{Y_{0t-2}|D_t=1}(Y_{0t-2})\right)|D_t = 1\right]$$

$$\mu_5(z_1,z_2;y) = \mathbb{1}\{\mathrm{F}^{-1}_{\Delta Y_{0t}|D_t=0}(\mathrm{F}_{\Delta Y_{0t-1}|D_t=1}(z_1)) \le y - \mathrm{F}^{-1}_{Y_{0t-1}|D_t=1}(\mathrm{F}_{Y_{0t-2}|D_t=1}(z_2))\}$$
$$- \mathrm{E}\left[\mathbb{1}\{\mathrm{F}^{-1}_{\Delta Y_{0t}|D_t=0}(\mathrm{F}_{\Delta Y_{0t-1}|D_t=1}(z_1)) \le y - \mathrm{F}^{-1}_{Y_{0t-1}|D_t=1}(\mathrm{F}_{Y_{0t-2}|D_t=1}(z_2))\}|D_t = 1\right]$$

and

$$f_{Y_{0t}|D_t=1}(y) = \int_{\mathcal{Y}_{0t-2}|D_t=1} f_{\Delta Y_{0t-1},Y_{0t-2}|D_t=1}(\mathrm{F}^{-1}_{\Delta Y_{0t-1}|D_t=1}(\mathrm{F}_{\Delta Y_{0t}|D_t=0}(y - \mathrm{F}^{-1}_{Y_{0t-1}|D_t=1}(\mathrm{F}_{Y_{0t-2}|D_t=1}(v)))), v)$$
$$\times \frac{f_{\Delta Y_{0t}|D_t=0}(y - \mathrm{F}^{-1}_{Y_{0t-1}|D_t=1}(\mathrm{F}_{Y_{0t-2}|D_t=1}(v)))}{f_{\Delta Y_{0t-1}|D_t=1}(\mathrm{F}^{-1}_{\Delta Y_{0t-1}|D_t=1}(\mathrm{F}_{\Delta Y_{0t}|D_t=0}(y - \mathrm{F}^{-1}_{Y_{0t-1}|D_t=1}(\mathrm{F}_{Y_{0t-2}|D_t=1}(Y_{0t-2})))))} \, dv$$

Since

$$\sqrt{n}\left(\hat{\mathrm{QTET}}(\tau) - \mathrm{QTET}(\tau)\right)$$
$$= \sqrt{n}\left(\hat{F}^{-1}_{Y_{1t}|D_t=1}(\tau) - F^{-1}_{Y_{1t}|D_t=1}(\tau)\right) - \sqrt{n}\left(\hat{F}^{-1}_{Y_{0t}|D_t=1}(\tau) - F^{-1}_{Y_{0t}|D_t=1}(\tau)\right)$$
$$= \sqrt{n}\left\{\frac{1}{f_{Y_{1t}|D_t=1}(F^{-1}_{Y_{1t}|D_t=1}(\tau))}\frac{1}{p}\frac{1}{n}\sum_{i=1}^{n} D_t\psi(Y_{1t,i}; F^{-1}_{Y_{1t}|D_t=1}(\tau))\right.$$
$$+ \frac{1}{f_{Y_{0t}|D_t=1}(F^{-1}_{Y_{0t}|D_t=1}(\tau))}\frac{1}{1-p}\frac{1}{n}\sum_{i=1}^{n}(1 - D_t)\mu_{12}(\Delta Y_{t,i}; F^{-1}_{Y_{0t}|D_t=1}(\tau))$$
$$+ \frac{1}{f_{Y_{0t}|D_t=1}(F^{-1}_{Y_{0t}|D_t=1}(\tau))}\frac{1}{p}\frac{1}{n}\sum_{i=1}^{n} D_t\left(\mu_{22}(\Delta Y_{t-1,i}; F^{-1}_{Y_{0t}|D_t=1}(\tau))\right.$$
$$+ \mu_{32}(Y_{t-1,i}; F^{-1}_{Y_{0t}|D_t=1}(\tau)) + \mu_{42}(Y_{t-2,i}; F^{-1}_{Y_{0t}|D_t=1}(\tau))$$
$$\left.\left.+ \mu_5(\Delta Y_{t-1,i}, Y_{0t-2,i}; F^{-1}_{Y_{0t}|D_t=1}(\tau))\right)\right\}$$

the result then follows from an application of the Central Limit Theorem. $\qquad\square$

Asymptotic Normality of propensity score reweighted estimator This section shows that the estimate of the QTET is still $\sqrt{n}$-asymptotically normal when the Distributional Difference in Differences assumption is made conditional on covariates. Under this variation, the only distribution that changes is $F_{\Delta Y_{0t}|D_t=1}(\Delta y)$ which is now given by $\mathrm{E}\left[\frac{1-D_t}{1-p(X)}\frac{p(X)}{P(D_t=1)}\mathbb{1}\{\Delta Y_t \le \Delta y\}\right]$ instead of replaced directly by the distribution of the change in untreated outcomes for the untreated group. We provide an asymptotically linear representation of this distribution which can easily be combined with the previous results to show asymptotic normality.

We consider two cases that are likely to be most useful to empirical researchers: (i) when the propensity score is known up to a finite number of parameters, and (ii) when the propensity score is estimated nonparametrically using a series logit estimator as in Hirano, Imbens, and Ridder (2003). We also have results (available upon request) that provides the conditions and proof of asymptotic normality when the propensity score is semiparametrically using the method of Klein and Spady (1993).

Proof of Theorem 6. At a high level, almost all of the proof of Theorem 4 carries over to Theorem 6). Only Equation 33 and $\hat{\mu}_1$ need to be changed. As a first step, we find an asymptotically linear representation of $\sqrt{n}(\hat{\mathrm{F}}_{\Delta Y_{0t}|D_t=1}(\Delta) - \mathrm{F}_{\Delta Y_{0t}|D_t=1}(\Delta))$. Then, we show how this result can be combined with previous results to show asymptotic normality of the estimate of the QTET. When the propensity score is estimated nonparametrically,

$$\sqrt{n}(\hat{\mathrm{F}}_{\Delta Y_{0t}|D_t=1}(\Delta y) - \mathrm{F}_{\Delta Y_{0t}|D_t=1}(\Delta y))$$
$$= \sqrt{n}\left(\frac{1}{n_C}\frac{(1-p)}{p}\sum_{i \in \mathcal{C}}\frac{p(X_i)}{(1-p(X_i))}\mathbb{1}\{\Delta Y_{ti} \le \Delta y\} - \mathrm{E}\left[\frac{(1-p)}{p}\frac{p(X)}{(1-p(X))}\mathbb{1}\{\Delta Y_t \le \Delta y\}|D_t=0\right]\right)$$
$$+ \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n}\frac{\mathrm{E}[\mathbb{1}\{\Delta Y_{it} \le \Delta y|X=X_i, D_t=0\}]}{p(1-p(X_i))}(D_{ti}-p(X_i))\right) + o_p(1)$$
$$\equiv \sqrt{n}\left(\frac{1}{n_C}\frac{(1-p)}{p}\sum_{i \in \mathcal{C}}\Psi_{N1}(Z_i;\Delta y) + \frac{1}{n}\sum_{i=1}^{n}\Psi_{N2}(Z_i;\Delta y)\right)$$

which follows using the results in Hirano, Imbens, and Ridder (2003) with $\mathbb{1}\{\Delta Y_t \le \Delta\}$ replacing Y_t in their model. The first line is the variance that would obtain if $p(x)$ were known. The second line gives the additional variance that comes from estimating $p(x)$.

When the propensity score is estimated parametrically,

$$\sqrt{n}(\hat{\mathrm{F}}_{\Delta Y_{0t}|D_t=1}(\Delta y) - \mathrm{F}_{\Delta Y_{0t}|D_t=1}(\Delta y))$$
$$= \sqrt{n}\left(\frac{1}{n_C}\frac{(1-p)}{p}\sum_{i \in \mathcal{C}}\frac{G(X_i^\top \zeta_0)}{(1-G(X_i^\top \zeta_0))}\mathbb{1}\{\Delta Y_{ti} \le \Delta y\} - \mathrm{E}\left[\frac{(1-p)}{p}\frac{G(X^\top \zeta_0}{(1-G(X^\top \zeta_0)}\mathbb{1}\{\Delta Y_t \le \Delta y\}|D_t=0\right.\right.$$
$$+ \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n}\mathrm{E}_{Y_{0t-2}|D_t=1}\left\{\mathrm{E}\left[\frac{1}{1-G(X^\top \zeta_0)}\left(1+\frac{G(X^\top \zeta_0)}{1-G(X^\top \zeta_0)}\right)\right.\right.$$

53

$$\times \mathrm{E}\left[\left(\frac{D - G(X^\top \zeta_0)}{G(X^\top \zeta_0)(1 - G(X^\top \zeta_0))}\right)^2 g(X^\top \zeta_0)^2 X X^\top\right]$$

$$\times \left. \frac{D_{it} - G(X_i^\top \zeta_0)}{G(X_i, \zeta_0)(1 - G(X_i^\top \zeta_0))} g(X_i^\top \zeta_0)\right\}\right)$$

$$\equiv \sqrt{n}\left(\frac{1}{n_C}\frac{(1-p)}{p}\sum_{i \in \mathcal{C}} \Psi_{P1}(Z_i; \Delta y) + \frac{1}{n}\sum_{i=1}^n \Psi_{P2}(Z_i; \Delta y)\right)$$

Let[14]

$$\hat{\mu}_{1N} = \sqrt{n}\frac{1}{n_C n_T}\frac{(1-p)}{p}\sum_{i \in \mathcal{C}}\sum_{j \in \mathcal{T}} \Psi_{N1}(Z_i; y - F^{-1}_{Y_{0t-1}|D_t=1}(F_{Y_{0t-2}|D_t=1}(Y_{0t-2,j})))$$

$$+ \sqrt{n}\frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n \frac{D_{tj}}{p}\Psi_{N2}(Z_i; y - F^{-1}_{Y_{0t-1}|D_t=1}(F_{Y_{0t-2}|D_t=1}(Y_{0t-2,j})))$$

where Ψ_{N1} and Ψ_{N2} are defined above.

Starting from Equation 33 except with $\hat{F}^{-1}_{\Delta Y_{0t}|D_t=0}(\cdot)$ replaced by the propensity score reweighted $\hat{F}^{-1}_{\Delta Y_{0t}|D_t=1}(\cdot)$,

$$\sqrt{n}\left(\frac{1}{n_T}\sum_{i \in \mathcal{T}}\left[\mathbb{1}\{\hat{F}^{-1}_{\Delta Y_{0t}|D_t=1}(\hat{F}_{\Delta Y_{0t-1}|D_t=1}(\Delta Y_{0t-1,i})) \le y - \hat{F}^{-1}_{Y_{0t-1}|D_t=1}(\hat{F}_{Y_{0t-2}|D_t=1}(Y_{0t-2,i}))\}\right]\right.$$

$$\left.- \frac{1}{n_T}\sum_{i \in \mathcal{T}}\left[\mathbb{1}\{F^{-1}_{\Delta Y_{0t}|D_t=1}(\hat{F}_{\Delta Y_{0t-1}|D_t=1}(\Delta Y_{0t-1,i})) \le y - \hat{F}^{-1}_{Y_{0t-1}|D_t=1}(\hat{F}_{Y_{0t-2}|D_t=1}(Y_{0t-2,i}))\}\right] - \hat{\mu}_{1N}\right)$$

$$\le \sup_z \sqrt{n}\left|\frac{1}{n_T}\sum_{i \in \mathcal{T}}\left[\mathbb{1}\{\Delta Y_{0t-1,i} \le \hat{F}^{-1}_{\Delta Y_{0t-1}|D_t=1}(\hat{F}_{\Delta Y_{0t}|D_t=1}(y - \hat{F}^{-1}_{Y_{0t-1}|D_t=1}(\hat{F}_{Y_{0t-2}|D_t=1}(z))))\}\right]\right.$$

$$- \frac{1}{n_T}\sum_{i \in \mathcal{T}}\left[\mathbb{1}\{\Delta Y_{0t-1,i} \le \hat{F}^{-1}_{\Delta Y_{0t-1}|D_t=1}(\mathrm{F}_{\Delta Y_{0t}|D_t=1}(y - \hat{F}^{-1}_{Y_{0t-1}|D_t=1}(\hat{F}_{Y_{0t-2}|D_t=1}(z))))\}\right]$$

$$- \frac{1}{n_C}\frac{(1-p)}{p}\sum_{i \in \mathcal{C}} \Psi_{N1}(Z_i; y - F^{-1}_{Y_{0t-1}|D_t=1}(F_{Y_{0t-2}|D_t=1}(z)))$$

$$\left.+ \frac{1}{n}\sum_{i=1}^n \Psi_{N2}(Z_i; y - F^{-1}_{Y_{0t-1}|D_t=1}(F_{Y_{0t-2}|D_t=1}(z)))\right| + o_p(1)$$

$$\le \sup_z \sqrt{n}\left|\hat{F}_{\Delta Y_{0t}|D_t=1}(y - \hat{F}^{-1}_{Y_{0t-1}|D_t=1}(\hat{F}_{Y_{0t-2}|D_t=1}(z)))\right.$$

$$- F_{\Delta Y_{0t}|D_t=1}(y - \hat{F}^{-1}_{Y_{0t-1}|D_t=1}(\hat{F}_{Y_{0t-2}|D_t=1}(z)))$$

$$- \left(\hat{F}_{\Delta Y_{0t}|D_t=1}(y - F^{-1}_{Y_{0t-1}|D_t=1}(F_{Y_{0t-2}|D_t=1}(z)))\right.$$

$$\left.\left.- F_{\Delta Y_{0t}|D_t=1}(y - F^{-1}_{Y_{0t-1}|D_t=1}(F_{Y_{0t-2}|D_t=1}(z)))\right)\right| + o_p(1)$$

---

[14]We show the remainder of the proof only for the nonparametric case; the argument for the parametric case is the same with $\Psi_{P1}(\cdot,\cdot)$ replacing $\Psi_{N1}(\cdot,\cdot)$ and $\Psi_{P2}(\cdot,\cdot)$ replacing $\Psi_{N2}(\cdot,\cdot)$.

which converges to 0 based on Lemma 13. $\qquad\square$

Finally, working with $\hat{\mu}_{1N}$, and using the result of Lemma 17 and Lemma 18, one can show that

$$\frac{1}{n_C n_T} \frac{(1-p)}{p} \sum_{i \in \mathcal{C}} \sum_{j \in \mathcal{T}} \Psi_{N1}(Z_i; y - F_{Y_{0t-1}|D_t=1}^{-1}(F_{Y_{0t-2}|D_t=1}(Y_{0t-2,j})))$$

$$= \frac{1}{n_C} \frac{(1-p)}{p} \sum_{i \in \mathcal{C}} E_{Y_{0t-2}|D_t=1} \left[\Psi_{N2}(Z_i; y - F_{Y_{0t-1}|D_t=1}^{-1}(F_{Y_{0t-2}|D_t=1}(Y_{0t-2}))) \right] + o_p(1)$$

$$\equiv \hat{\mu}_{1Ca} + o_p(1)$$

$$\frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \frac{D_{tj}}{p} \Psi_{P2}(Z_i; y - F_{Y_{0t-1}|D_t=1}^{-1}(F_{Y_{0t-2}|D_t=1}(Y_{0t-2,j})))$$

$$= \frac{1}{n} \sum_{i=1}^n E_{Y_{0t-2}|D_t=1} \left[\Psi_{P2}(Z_i; y - F_{Y_{0t-1}|D_t=1}^{-1}(F_{Y_{0t-2}|D_t=1}(Y_{0t-2}))) \right] + o_p(1)$$

$$\equiv \hat{\mu}_{1Cb} + o_p(1)$$

This implies that $\sqrt{n}(\hat{F}_{Y_{0t}|D_t=1}(y) - F_{Y_{0t}|D_t=1}(y)) = \sqrt{n} (\hat{\mu}_{1Ca} + \hat{\mu}_{1Cb} + \hat{\mu}_2 + \hat{\mu}_3 + \hat{\mu}_4 + \hat{\mu}_5) + o_p(1)$. And the result follows using the same ideas of the case with no covariates but with $\hat{\mu}_{1Ca} + \hat{\mu}_{1Cb}$ substituted for $\hat{\mu}_1$.

B   Tables

Table 1: Summary Statistics

	Treated		Randomized			Observational		
	mean	sd	mean	sd	nd	mean	sd	nd
RE 1978	6.35	7.87	4.55	5.48	0.19	21.55	15.56	− 0.87
RE 1975	1.53	3.22	1.27	3.10	0.06	19.06	13.60	− 1.25
RE 1974	2.10	4.89	2.11	5.69	0.00	19.43	13.41	− 1.21
Age	25.82	7.16	25.05	7.06	0.08	34.85	10.44	− 0.71
Education	10.35	2.01	10.09	1.61	0.10	12.12	3.08	− 0.48
Black	0.84	0.36	0.83	0.38	0.03	0.25	0.43	1.05
Hispanic	0.06	0.24	0.11	0.31	− 0.12	0.03	0.18	0.09
Married	0.19	0.39	0.15	0.36	0.07	0.87	0.34	− 1.30
No Degree	0.71	0.46	0.83	0.37	− 0.21	0.31	0.46	0.62
Unemployed in 1975	0.60	0.49	0.68	0.47	− 0.13	0.10	0.30	0.87
Unemployed in 1974	0.71	0.46	0.75	0.43	− 0.07	0.09	0.28	1.16

Notes: RE are real earnings in a given year in thousands of dollars. ND denotes the normalized difference between the Treated group and the Randomized group or Observational group, respectively.

Table 2: QTET Estimates for Job Training Program

	0.7	Diff	0.8	Diff	0.9	Diff	ATT	Diff
			PanelQTET Method					
PanelQTET SL	3.21*	1.40	5.80*	3.53*	7.25*	4.05*	2.96*	1.16
	(1.35)	(1.34)	(1.11)	(1.23)	(2.40)	(1.75)	(1.02)	(0.97)
PanelQTET Cov	1.46	−0.34	2.59*	0.32	2.45	−0.74	3.09*	1.29*
	(1.44)	(1.22)	(1.22)	(1.43)	(2.28)	(1.51)	(0.72)	(0.55)
PanelQTET UNEM	3.32*	1.51	5.80*	3.53*	7.92*	4.72*	3.23*	1.44
	(1.43)	(1.37)	(1.17)	(1.24)	(2.15)	(1.54)	(0.96)	(0.83)
PanelQTET No Cov	−0.77	−2.57*	0.58	−1.69	−0.25	−3.45*	2.33*	0.53
	(1.27)	(0.98)	(0.99)	(1.10)	(2.09)	(1.24)	(0.70)	(0.44)
			Conditional Independence Method					
CI SL	4.52*	2.71*	6.03*	3.76*	4.98	1.78	1.16	−0.63
	(1.47)	(1.19)	(1.92)	(1.84)	(4.00)	(3.25)	(1.13)	(1.04)
CI Cov	−5.13*	−6.93*	−6.97*	−9.25*	−10.54*	−13.74*	−4.70*	−6.50*
	(1.23)	(1.14)	(1.40)	(1.48)	(2.64)	(2.02)	(0.94)	(0.77)
CI UNEM	3.45*	1.64	5.14*	2.87	4.24	1.04	0.02	−1.77
	(1.40)	(1.22)	(1.54)	(1.53)	(3.22)	(2.48)	(1.16)	(0.99)
CI No Cov	−19.19*	−20.99*	−20.86*	−23.14*	−23.87*	−27.07*	−15.20*	−17.00*
	(0.89)	(0.75)	(0.92)	(1.08)	(1.92)	(1.12)	(0.69)	(0.49)
			Change in Changes					
CiC Cov	3.74*	1.94	4.32*	2.04	5.03*	1.84	3.84*	2.05*
	(0.88)	(1.01)	(1.02)	(1.23)	(1.54)	(1.76)	(0.81)	(0.53)
CiC UNEM	0.37	−1.44	1.84	−0.43	2.09	−1.10	1.92*	0.13
	(1.31)	(1.35)	(1.43)	(1.45)	(2.02)	(1.96)	(0.76)	(0.49)
CiC No Cov	8.16*	6.36*	9.83*	7.56*	10.07*	6.87*	5.08*	3.29*
	(0.80)	(0.60)	(1.04)	(1.08)	(2.57)	(1.97)	(0.69)	(0.40)
			Quantile D-i-D					
QDiD Cov	2.18*	0.37	2.85*	0.58	2.45	−0.75	2.48*	0.69
	(0.71)	(0.91)	(0.97)	(1.23)	(1.59)	(1.77)	(0.75)	(0.56)
QDiD UNEM	1.10	−0.70	2.66*	0.39	2.35	−0.84	2.40*	0.60
	(1.13)	(1.21)	(1.26)	(1.34)	(1.87)	(1.92)	(0.74)	(0.56)
QDiD No Cov	4.21*	2.41*	4.65*	2.38*	4.90*	1.70	1.68*	−0.11
	(0.97)	(0.87)	(1.09)	(1.04)	(2.05)	(1.31)	(0.79)	(0.61)
			Mean D-i-D					
MDiD Cov	3.09*	1.29	3.74*	1.47	4.80*	1.60	2.33*	0.53
	(0.67)	(0.85)	(0.94)	(1.20)	(1.46)	(1.66)	(0.70)	(0.44)
MDiD UNEM	2.41*	0.61	4.17*	1.90	4.85*	1.65	2.33*	0.53
	(1.14)	(1.21)	(1.22)	(1.30)	(1.78)	(1.79)	(0.70)	(0.44)
MDiD No Cov	4.47*	2.67*	5.58*	3.31*	6.65*	3.46*	2.33*	0.53
	(0.88)	(0.74)	(0.90)	(0.94)	(2.01)	(1.11)	(0.70)	(0.44)
Experimental	1.80		2.27*		3.20		1.79*	
	(0.93)		(1.13)		(2.04)		(0.69)	

Notes: This table provides estimates of the QTET for $\tau = c(0.7, 0.8, 0.9)$ using a variety of methods on the observational dataset. The reported estimates are in real terms and in 1000s of dollars. The columns labeled 'Diff' provide the difference between the estimated QTET and the QTET obtained from the experimental portion of the dataset. The columns identify the method (PanelQTET, CI, CiC, QDiD, or MDiD) and the set of covariates ((i) SL: Series Logit estimates of the propensity score (these specifications are slightly different as the CI method can condition on lags of real earnings while the PanelQTET does not include lags of real earnings as covariates; more details of method in text) (ii) COV: Age, Education, Black dummy, Hispanic dummy, Married dummy, and No HS Degree dummy; (iii) UNEM: all covariates in COV plus Unemployed in 1975 dummy and Unemployed in 1974 dummy (iv) NO COV: no covariates). The PanelQTET model and the CI model use propensity score re-weighting techniques based on the covariate set. The CiC, QDiD, and MDiD method "residualize" (as outlined in the text) the outcomes based on the covariate set; the estimates come from using the no covariate method on the "residualized" outcome. Standard errors are produced using 100 bootstrap iterations. The significance level is 5%.