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

Brantly Callaway†
Tong Li‡

First Version: April 2015
This Version: October 2016

Abstract

This paper considers identification and estimation of the Quantile Treatment Effect on the Treated (QTT) under a straightforward distributional extension of the most commonly invoked Mean Difference in Differences assumption used for identifying the Average Treatment Effect on the Treated (ATT). Identification of the QTT is more complicated than the ATT though because it depends on the unknown dependence between the change in untreated potential outcomes and the initial level of untreated potential outcomes for the treated group. To address this issue, we introduce a new Copula Stability Assumption that says that the missing dependence is constant over time. Under this assumption and when panel data is available, the missing dependence can be recovered, and the QTT is identified. Second, we allow for identification to hold only after conditioning on covariates and provide very simple estimators based on propensity score re-weighting for this case. We compare the performance of our method to existing methods for estimating QTTs using Lalonde (1986)'s job training dataset. Using this dataset, we find the performance of our method compares favorably to the performance of existing methods.

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

*We would like to thank Don Andrews, Stephane Bonhomme, Sergio Firpo, Antonio Galvao, Federico Gutierrez, John Ham, Magne Mogstad, Derek Neal, John Pepper, Peter Phillips, Pedro Sant'Anna, Azeem Shaikh, Youngki Shin, Steve Stern, Ed Vytlacil, Kaspar Wuthrich, and participants in seminars at Lancaster University, the National University of Singapore, the University of Chicago, the University of Iowa, the University of Sydney, the University of Virginia, UC San Diego, Vanderbilt University, Yale University, at the conference in honor of Takeshi Amemiya in Xiamen, China, June 2015, and at the 11th World Congress of the Econometric Society for their comments and suggestions. We are especially grateful to James Heckman for his insightful discussion and comments that have greatly improved the paper. Li acknowledges gratefully the hospitality and support of Becker Friedman Institute at the University of Chicago.

†Department of Economics, Temple University, 1301 Cecil B. Moore Avenue, Ritter Annex 867, Philadelphia, PA 19122. Email: brantly.callaway@temple.edu

‡Department of Economics, Vanderbilt University, VU Station B #351819, Nashville, TN 37235-1819. Phone: (615) 322-3582, Fax: (615) 343-8495, Email: tong.li@vanderbilt.edu

1 Introduction

Although most research using program evaluation techniques focuses on estimating the average effect of participating in a program or treatment, in some cases a researcher may be interested in understanding the distributional impacts of treatment participation. For example, for two labor market policies with the same mean impact, policymakers are likely to prefer a policy that tends to increase income in the lower tail of the income distribution to one that tends to increase income in the middle or upper tail of the income distribution. In contrast to the standard linear model, the treatment effects literature explicitly recognizes that the effect of treatment can be heterogeneous across different individuals (Heckman and Robb, 1985; Heckman, Smith, and Clements, 1997). Recently, many methods have been developed that identify distributional treatment effect parameters under common identifying assumptions such as selection on observables (Firpo, 2007), access to a an instrumental variable (Abadie, Angrist, and Imbens, 2002; Chernozhukov and Hansen, 2005; Carneiro and Lee, 2009; Frölich and Melly, 2013), or access to repeated observations over time (Athey and Imbens, 2006; Bonhomme and Sauder, 2011; Chernozhukov, Fernández-Val, Hahn, and Newey, 2013; Jun, Lee, and Shin, 2016). This paper focuses on identifying and estimating a particular distributional treatment effect parameter called the Quantile Treatment Effect on the Treated (QTT) using a Difference in Differences assumption for identification.

Empirical researchers commonly employ Difference in Differences assumptions to credibly identify the Average Treatment Effect on the Treated (ATT) (early examples include Card, 1990; Card and Krueger, 1994). Despite the prevalence of DID methods in applied work, there has been very little empirical work studying the distributional effects of a treatment with identification that exploits have access to repeated observations over time (Recent exceptions include Meyer, Viscusi, and Durbin, 1995; Finkelstein and McKnight, 2008; Pomeranz, 2015; Havnes and Mogstad, 2015).

The first contribution of the current paper is to provide identification and estimation results for the QTT under a straightforward extension of the most common Mean Difference in Differences Assumption (Heckman and Robb, 1985; Heckman, Ichimura, Smith, and Todd, 1998; Abadie, 2005). In particular, we strengthen the assumption of mean independence between (i) the change in untreated potential outcomes over time and (ii) whether or not an individual is treated to full independence. We call this assumption the Distributional Difference in Differences Assumption.

For empirical researchers, methods developed under the Distributional Difference in Differences Assumption are valuable precisely because the identifying assumptions are straightforward extensions of the Mean Difference in Differences assumptions that are frequently

employed in applied work. This means that almost all of the intuition for applying a Difference in Differences method for the ATT will carry over to identifying the QTT using our method.

Although applying a Mean Difference in Differences assumption leads straightforwardly to identification of the ATT, using the Distributional Difference in Differences Assumption to identify the QTT faces some additional challenges. The reason for the difference is that Mean Difference in Differences exploits the linearity of the expectation operator. In fact, with only two periods of data (which can be either repeated cross sections or panel) and under the same Distributional Difference in Differences assumption considered in the current paper, the QTT is known to be partially identified (Fan and Yu, 2012) without further assumptions. In practice, these bounds tend to be quite wide. Lack of point identification occurs because the dependence between (i) the distribution of the change in untreated outcomes for the treated group and (ii) the initial level of untreated potential outcomes for the treated group is unknown. For identifying the ATT, knowledge of this dependence is not required and point identification results can be obtained.

To move from partial identification back to point identification, we introduce a new assumption which we call the Copula Stability Assumption. This assumption says that the copula, which captures the unknown dependence mentioned above, does not change over time. To give an example, consider the case where the outcome of interest is earnings. The Copula Stability Assumption says that if we observe in the past that the largest earnings increases tended to go to those with the highest earnings, then, in the present (and in the absence of treatment), the largest earnings increase would have gone to those with the highest earnings. Importantly, this does not place any restrictions on the marginal distributions of outcomes over time allowing, for example, the outcomes to be non-stationary. There are two additional requirements for invoking this assumption relative to the Mean Difference in Differences Assumption: (i) access to panel data (repeated cross sections is not enough) and (ii) access to at least three periods of data (rather than at least two periods of data) where two of the periods must be pre-treatment periods and the third period is post-treatment. We show that the additional requirements that the Copula Stability Assumption places on the type of model that is consistent with the Distributional Difference in Differences Assumption are small.

The second contribution of the paper is to extend the results to the case where the identifying assumptions hold conditional on covariates. There are several reasons why this is an important contribution. First, we show that, for many models where an unconditional Mean Difference in Differences assumption holds, the Distributional Difference in Differences Assumption is likely to require conditioning on covariates. Second, conditional on covariates

versions of our assumptions can allow the path of untreated potential outcomes to depend on observed characteristics.

Having simple identification results when identification holds conditional on some covariates stands in contrast to the existing methods for estimating QTTs. The methods are either (i) unavailable or at least computationally challenging 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 QTT and is available under weaker assumptions, a researcher's primary interest in studying the QTT is likely to be in the shape of the QTT rather than the location of the QTT. In this regard, the parametric assumptions required by other methods to accommodate covariates may be restrictive because nonlinearities or misspecification of the parametric model could easily be confused with the shape of the QTT. 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.

Based on our identification results, estimation of the QTT is straightforward and computationally fast. Without covariates, estimating the QTT 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, but second step estimation is simple and fast. We show that our estimate of the QTT converges to a Gaussian process at the parametric rate $\sqrt{n}$ even when the propensity score is estimated nonparametrically. This result allows us to conduct uniform inference over a range of quantiles and can test, for example, whether the distribution of treated potential outcomes stochastically dominates the distribution of untreated potential outcomes.

We conclude the paper by comparing the performance of our method with alternative estimators of the QTT: the Quantile Difference in Differences model, the Change in Changes model, and a model based on selection on observables (Firpo, 2007) in an application to estimating the QTT of participating in a job training program using a well known dataset from 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 considers the main challenges for identification of the QTT while allowing for time-invariant unobserved heterogeneity. Section 4 provides our main identification result in the case where the Distributional Difference in Differences assumption holds with no covariates. Section 5 extends this result to the case with covariates and provides a propensity score re-weighting procedure to make estimation more feasible. Section 6 details our estimation strategy and the asymptotic properties of our estimation procedure. Section 7 compares our method to existing methods for estimating QTTs. Section 8 provides additional evidence on our Copula Stability Assumption. Section 9 contains the job training application. Section 10 concludes. Identification and estimation under a conditional Copula Stability Assumption is included in Appendix A. All the proofs are included in Appendix B.

2 Background

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.

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. Comparing the distribution of observed outcomes to a counterfactual distribution of untreated potential outcomes is a very important ingredient for evaluating the effect of a program or policy (Sen, 1997; Carneiro, Hansen, and Heckman, 2001) and provides more information than the average effect of the program alone. For example, a policy maker may be in favor of implementing a job training program that increases the lower tail of the distribution of earnings while decreasing the upper tail of the distribution of earnings even if the average effect of the program is zero.

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

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

An example is the 0.5-quantile – the median.[1] 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 0.95-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 on the Treated (QTT)[2] is defined as

$$\text{QTT}(\tau) = F^{-1}_{Y_{1t}|D_t=1}(\tau) - F^{-1}_{Y_{0t}|D_t=1}(\tau) \tag{2}$$

The QTT 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. Difference in Differences methods for the average effect of participating in a treatment also identify the Average Treatment Effect on the Treated, not the average treatment effect for the population at large.

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

[2] Quantile Treatment Effects were first studied by Doksum, (1974) and Lehmann, (1974)

3 Identification Challenges

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

Assumption 3.1 (Mean Difference in Differences)**.**

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

This is the "parallel 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 QTT, Assumption 3.1 needs to be strengthened because the QTT depends on the entire distribution of untreated outcomes for the treated group rather than only the mean of this distribution.

The next assumption strengthens Assumption 3.1 and this is the assumption maintained throughout the paper.

Distributional Difference in Differences Assumption.

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

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 QTT stands in contrast to the existing literature on identifying the QTT in similar models which generally require less familiar assumptions on the relationship between observed and unobserved outcomes.

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. In practice, these bound tend to be very wide – too wide to be useful in most applications.

4 Main Results: Identifying QTT in Difference in Differences Models

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 QTT. 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 – that is, the copula – between (i) the distribution of $(\Delta Y_{0t}|D_t = 1)$ (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 says that if, in the past, the largest increases in outcomes tend to go to those at the top of the distribution, then in the present, the largest increases in outcomes will tend to go to those who start out at the top of the distribution. It does not restrict what the distribution of the change in outcomes over time is nor does it restrict the distribution of outcomes in the previous period; instead, it restricts the dependence between these two marginal distributions. We discuss this assumption in more detail and show how it can be used to point identify the QTT 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 $F_{Y_{0t}|D_t=1}(y)$ (the distribution of interest) could be derived from it. The marginal distributions $F_{\Delta Y_{0t}|D_t=1}(\Delta)$ (through the Distributional Difference in Differences assumption) and $F_{Y_{0t-1}|D_t=1}(y')$ (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.[3] We use the idea that the dependence is the same between period t and period $(t-1)$. With this additional information, $\mathrm{F}_{\Delta Y_{0t}, Y_{t-1}|D_t=1}(\Delta, y')$ is identified and therefore the counterfactual distribution of untreated potential outcomes for the treated group, $\mathrm{F}_{Y_{0t}|D_t=1}(y)$ is identified.

The time invariance of the dependence between $F_{\Delta Y_{0t}|D_t=1}(\Delta)$ and $F_{Y_{0t-1}|D_t=1}(y)$ can

[3]For a continuous distribution, the copula representation is unique. Joe, (1997), Nelsen, (2007), and Joe, (2015) are useful references for more details on copulas.

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

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

where $C_{\Delta Y_{0t},Y_{0t-1}|D_t=1}(\cdot,\cdot)$ is a copula function.[4] Next, we state the second main assumption 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, such as independence or perfect positive dependence, 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 $(\Delta Y_{0t-1}|D_t = 1)$ 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.

[4] 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.

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

Assumption 4.1. *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 group in period $(t-2)$, respectively. We assume that $\Delta\mathcal{Y}_{t|D_t=0}$, $\Delta\mathcal{Y}_{t-1|D_t=1}$, $\mathcal{Y}_{t-1|D_t=1}$, and $\mathcal{Y}_{t-2|D_t=1}$ are compact. Also, 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 their support with densities that are bounded from above and bounded away from 0.*

Assumption 4.2. *The observed data $(Y_{dt,i}, Y_{t-1,i}, Y_{t-2,i}, X_i, D_{it})$ are independently and identically distributed.*

Assumption 4.1 says that outcomes are continuously distributed. Copulas are unique on the range of their marginal distributions; thus, continuously distributed outcomes guarantee that the copula is unique. However, for the CSA, one could weaken this assumption to $Range(\mathrm{F}_{\Delta Y_{0t}|D_t=1}) \subseteq Range(\mathrm{F}_{\Delta Y_{t-1}|D_t=1})$ and $Range(\mathrm{F}_{Y_{t-1}|D_t=1}) \subseteq Range(\mathrm{F}_{Y_{t-2}|D_t=1})$ and still obtain point identification. On the other hand, although neither our DDID Assumption nor the standard mean DID Assumption explicitly require continuously distributed outcomes, it should be noted that standard limited dependent variable models with unobserved heterogeneity would not generally satisfy either of these DID assumptions. Assumption 4.2 could potentially be relaxed in several ways. More periods of data could be available – our method requires at least three periods of data, but more periods could be incorporated in a straightforward way. Also, our setup could allow for some individuals to be treated in earlier periods than the last one though our results would continue to go through for the group of individuals that are first treated in the last period; considering the case where no one is treated before the last period is standard DID setup. Assumption 4.2 also says that other covariates X are time invariant. This assumption can be relaxed by focusing on the subset of individuals whose covariates do not change over time. Appendix A also discusses the possibility of including time varying covariates though they must enter our model is a more restrictive way than time invariant covariates. Essentially, the problem with time varying covariates is that that one cannot separate individuals changing ranks in the distribution of outcomes over time due to changes in covariates or due to unobservables. Finally, we assume that we observe treatment status for each individual; however, in many DID applications, treatments may be defined by location and individuals may move between treatment regimes over time (Lee and Kang, 2006) though we do not consider this complication.

Theorem 1. *Under the Distributional Difference in Differences Assumption, the Copula Stability Assumption, and Assumptions 4.1 and 4.2*

$$\begin{aligned}&\mathrm{F}_{Y_{0t}|D_t=1}(y)\\&\quad=\mathrm{E}\left[\mathbb{1}\{F^{-1}_{\Delta Y_t|D_t=0}(F_{\Delta Y_{t-1}|D_t=1}(\Delta Y_{t-1}))\leq y-F^{-1}_{Y_{t-1}|D_t=1}(F_{Y_{t-2}|D_t=1}(Y_{t-2}))\}|D_t=1\right]\end{aligned} \tag{3}$$

and

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

which is identified

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 $\mathrm{P}(Y_{0t}\leq y|D_t=1)=\mathrm{E}[\mathbb{1}\{\Delta Y_{0t}+Y_{0t-1}\leq y\}|D_t=1]$[5] 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 $\mathrm{F}^{-1}_{\Delta Y_{0t}|D_t=1}(\cdot)$ with $\mathrm{F}^{-1}_{\Delta Y_{0t}|D_t=0}(\cdot)$ which is just the quantiles of the distribution of the change in (observed) untreated outcomes for the untreated group.

5 Allowing for covariates

In our view, the key reason that there has been little use of distributional methods with panel data is that existing work has focused primarily on the case without conditioning on other covariates.[6] This section extends the previous results to the case where a Conditional DDID assumption holds.

Conditional Distributional Difference in Differences Assumption.

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

[5]Adding and subtracting Y_{0t-1} is also the first step for showing that the Mean Difference in Differences Assumption identifies $\mathrm{E}[Y_{0t}|D_t=1]$; the problem is much easier in the mean case though due to the linearity of expectations and no indicator function.

[6]Recent work such as Melly and Santangelo, (2015) and Callaway, Li, and Oka, (2016) has begun relaxing this restriction.

This assumption says that, 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. The next example shows that having the conditional DDID assumption may be important even in cases where an unconditional mean DID assumption holds and would identify the ATT

Example 1. *Consider the following model*

$$Y_{it} = q(U_{it}, D_{it}, X_i) + C_i$$

with (U_{it}, U_{it-1}), $(U_{it-1}, U_{it-2})|X, C, D \sim F_{U1,U2}$ *where* $F_{U1,U2}$ *is a bivariate distribution with uniform marginals,* C *is time invariant unobserved heterogeneity that may be correlated with observables, and* $q(\tau, d, x)$ *is strictly increasing in* τ *for all* $(d, x) \in \{0, 1\} \times \mathcal{X}$.

In this model,

- *The Unconditional Mean Difference in Differences Assumption holds*
- *The Unconditional Distributional Difference in Differences Assumption does not hold*
- *The Conditional Distributional Difference in Differences Assumption holds*
- *The Unconditional Copula Stability Assumption holds*

Example 1 is a Skorohod representation for panel quantile regression while allowing for serial correlation among U. This model allows the effect of covariates to be different at different parts of the conditional distribution. For example, if Y is wages, it is well known that the effect of education is different at different parts of the conditional distribution. One sufficient condition for the unconditional DDID assumption is that X has only a location effect on outcomes. Another sufficient condition is that the distribution of X is the same for the treated and untreated groups. Neither of these conditions seems likely to hold in the types of applications where a researcher is interested in understanding the distributional effect of a program or policy.

Example 1 is a leading case for using distributional methods to understand heterogeneity in the effect of a treatment, and the main conclusion to be reached from this example is that even when an unconditional mean DID assumption holds, one is likely to need to condition on covariates to justify the DDID assumption. On the other hand, in this model, the unconditional CSA continues to hold.[7]

[7] Appendix A discusses the possibility of using the conditional DDID assumption along with a conditional CSA. Identification continues to go through in this case. The advantage of this approach is that it could be used in the case where the trend in outcomes depends on covariates. This could be important in many

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 $F_{\Delta Y_{0t}|D_t=1}(\Delta)$. Under the Distributional Difference in Differences Assumption, this distribution could be replaced directly by $F_{\Delta Y_t|D_t=0}(\Delta)$; 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.

With covariates, we also require an additional standard assumption for identification.

Assumption 5.1. *$p \equiv P(D_t = 1) > 0$ and $p(x) \equiv P(D_t = 1|X = x) < 1$.*

The first part of this assumption says that there is some positive probability that individuals are treated. The second part says 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. This is a standard overlap assumption used in the treatment effects literature.

Theorem 2. *Under the Conditional Distributional Difference in Differences Assumption, the Copula Stability Assumption, and Assumptions 4.1, 4.2 and 5.1*

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

where

$$F^{p}_{\Delta Y_{0t}|D_t=1}(\Delta) = \mathrm{E}\left[\frac{1-D_t}{p}\frac{p(X)}{1-p(X)}\mathbb{1}\{\Delta Y_t \leq \Delta\}\right] \tag{4}$$

and

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

which is identified

This result is very similar to the main identification result in Theorem 1. The only difference is that $\mathrm{F}_{\Delta Y_{0t}|D_t=1}(\cdot)$ is no longer identified by the distribution of untreated potential

applications; for example, suppose that the outcome of interest is wages, the trend in wages may be different for individuals with different education levels. The cost of this approach is that nonparametric estimation would be very challenging in many applications.

outcomes for the untreated group; instead, it is replaced by the re-weighted distribution in Equation 4. Equation 4 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 4 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 $\mathbb{1}\{\Delta Y_t \leq \Delta\}$ in our case is given by Y_t, ΔY_t, and $\mathbb{1}\{Y_t \leq y\}$ in each of the other cases, respectively.

6 Estimation

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

We estimate

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

The first term is estimated directly from the data by inverting the estimated empirical distribution of observed outcomes for the treated group.

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

We estimate counterfactual quantiles by

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

where

$$\hat{\text{F}}_{Y_{0t}|D_t=1}(y) = \frac{1}{n_T}\sum_{i\in\mathcal{T}} \mathbb{1}\{\hat{\text{F}}^{-1}_{\Delta Y_t|D_t=0}(\hat{\text{F}}_{\Delta Y_{t-1}|D_t=1}(\Delta Y_{it-1})) \leq y - \hat{\text{F}}^{-1}_{Y_{t-1}|D_t=1}(\hat{\text{F}}_{Y_{t-2}|D_t=1}(Y_{it-2}))\}$$

which follows from the identification result in Theorem 1 and where distribution functions are estimated by empirical distribution functions and quantile functions are estimated by inverting empirical distribution functions.

The final issue is estimating $\text{F}^{-1}_{\Delta Y_{0t}|D_t=1}(\nu)$ when identification depends on covariates. Using the identification result in Theorem 2, we can easily construct an estimator of the distribution function

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

where the last term in the denominator ensures that $\hat{\mathrm{F}}_{\Delta Y_{0t}|D_t=1}$ is a distribution function and is asymptotically negligible. One can invert this distribution to obtain its quantiles.

When identification depends on covariates X, then there must be a first step estimation of the propensity score. We consider the case where the propensity score is estimated non-parametrically and show that, even though the propensity score itself converges at a slower rate, our estimator of the QTT converges at the parametric $\sqrt{n}$ rate. Also, simpler parametric estimators of the propensity score such as logit or probit can be used instead. All of our main results continue to go though – particularly, the empirical bootstrap can still be used for inference when the propensity score is estimated either parametrically under some mild regularity conditions.

6.1 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 proofs for each of the results in this section are given in the Appendix.

6.1.1 No Covariates Case

This section shows that our estimator of the QTT obeys a functional central limit theorem. In order to show this, the key step is to show that the counterfactual distribution of untreated potential outcomes for the treated group is Hadamard Differentiable.

We denote empirical processes by

$$\hat{G}_X(x) = \sqrt{n}(\hat{F}_X(x) - F_X(x))$$

Next, let $\tilde{Y}_{it} = \mathrm{F}^{-1}_{\Delta Y_t|D_t=0}(\mathrm{F}_{\Delta Y_{t-1}|D_t=1}(\Delta Y_{it-1})) + \mathrm{F}^{-1}_{Y_{t-1}|D_t=1}(\mathrm{F}_{Y_{t-2}|D_t=1}(Y_{it-2}))$; these are pseudo-observations if each distribution and quantile function were known. Let $\tilde{\mathrm{F}}_{Y_{0t}|D_t=1}(y) = \frac{1}{n_T}\sum_{i\in\mathcal{T}}\mathbb{1}\{\tilde{Y}_{it} \leq y\}$. Then, define

$$\tilde{G}_{Y_{0t}|D_t=1}(y) = \sqrt{n}(\tilde{\mathrm{F}}_{Y_{0t}|D_t=1}(y) - \mathrm{F}_{Y_{0t}|D_t=1}(y))$$

As a first step, we establish a functional central limit theorem for the empirical processes of each of the terms used in our identification result.

Proposition 1. *Under the Distributional Difference in Differences Assumption, Copula Stability Assumption, and Assumptions 4.1 and 4.2*

$$(\hat{G}_{\Delta Y_t|D_t=0}, \hat{G}_{\Delta Y_{t-1}|D_t=1}, \tilde{G}_{Y_{0t}|D_t=1}, \hat{G}_{Y_t|D_t=1}, \hat{G}_{Y_{t-1}|D_t=1}, \hat{G}_{Y_{t-2}|D_t=1}) \rightsquigarrow (\mathbb{W}_1, \mathbb{W}_2, \mathbb{V}_0, \mathbb{V}_1, \mathbb{W}_3, \mathbb{W}_4)$$

in the space $\mathcal{S} = l^\infty(\Delta\mathcal{Y}_{t|D_t=0}) \times l^\infty(\Delta\mathcal{Y}_{t-1|D_t=1}) \times l^\infty(\mathcal{Y}_{0t|D_t=1}) \times l^\infty(\mathcal{Y}_{t|D_t=1}) \times l^\infty(\mathcal{Y}_{t-1|D_t=1}) \times l^\infty(\mathcal{Y}_{t-2|D_t=1})$ *where* $(\mathbb{W}_1, \mathbb{W}_2, \mathbb{V}_0, \mathbb{V}_1, \mathbb{W}_3, \mathbb{W}_4)$ *is a tight Gaussian process with mean 0 and block diagonal covariance matrix* $V(y, y') = diag(V_1(y, y'), V_2(y, y'))$ *where*

$$V_1(y, y') = \left(F_{\Delta Y_t|D_t=0}(y_1 \wedge y_1') - F_{\Delta Y_t|D_t=0}(y_1)F_{\Delta Y_t|D_t=0}(y_1')\right)/(1-p)$$

and

$$V_2(y, y') = \mathrm{E}[\psi(y)\psi(y')']$$

for $y = (y_1, y_2, y_3, y_4, y_5, y_6) \in \mathcal{S}$ *and* $y' = (y_1', y_2', y_3', y_4', y_5', y_6') \in \mathcal{S}$ *and*

$$\psi(y) = 1/\sqrt{p} \begin{pmatrix} \mathbb{1}\{\Delta Y_{t-1} \leq y_2\} - \mathrm{F}_{\Delta Y_{t-1}|D_t=1}(y_2) \\ \mathbb{1}\{\tilde{Y}_t \leq y_3\} - \mathrm{F}_{\tilde{Y}_t|D_t=1}(y_3) \\ \mathbb{1}\{Y_t \leq y_4\} - \mathrm{F}_{Y_t|D_t=1}(y_4) \\ \mathbb{1}\{Y_{t-1} \leq y_5\} - \mathrm{F}_{Y_{t-1}|D_t=1}(y_5) \\ \mathbb{1}\{Y_{t-2} \leq y_6\} - \mathrm{F}_{Y_{t-2}|D_t=1}(y_6) \end{pmatrix}$$

The next result establishes the joint limiting distribution of observed treated outcomes and counterfactual untreated potential outcomes for the treated group.

Proposition 2. *Let* $\hat{G}_0(y) = \sqrt{n}(\hat{\mathrm{F}}_{Y_{0t}|D_t=1}(y) - \mathrm{F}_{Y_{0t}|D_t=1}(y))$ *and let* $\hat{G}_1(y) = \sqrt{n}(\hat{\mathrm{F}}_{Y_t|D_t=1}(y) - \mathrm{F}_{Y_t|D_t=1}(y))$. *Under Assumptions Distributional Difference in Differences Assumption, Copula Stability Assumption, and Assumptions 4.1 and 4.2*

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

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

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

and

$$\mathbb{G}_0 = \mathbb{V}_0 + \int \left\{ \mathbb{W}_1 \circ F^{-1}_{Y_{t-1}|D_t=1} \circ F_{Y_{t-2}|D_t=1}(v) - F_{\Delta Y_t|D_t=0}\left(y - \frac{\mathbb{W}_4 - \mathbb{W}_3 \circ F^{-1}_{Y_{t-1}|D_t=1} \circ F_{Y_{t-2}|D_t=1}(v)}{f_{Y_{t-1}|D_t=1} \circ F^{-1}_{Y_{t-1}|D_t=1} \circ F_{Y_{t-2}|D_t=1}(v)} \right) \right.$$
$$\left. - \mathbb{W}_2 \circ F^{-1}_{\Delta Y_{t-1}|D_t=1} \circ F_{\Delta Y_t|D_t=0}(y - F^{-1}_{Y_{t-1}|D_t=1} \circ F_{Y_{t-2}|D_t=1}(v)) \right\} K(y,v) \, \mathrm{d}F_{Y_{t-2}|D_t=1}(v)$$

where

$$K(y,v) = \frac{f_{\Delta Y_{t-1}|Y_{t-2},D_t=1}(F^{-1}_{\Delta Y_{t-1}|D_t=1} \circ F_{\Delta Y_t|D_t=0}(y - F^{-1}_{Y_{t-1}|D_t=1} \circ F_{Y_{t-2}|D_t=1}(v)))}{f_{\Delta Y_{t-1}|D_t=1} \circ F^{-1}_{\Delta Y_{t-1}|D_t=1} \circ F_{\Delta Y_t|D_t=0} \circ (y - F^{-1}_{Y_{t-1}|D_t=1} \circ F_{Y_{t-2}|D_t=1}(v))}$$

The key step in showing Proposition 2 is establishing the Hadamard Differentiability of the counterfactual distribution of untreated potential outcomes for the treated group. Here, $\mathbb{V}_0$ is the variance that would obtain if each distribution and quantile function were known. The second term comes from having to estimate each of these distribution and quantile functions in a first step. With Proposition 2 in hand, our main result for the QTT follows straightforwardly by the Hadamard Differentiability of quantiles.

Theorem 3. *Suppose* $\mathrm{F}_{Y_{0t}|D_t=1}$ *admits a positive continuous density* $f_{Y_{0t}|D_t=1}$ *on an interval* $[a,b]$ *containing an* ε*-enlargement of the set* $\{\mathrm{F}^{-1}_{Y_{0t}|D_t=1}(\tau) : \tau \in \mathcal{T}\}$ *in* $\mathcal{Y}_{0t|D_t=1}$ *with* $\mathcal{T} \subset (0,1)$*. Under the Distributional Difference in Differences Assumption, the Copula Stability Assumption, and Assumptions 4.1 and 4.2*

$$\sqrt{n}(\hat{QTT}(\tau) - QTT(\tau)) \rightsquigarrow \bar{\mathbb{G}}_1(\tau) - \bar{\mathbb{G}}_0(\tau)$$

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

$$\bar{\mathbb{G}}_0(\tau) = \frac{\mathbb{G}_0(\mathrm{F}^{-1}_{Y_{0t}|D_t=1}(\tau))}{f_{Y_{0t}|D_t=1}(\mathrm{F}^{-1}_{Y_{0t}|D_t=1}(\tau))} \quad \text{and} \quad \bar{\mathbb{G}}_1(\tau) = \frac{\mathbb{G}_1(\mathrm{F}^{-1}_{Y_t|D_t=1}(\tau))}{f_{Y_t|D_t=1}(\mathrm{F}^{-1}_{Y_t|D_t=1}(\tau))}$$

Estimating the asymptotic variance of our estimator is likely to be quite complicated particularly due to the presence of density functions which would require smoothing and choosing some tuning parameters. Instead, we conduct inference using the nonparametric bootstrap.

Algorithm 1. *Let* B *be the number of bootstrap iterations. For* $b = 1, \ldots, B$*,*

1. *Draw a sample of size* n *with replacement from the original data*

2. *Compute*

$$\hat{QTT}^b(\tau) = \hat{\mathrm{F}}^{-1b}_{Y_t|D_t=1}(\tau) - \hat{\mathrm{F}}^{-1b}_{Y_{0t}|D_t=1}(\tau)$$

where

$$\hat{\mathrm{F}}^b_{Y_{0t}|D_t=1}(y) = \frac{1}{n^b_T}\sum_{i\in\mathcal{T}} \mathbb{1}\{\hat{\mathrm{F}}^{-1b}_{\Delta Y_t|D_t=0}(\hat{\mathrm{F}}^b_{\Delta Y_{t-1}|D_t=1}(\Delta Y^b_{it-1})) \leq y - \hat{\mathrm{F}}^{-1b}_{Y_{t-1}|D_t=1}(\hat{\mathrm{F}}^b_{Y_{t-2}|D_t=1}(Y^b_{it-2}))\}$$

and the superscript b indicates that the distribution or quantile function is computed using the bootstrap data.

3. *Compute* $I^b = \sup_{\tau\in\mathcal{T}} \left|\hat{QTT}^b(\tau) - \hat{QTT}(\tau)\right|$

Then, a $(1-\alpha)$ confidence band is given by

$$\hat{QTT}(\tau) - c^B_{1-\alpha}/\sqrt{n} \leq QTT(\tau) \leq \hat{QTT}(\tau) + c^B_{1-\alpha}/\sqrt{n}$$

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

The next result shows the validity of the nonparametric bootstrap for our procedure.

Theorem 4. *Under the Distributional Difference in Differences Assumption, Copula Stability Assumption, and Assumptions 4.1 and 4.2,*

$$\sqrt{n}\left(\hat{QTT}^*(\tau) - \hat{QTT}(\tau)\right) \rightsquigarrow_* \bar{\mathbb{G}}_0(\tau) - \bar{\mathbb{G}}_1(\tau)$$

where $(\bar{\mathbb{G}}_0, \bar{\mathbb{G}}_1)$ *are as in Theorem 3 and* $\rightsquigarrow_*$ *indicates weak convergence in probability under the bootstrap law (Giné and Zinn, 1990)*

Theorem 4 follows because our estimate of the QTT is Donsker and by Van Der Vaart and Wellner, (1996, Theorem 3.6.1)

6.1.2 Distributional Difference in Differences Assumption holds conditional on covariates

This section develops the asymptotic properties of our estimator in the case where the Distributional Difference in Differences Assumption holds conditional on covariates and consider the case where the propensity score is estimated nonparametrically by using series logit methods. Following Hirano, Imbens, and Ridder, (2003), we make the following assumptions on the propensity score

Assumption 6.1. $\mathrm{E}[\mathbb{1}\{\Delta Y_{0t} \leq y\}|X, D_t = 0]$ *is continuously differentiable for all* $x \in \mathcal{X}$.

Assumption 6.2. *(Distribution of X)*

(i) The support $\mathcal{X}$ *of* X *is a Cartesian product of compact intervals; 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 6.3. *(Assumptions on the propensity score)*

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

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

Assumption 6.4. *(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* $\frac{1}{4(s/r-1)} < \nu < \frac{1}{9}$.

Remark. Assumptions Assumptions 6.1 to 6.4 are standard assumptions in the literature which depends on first step estimation of the propensity score. Hirano, Imbens, and Ridder, (2003) developed the properties of the series logit estimator under the same set of assumptions. Similar assumptions have been used in, for example, Firpo, (2007) and Donald and Hsu, (2014). Assumption 6.2 says that X is continuously distributed though our setup can easily handle discrete covariates as well by splitting the sample based on the discrete covariates. Assumption 6.3(i) is a standard assumption on differentiability of the propensity score and guarantees the existence of ν that satisfies the conditions of Assumption 6.4. Assumption 6.3(ii) is a standard overlap condition.

Proposition 3. *Let* $\hat{G}^p_{\Delta Y_{0t}|D_t=1}(\Delta) = \sqrt{n}\left(\hat{\mathrm{F}}^p_{\Delta Y_{0t}|D_t=1}(\Delta) - \mathrm{F}^p_{\Delta Y_{0t}|D_t=1}(\Delta)\right)$ *where* $\mathrm{F}^p_{Y_{0t}|D_t=1}(\Delta)$ *is given in Equation* (4). *Let* $\tilde{Y}^p_{it} = \mathrm{F}^{-1p}_{\Delta Y_{0t}|D_t=1}(\mathrm{F}_{\Delta Y_{t-1}|D_t=1}(\Delta Y_{it-1}))+\mathrm{F}^{-1}_{Y_{t-1}|D_t=1}(\mathrm{F}_{Y_{t-2}|D_t=1}(Y_{it-2}))$, *let* $\tilde{\mathrm{F}}^p_{Y_{0t}|D_t=1}(y) = \frac{1}{n_T}\sum_{i\in\mathcal{T}}\mathbb{1}\{\tilde{Y}^p_{it} \leq y\}$, *and let* $\tilde{G}^p_{Y_{0t}|D_t=1}(y) = \left(\tilde{\mathrm{F}}^p_{Y_{0t}|D_t=1}(y) - \mathrm{F}_{Y_{0t}|D_t=1}(y)\right)$. *Under the Conditional Distributional Difference in Differences Assumption, the Copula Stability Assumption, Assumptions 4.1, 4.2, 5.1 and 6.1 to 6.4*

$$(\hat{G}^p_{\Delta Y_{0t}|D_t=1}, \hat{G}_{\Delta Y_{t-1}|D_t=1}, \tilde{G}^p_{Y_{0t}|D_t=1}, \hat{G}_{Y_t|D_t=1}, \hat{G}_{Y_{t-1}|D_t=1}, \hat{G}_{Y_{t-2}|D_t=1}) \rightsquigarrow (\mathbb{W}^p_1, \mathbb{W}^p_2, \mathbb{V}^p_0, \mathbb{V}^p_1, \mathbb{W}^p_3, \mathbb{W}^p_4)$$

in the space $\mathcal{S} = l^{\infty}(\Delta\mathcal{Y}_{t|D_t=0}) \times l^{\infty}(\Delta\mathcal{Y}_{t-1|D_t=1}) \times l^{\infty}(\mathcal{Y}_{0t|D_t=1}) \times l^{\infty}(\mathcal{Y}_{t|D_t=1}) \times l^{\infty}(\mathcal{Y}_{t-1|D_t=1}) \times l^{\infty}(\mathcal{Y}_{t-2|D_t=1})$ *where* $(\mathbb{W}^p_1, \mathbb{W}^p_2, \mathbb{V}^p_0, \mathbb{V}^p_1, \mathbb{W}^p_3, \mathbb{W}^p_4)$ *is a tight Gaussian process with mean 0 and covariance* $V(y, y') = \mathrm{E}[\psi^p(y)\psi^p(y')']$ *for* $y = (y_1, y_2, y_3, y_4, y_5, y_6) \in \mathcal{S}$ *and* $y' = (y'_1, y'_2, y'_3, y'_4, y'_5, y'_6) \in$

$\mathcal{S}$ *and with* $\psi^p(y)$ *given by*

$$\psi^p(y) = \begin{pmatrix} \frac{\mathbb{1}\{\Delta Y \leq y_1|X\}}{p(1-p(X))}(D-p(X)) + \frac{1-D}{p}\frac{p(X)}{1-p(X)}\mathbb{1}\{\Delta Y_t \leq y_1\} - \mathrm{F}^p_{\Delta Y_{0t}|D_t=1}(y_1) \\ \frac{D}{p}\mathbb{1}\{\Delta Y_{t-1} \leq y_2\} - \mathrm{F}_{\Delta Y_{t-1}|D_t=1}(y_2) \\ \frac{D}{p}\mathbb{1}\{\tilde{Y}_t \leq y_3\} - \mathrm{F}_{\tilde{Y}_t|D_t=1}(y_3) \\ \frac{D}{p}\mathbb{1}\{Y_t \leq y_4\} - \mathrm{F}_{Y_t|D_t=1}(y_4) \\ \frac{D}{p}\mathbb{1}\{Y_{t-1} \leq y_5\} - \mathrm{F}_{Y_{t-1}|D_t=1}(y_5) \\ \frac{D}{p}\mathbb{1}\{Y_{t-2} \leq y_6\} - \mathrm{F}_{Y_{t-2}|D_t=1}(y_6) \end{pmatrix}$$

The next result establishes an analogous result to Proposition 2 for the case where identification depends on covariates.

Proposition 4. *Let* $\hat{G}^p_0(y) = \sqrt{n}(\hat{\mathrm{F}}^p_{Y_{0t}|D_t=1}(y) - \mathrm{F}^p_{Y_{0t}|D_t=1}(y))$ *and let* $\hat{G}^p_1(y) = \sqrt{n}(\hat{\mathrm{F}}_{Y_t|D_t=1}(y) - \mathrm{F}_{Y_t|D_t=1}(y))$. *Under the Conditional Distributional Difference in Differences Assumption, Copula Stability Assumption, and Assumptions 4.1, 4.2, 5.1 and 6.1 to 6.4*

$$(\hat{G}^p_0, \hat{G}^p_1) \rightsquigarrow (\mathbb{G}^p_0, \mathbb{G}^p_1)$$

where $\mathbb{G}^p_0$ *and* $\mathbb{G}^p_1$ *are tight Gaussian processes with mean 0 with almost surely uniformly continuous paths on the space* $\mathcal{Y}_{0t|D_t=1} \times \mathcal{Y}_{t|D_t=1}$ *given by*

$$\mathbb{G}^p_1 = \mathbb{V}^p_1$$

and

$$\mathbb{G}^p_0 = \mathbb{V}^p_0 + \int \left\{ \mathbb{W}^p_1 \circ F^{-1}_{Y_{t-1}|D_t=1} \circ F_{Y_{t-2}|D_t=1}(v) - F^p_{\Delta Y_t|D_t=1}\left(y - \frac{\mathbb{W}^p_4 - \mathbb{W}^p_3 \circ F^{-1}_{Y_{t-1}|D_t=1} \circ F_{Y_{t-2}|D_t=1}(v)}{f_{Y_{t-1}|D_t=1} \circ F^{-1}_{Y_{t-1}|D_t=1} \circ F_{Y_{t-2}|D_t=1}(v)}\right) \right.$$
$$\left. - \mathbb{W}^p_2 \circ F^{-1}_{\Delta Y_{t-1}|D_t=1} \circ F^p_{\Delta Y_t|D_t=1}(y - F^{-1}_{Y_{t-1}|D_t=1} \circ F_{Y_{t-2}|D_t=1}(v)) \right\} K(y,v)\ \mathrm{d}F_{Y_{t-2}|D_t=1}(v)$$

where

$$K(y,v) = \frac{f_{\Delta Y_{t-1}|Y_{t-2},D_t=1}(F^{-1}_{\Delta Y_{t-1}|D_t=1} \circ F^p_{\Delta Y_t|D_t=1}(y - F^{-1}_{Y_{t-1}|D_t=1} \circ F_{Y_{t-2}|D_t=1}(v)))}{f_{\Delta Y_{t-1}|D_t=1} \circ F^{-1}_{\Delta Y_{t-1}|D_t=1} \circ F^p_{\Delta Y_t|D_t=1} \circ (y - F^{-1}_{Y_{t-1}|D_t=1} \circ F_{Y_{t-2}|D_t=1}(v))}$$

Theorem 5. *Suppose* $\mathrm{F}_{Y_{0t}|D_t=1}$ *admits a positive continuous density* $f_{Y_{0t}|D_t=1}$ *on an interval* $[a,b]$ *containing an* ε*-enlargement of the set* $\{\mathrm{F}^{-1p}_{Y_{0t}|D_t=1}(\tau) : \tau \in \mathcal{T}\}$ *in* $\mathcal{Y}_{0t|D_t=1}$ *with* $\mathcal{T} \subset (0,1)$. *Under the Conditional Distributional Difference in Differences Assumption, the*

Copula Stability Assumption, and Assumptions 4.1, 4.2, 5.1 and 6.1 to 6.4

$$\sqrt{n}(Q\hat{T}T^p(\tau) - QTT^p(\tau)) \rightsquigarrow \bar{\mathbb{G}}_1^p(\tau) - \bar{\mathbb{G}}_0^p(\tau)$$

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

$$\bar{\mathbb{G}}_0^p(\tau) = \frac{\mathbb{G}_0^p(\mathrm{F}^{-1}_{Y_{0t}|D_t=1}(\tau))}{f_{Y_{0t}|D_t=1}(\mathrm{F}^{-1}_{Y_{0t}|D_t=1}(\tau))} \quad \text{and} \quad \bar{\mathbb{G}}_1^p(\tau) = \frac{\mathbb{G}_1^p(\mathrm{F}^{-1}_{Y_t|D_t=1}(\tau))}{f_{Y_t|D_t=1}(\mathrm{F}^{-1}_{Y_t|D_t=1}(\tau))}$$

Finally, we show that the empirical bootstrap can be used to construct asymptotically valid confidence bands. The steps for the bootstrap are the same as in Algorithm 1 – only the $\mathrm{F}_{\Delta Y_{0t}|D_t=1}(\Delta)$ should be calculated using the result on re-weighting rather than replacing it directly with $\mathrm{F}_{\Delta Y_t|D_t=0}(\Delta)$. The same series terms used to estimate the propensity score can be reused in each bootstrap iteration. Theorem 6 follows essentially using the same arguments as in Chen, Linton, and Van Keilegom, (2003).

Theorem 6. *Under the Conditional Distributional Difference in Differences Assumption, Copula Stability Assumption, and Assumptions 4.1, 4.2, 5.1 and 6.1 to 6.4,*

$$\sqrt{n}\left(Q\hat{T}T^{p*}(\tau) - Q\hat{T}T^p(\tau)\right) \rightsquigarrow_* \bar{\mathbb{G}}_1^p(\tau) - \bar{\mathbb{G}}_0^p(\tau)$$

where $(\bar{\mathbb{G}}_0^p, \bar{\mathbb{G}}_1^p)$ *are as in Theorem 5.*

7 Comparison with Existing Methods

Our method is related to the work on quantile regression with panel data (Koenker, 2004; Abrevaya and Dahl, 2008; Lamarche, 2010; Canay, 2011; Rosen, 2012; Galvao, Lamarche, and Lima, 2013; Chen, 2015) though our method is distinct in several ways. First, because we do not impose a parametric model, our method allows for the effect of treatment to vary across individuals with different covariates in an unspecified way. Second, our method is consistent under fixed-T asymptotics while the papers mentioned above generally require $T \to \infty$.[8] Third, we focus on an unconditional QTT whereas the quantile treatment effects identified in these models are conditional – both on covariates and on unobserved heterogeneity. This means that the results from our method should be interpreted in the same way as the difference between treated and untreated quantiles if individuals were randomly

[8]The two exceptions are Abrevaya and Dahl, (2008) which uses a correlated random effects structure to obtain identification without $T \to \infty$ and Rosen, (2012) which deals with partial identification under quantile restrictions.

assigned to treatment. See Frölich and Melly, (2013) for a good discussion of the difference between conditional and unconditional quantile treatment effects. On the other hand, our method only applies to the case where the researcher is interested only in the effect of a binary treatment; quantile regression methods can can deliver estimates for multiple, possibly continuous variables.

Because we focus on nonparametric identifying assumptions, the current paper is also related to the literature on nonseparable panel data models (Altonji and Matzkin, 2005; Evdokimov, 2010; Bester and Hansen, 2012; Graham and Powell, 2012; Hoderlein and White, 2012; Chernozhukov, Fernández-Val, Hahn, and Newey, 2013). The most similar of these is Chernozhukov, Fernández-Val, Hahn, and Newey, (2013) which considers a nonseparable model and, similarly to our paper, obtains point identification for observations that are observed in both treated and untreated states. Relative to Chernozhukov, Fernández-Val, Hahn, and Newey, (2013), we exploit having access to a control group much more – their approach either does not use the control group or uses it to adjust the mean and variance only – and our setup is compatible with more complicated distributional shifts in outcomes over time such as the top of the income distribution increasing more than the bottom of the income distribution.

Perhaps the most similar work to ours is Athey and Imbens, (2006). Their Change in Changes model identifies the QTT for models that are monotone is a scalar unobservable. They assume that the distribution of unobservables does not change over time (though the distribution of unobservables can be different for the treated group and untreated group) but allow for the return to unobservables to change over time. However, even a mean Difference in Differences Assumption does not hold in general in their model. Interestingly, one model that satisfies the Change in Changes model and our setup is when untreated potential outcomes at period s are generated by $Y_{0is} = C_i + V_{is} + \theta_s$ for $s = t, t-1, t-2$ where C_i is an individual specific fixed effect, θ_s is a time fixed effect and V_{is} is an idiosyncratic error term such that $V_s|C \sim F_V$ for all s.

8 Evidence on the Validity of the Copula Stability Assumption

In order to assess the validity of the Copula Stability Assumption, we first provide some general empirical evidence testing whether or not the dependence between the change in outcomes and the initial level of outcomes is constant over time. In order to do this, we use bi-annual earnings data from the National Longitudinal Study of Youths and estimate

Spearman's Rho – a common dependence measure Nelsen, (2007) – for each year. Spearman's Rho is bounded between -1 and 1. If Spearman's Rho is constant over time, this provides evidence in favor of the CSA. Spearman's Rho fluctuating over time would indicate that the CSA is violated.

Figure 1 plots our estimates of the Spearman's Rho for each even year from 1992 to 2012 using a sample of 2,283 NLSY participants with positive earnings in each period. Standard errors are calculated using the block bootstrap. Spearman's Rho is essentially constant across all periods and close to 0.

Second, in a particular application, 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 copula function is stable over time.

For an applied researcher looking for a simpler test, another idea would be to simply test whether a dependence measure, such as Spearman's Rho or Kendall's Tau is constant over time. With only three periods, another pseudo-test would be to test whether or not the Copula Stability Assumption holds for the untreated group.

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: $F_{\Delta Y_{t-1}|D_t=1}(\Delta)$ and $F_{\Delta Y_{t-1}|D_t=0}(\Delta)$. 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.

9 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 QTT estimates to those from the randomized experiment and the various other procedures available to estimate quantile treatment effects.

9.1 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 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.

9.2 Results

The PanelQTT 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

[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.

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 2 provides estimates of the 0.7-, 0.8-, and 0.9-QTT using the method of this paper (which we hereafter term PanelQTT), the conditional independence (CI) method (Firpo, 2007), the Change in Changes method (Athey and Imbens, 2006), and the Quantile Difference in Differences (QDiD) 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 PanelQTT 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 PanelQTT method and the CI method use slightly different series logit estimates. For the PanelQTT 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 and QDiD, 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 QTT using the PanelQTT method. The first row provides results where the propensity score is estimated nonparametrically using series logit. The estimated QTT 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 QTT 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 PanelQTT method. Out of all 16 method-covariate set estimates presented in Table 2, the QTTs come closest to matching the experimental results using the PanelQTT method and the COV conditioning set. When using the UNEM conditioning set, the estimates of the QTT are very similar to the nonparametric specification. Finally, the NO COV conditioning set tends to perform the most poorly. The QTT 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 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 PanelQTT model is that the CI model would include lags of the dependent variable but no fixed effect while the PanelQTT 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 PanelQTT 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 QTTs are estimated to be large

and negative indicating that participating in the job training program tended to strongly decrease earnings. 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 two sections of Table 2 provide results using CiC and QDiD. We briefly summarize these results. Broadly speaking, each of these 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 methods, the QDiD method performs slightly better than the CiC model. Comparing the results of these three models to the results from the PanelQTT method, the PanelQTT method performs slightly better than the CiC model. With the COV specification, it performs evenly with the QDiD method. With the UNEM specification, it performs slightly worse than the QDiD method.

10 Conclusion

This paper has considered identification and estimation of the QTT under a distributional extension of the most common Mean Difference in Differences Assumption used to identify the ATT. Even under this Distributional Difference in Differences Assumption, the QTT is still only partially identified because it depends on the unknown dependence between the change in untreated potential outcomes and the initial level of untreated potential outcomes for the treated group. We introduced the Copula Stability Assumption which says that the missing dependence is constant over time. Under this assumption and when panel data is available, the QTT is point identified. We show that the Copula Stability Assumption is likely to hold in exactly the type of models that are typically estimated using Difference in Differences techniques.

In many applications it is important to invoke identifying assumptions that hold only after conditioning on some covariates. We developed very simple estimators of the QTT using propensity score re-weighting. In an application where we compare the results using several available methods to estimate the QTT on observational data to results obtained

from an experiment, we find that our method performs at least as well as other available methods.

In ongoing work, we are using similar ideas about the time invariance of a copula function to study the joint distribution of treated and untreated potential outcomes when panel data is available. Also, we are working on using the same type of assumption to identify the QTT in more complicated situations such as when outcomes are censored or in dynamic panel data models. The idea of a time invariant copula may also be valuable in other areas of microeconometric research especially when a researcher has access to panel data.

References

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

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

[3] Abrevaya, Jason and Christian Dahl. "The effects of birth inputs on birthweight: Evidence from quantile estimation on panel data". *Journal of Business & Economic Statistics* 26.4 (2008), pp. 379–397.

[4] Altonji, Joseph and Rosa Matzkin. "Cross section and panel data estimators for nonseparable models with endogenous regressors". *Econometrica* (2005), pp. 1053–1102.

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

[6] Bester, C Alan and Christian Hansen. "Identification of marginal effects in a nonparametric correlated random effects model". *Journal of Business & Economic Statistics* (2012).

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

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

[9] Callaway, Brantly, Tong Li, and Tatsushi Oka. "Quantile Treatment Effects in Difference in Differences Models under Dependence Restrictions and with only Two Time Periods". Working Paper. 2016.

[10] Canay, Ivan. "A simple approach to quantile regression for panel data". *The Econometrics Journal* 14.3 (2011), pp. 368–386.

[11] Card, David. "The impact of the mariel boatlift on the miami labor market". *Industrial & Labor Relations Review* 43.2 (1990), pp. 245–257.

[12] Card, David and Alan Krueger. "Minimum wages and employment: A case study of the fast-food industry in new jersey and pennsylvania". *The American Economic Review* 84.4 (1994), p. 772.

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

[14] Carneiro, Pedro and Sokbae Lee. "Estimating distributions of potential outcomes using local instrumental variables with an application to changes in college enrollment and wage inequality". *Journal of Econometrics* 149.2 (2009), pp. 191–208.

[15] Chen, Heng. "Within-group estimators for fixed effects quantile models with large N and large T". Working Paper, Bank of Canada. 2015.

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

[17] Chernozhukov, Victor, Iván Fernández-Val, Jinyong Hahn, and Whitney Newey. "Average and quantile effects in nonseparable panel models". *Econometrica* 81.2 (2013), pp. 535–580.

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

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

[20] Doksum, Kjell. "Empirical probability plots and statistical inference for nonlinear models in the two-sample case". *The Annals of Statistics* (1974), pp. 267–277.

[21] Donald, Stephen G and Yu-Chin Hsu. "Estimation and inference for distribution functions and quantile functions in treatment effect models". *Journal of Econometrics* 178 (2014), pp. 383–397.

[22] Evdokimov, Kirill. "Identification and estimation of a nonparametric panel data model with unobserved heterogeneity". Working Paper, Princeton University. 2010.

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

[24] Finkelstein, Amy and Robin McKnight. "What did medicare do? The initial impact of medicare on mortality and out of pocket medical spending". *Journal of Public Economics* 92.7 (2008), pp. 1644–1668.

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

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

[27] Galvao, Antonio F, Carlos Lamarche, and Luiz Renato Lima. "Estimation of censored quantile regression for panel data with fixed effects". *Journal of the American Statistical Association* 108.503 (2013), pp. 1075–1089.

[28] Giné, Evarist and Joel Zinn. "Bootstrapping general empirical measures". *The Annals of Probability* (1990), pp. 851–869.

[29] Graham, Bryan and James Powell. "Identification and estimation of average partial effects in irregular correlated random coefficient panel data models". *Econometrica* (2012), pp. 2105–2152.

[30] Havnes, Tarjei and Magne Mogstad. "Is universal child care leveling the playing field?" *Journal of Public Economics* 127 (2015), pp. 100–114.

[31] Heckman, James and V Joseph Hotz. "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 (1989), pp. 862–874.

[32] Heckman, James, Hidehiko Ichimura, Jeffrey Smith, and Petra Todd. "Characterizing selection bias using experimental data". *Econometrica* 66.5 (1998), pp. 1017–1098.

[33] Heckman, James, Hidehiko Ichimura, and Petra Todd. "Matching as an econometric evaluation estimator: Evidence from evaluating a job training programme". *The Review of Economic Studies* 64.4 (1997), pp. 605–654.

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

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

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

[37] Hoderlein, Stefan and Halbert White. "Nonparametric identification in nonseparable panel data models with generalized fixed effects". *Journal of Econometrics* 168.2 (2012), pp. 300–314.

[38] Hollister, Robinson, Peter Kemper, and Rebecca Maynard. *The National Supported Work Demonstration*. Univ of Wisconsin Pr, 1984.

[39] Joe, Harry. *Multivariate Models and Multivariate Dependence Concepts*. CRC Press, 1997.

[40] Joe, Harry. *Dependence Modeling with Copulas*. CRC Press, Boca Raton, FL, 2015.

[41] Jun, Sung Jae, Yoonseok Lee, and Youngki Shin. "Treatment effects with unobserved heterogeneity: A set identification approach". *Journal of Business & Economic Statistics* 34.2 (2016), pp. 302–311.

[42] Koenker, Roger. "Quantile regression for longitudinal data". *Journal of Multivariate Analysis* 91.1 (2004), pp. 74–89.

[43] Koenker, Roger. *Quantile Regression*. Cambridge University Press, 2005.

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

[45] Lamarche, Carlos. "Robust penalized quantile regression estimation for panel data". *Journal of Econometrics* 157.2 (2010), pp. 396–408.

[46] Lee, Myoung-jae and Changhui Kang. "Identification for difference in differences with cross-section and panel data". *Economics Letters* 92.2 (2006), pp. 270–276.

[47] Lehmann, Erich. *Nonparametrics: Statistical Methods Based on Ranks*. 1974.

[48] Melly, Blaise and Giulia Santangelo. "The changes-in-changes model with covariates". Working Paper. 2015.

[49] Meyer, Bruce, Kip Viscusi, and David Durbin. "Workers' compensation and injury duration: Evidence from a natural experiment". *The American Economic Review* (1995), pp. 322–340.

[50] Nelsen, Roger. *An Introduction to Copulas*. 2nd ed. Springer, 2007.

[51] Pomeranz, Dina. "No taxation without information: deterrence and self-enforcement in the value added tax". *The American Economic Review* 105.8 (2015), pp. 2539–2569.

[52] Rosen, Adam. "Set identification via quantile restrictions in short panels". *Journal of Econometrics* 166.1 (2012), pp. 127–137.

[53] Sen, Amartya. *On Economic Inequality*. Clarendon Press, 1997.

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

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

[56] Vaart, Aad W van der and Jon A Wellner. "Empirical processes indexed by estimated functions". *Lecture Notes-Monograph Series* (2007), pp. 234–252.

[57] Van der Vaart, Aad. *Asymptotic Statistics*. Vol. 3. Cambridge University Press, 2000.

[58] Van Der Vaart, Aad W and Jon A Wellner. *Weak Convergence*. Springer, 1996.

[59] Wüthrich, Kaspar. "Semiparametric estimation of quantile treatment effects with endogeneity". Working Paper. 2015.

A Identification and Estimation under a Conditional CSA

Our main results dealt with the case where the Distributional Difference in Differences Assumption held conditional on covariates, but the Copula Stability Assumption held unconditionally. We showed that this combination of assumptions is likely to hold in the most common type of model where empirical researchers use Difference in Differences to identify the ATT. We also provided some empirical evidence in favor of the Unconditional Copula Stability Assumption.

However, in some applications, a researcher may wish to make the Copula Stability Assumption hold after conditioning on covariates. This assumption says that the copula between the change in untreated potential outcomes and the initial level of untreated potential outcomes does not change over time after conditioning on some covariates X.

Conditional Copula Stability Assumption.

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

Importantly, the QTT continues to be identified under the Conditional Copula Stability Assumption.

Proposition 5. *Assume that, for all $x \in \mathcal{X}$, ΔY_t for the untreated group, ΔY_{t-1}, Y_{t-1}, and Y_{t-2} for the treated group are continuously distributed conditional on x. Under the Conditional Distributional Difference in Differences Assumption, the Conditional Copula Stability Assumption, and Assumption 4.2*

$$\begin{aligned}
&\mathrm{P}(Y_{0t} \le y|X=x, D_t=1) \\
&\quad = \mathrm{E}\Big[\mathbb{1}\{F^{-1}_{\Delta Y_{0t}|X,D_t=0}(F_{\Delta Y_{0t-1}|X,D_t=1}(\Delta Y_{0t-1}|x)) \\
&\qquad\qquad \le y - F^{-1}_{Y_{0t-1}|X,D_t=1}(F_{Y_{0t-2}|X,D_t=1}(Y_{0t-2}|x))\}|X=x, D_t=1\Big]
\end{aligned}$$

and

$$\mathrm{QTT}(\tau;x) = \mathrm{F}^{-1}_{Y_{1t}|X,D_t=1}(\tau|x) - \mathrm{F}^{-1}_{Y_{0t}|X,D_t=1}(\tau|x)$$

which is identified, and

$$\mathrm{P}(Y_{0t} \le y|D_t=1) = \int_{\mathcal{X}} \mathrm{P}(Y_{0t} \le y|X=x, D_t=1)\ \mathrm{dF}(x|D_t=1)$$

and

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

which is identified.

There are several advantages to this approach. First, under the Conditional Copula Stability Assumption, the path of untreated potential outcomes can depend on the covariates.

This could be important in applications where the return to some covariate – for example, the return to education – changes over time. Conditional Difference in Differences assumptions for the ATT (Heckman, Ichimura, Smith, and Todd, 1998; Abadie, 2005) allow for this pattern. Second, under the Conditional Copula Stability Assumption, it is possible to allow for time varying covariates; however, the effect of time varying covariates must be a location-shift. Finally, under the Conditional Copula Stability Assumption, one can obtain estimates of conditional quantile treatment effects.

On the other hand, there are some costs associated with the Conditional Copula Stability Assumption. Primarily, estimation becomes potentially much more challenging. Nonparametric estimation would require estimating five conditional distribution functions and conditional quantile functions which is likely to be quite challenging in practice. One could replace nonparametric estimation by assuming a parametric model for each conditional quantile function though parametric assumptions are unattractive in our model because it is not clear how misspecification in any of the first step conditional distribution/quantile functions would affect our estimates of the QTT.

In ongoing work (Callaway, Li, and Oka, 2016), we consider a conditional copula assumption in a related model. Those results are likely to go through with minor adaptations to the current model. Melly and Santangelo, (2015) use parametric quantile regressions to estimate a conditional version of the Change in Changes model (Athey and Imbens, 2006); Wüthrich, (2015) uses a similar approach to estimate quantile treatment effects with endogeneity. One could also adapt those types of results to our setup in a straightforward way.

B Proofs

B.1 Identification

B.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 B.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 B.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,

$$
\begin{aligned}
P(Y_{0t} \leq y | D_t = 1) &= P(\Delta Y_{0t} + Y_{0t-1} \leq y | D_t = 1) \\
&= \mathrm{E}[\mathbb{1}\{\Delta Y_{0t} \leq y - Y_{0t-1}\} | D_t = 1) \\
&= \int_{\mathcal{Y}_{t-1|D_t=1}} \int_{\Delta\mathcal{Y}_{t|D_t=1}} \mathbb{1}\{\Delta y_{0t} \leq y - y_{0t-1}\} \varphi_t(\Delta y_{0t}, y_{0t-1} | D_t = 1) \mathrm{d}\Delta y_{0t} \mathrm{d}y_{0t-1} \\
&= \int_{\mathcal{Y}_{t-1|D_t=1}} \int_{\Delta\mathcal{Y}_{t|D_t=1}} \mathbb{1}\{\Delta y_{0t} \leq y - y_{0t-1}\} \qquad (5) \\
&\qquad \times \quad c_t(F_{\Delta Y_{0t}|D_t=1}(\Delta y_{0t}), F_{Y_{0t-1}|D_t=1}(y_{0t-1})) \\
&\qquad \times \quad f_{\Delta Y_{0t}|D_t=1}(\Delta y_{0t}) f_{Y_{0t-1}|D_t=1}(y_{0t-1}) \mathrm{d}\Delta y_{0t} \mathrm{d}y_{0t-1} \\
&= \int_{\mathcal{Y}_{t-1|D_t=1}} \int_{\Delta\mathcal{Y}_{t|D_t=1}} \mathbb{1}\{\Delta y_{0t} \leq y - y_{0t-1}\} \qquad (6) \\
&\qquad \times \quad c_{t-1}(F_{\Delta Y_{0t}|D_t=1}(\Delta y_{0t}), F_{Y_{0t-1}|D_t=1}(y_{0t-1})) \\
&\qquad \times \quad f_{\Delta Y_{0t}|D_t=1}(\Delta y_{0t}) f_{Y_{0t-1}|D_t=1}(y_{0t-1}) \mathrm{d}\Delta y_{0t} \mathrm{d}y_{0t-1} \\
&= \int_{\mathcal{Y}_{t-1|D_t=1}} \int_{\Delta\mathcal{Y}_{t|D_t=1}} \mathbb{1}\{\Delta y_{0t} \leq y - y_{0t-1}\} \qquad (7) \\
&\qquad \times \quad \varphi_{t-1}\left\{F^{-1}_{\Delta Y_{0t-1}|D_t=1}(F_{\Delta Y_{0t}|D_t=1}(\Delta y_{0t})), F^{-1}_{Y_{0t-2}|D_t=1}(F_{Y_{0t-1}|D_t=1}(y_{0t-1}))\right\} \\
&\qquad \times \quad \frac{f_{\Delta Y_{0t}|D_t=1}(\Delta y_{0t})}{f_{\Delta Y_{0t-1}|D_t=1}(F^{-1}_{\Delta Y_{0t-1}|D_t=1}(F_{\Delta Y_{0t}|D_t=1}(\Delta y_{0t})))} \\
&\qquad \times \quad \frac{f_{Y_{0t-1}|D_t=1}(y_{0t-1})}{f_{Y_{0t-2}|D_t=1}(F^{-1}_{Y_{0t-2}|D_t=1}(F_{Y_{0t-1}|D_t=1}(y_{0t-1})))} \mathrm{d}\Delta y_{0t} \mathrm{d}y_{0t-1}
\end{aligned}
$$

Equation 5 rewrites the joint distribution in terms of the copula pdf using Lemma B.1; Equation 6 uses the copula stability assumption; Equation 7 rewrites the copula pdf as the joint distribution (now in period $t-1$) using Lemma B.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 $v = F^{-1}_{Y_{0t-2}|D_t=1}(F_{Y_{0t-1}|D_t=1}(y_{0t-}$ 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 7 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 7 implies

$$\text{Equation 7} = \int_{\mathcal{Y}_{t-2|D_t=1}} \int_{\Delta\mathcal{Y}_{t-1|D_t=1}} \mathbb{1}\{F^{-1}_{\Delta Y_{0t}|D_t=1}(F_{\Delta Y_{0t-1}|D_t=1}(u)) \leq y - F^{-1}_{Y_{0t-1}|D_t=1}(F_{Y_{0t-2}|D_t=1}(v))\} \tag{8}$$

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

$$= \mathrm{E}\left[\mathbb{1}\{F^{-1}_{\Delta Y_{0t}|D_t=1}(F_{\Delta Y_{0t-1}|D_t=1}(\Delta Y_{0t-1})) \leq y - F^{-1}_{Y_{0t-1}|D_t=1}(F_{Y_{0t-2}|D_t=1}(Y_{0t-2}))\}|D_t = 1\right] \tag{9}$$

$$= \mathrm{E}\left[\mathbb{1}\{F^{-1}_{\Delta Y_{0t}|D_t=0}(F_{\Delta Y_{0t-1}|D_t=1}(\Delta Y_{0t-1})) \leq y - F^{-1}_{Y_{0t-1}|D_t=1}(F_{Y_{0t-2}|D_t=1}(Y_{0t-2}))\}|D_t = 1\right] \tag{10}$$

where Equation 8 follows from the discussion above, Equation 9 follows by the definition of expectation, and Equation 10 follows from the Distributional Difference in Differences Assumption. □

B.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 4 holds. Notice,

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

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

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

$$= \mathrm{E}\left[\frac{p(X)}{p}\mathrm{P}(\Delta Y_{0t} \leq \Delta y|X, D_t = 0)\right] \tag{11}$$

$$= \mathrm{E}\left[\frac{p(X)}{p}\mathrm{E}[(1-D_t)\mathbb{1}\{\Delta Y_t \leq \Delta y)\}|X, D_t = 0]\right] \tag{12}$$

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

$$= \mathrm{E}\left[\frac{1-D_t}{1-p(X)}\frac{p(X)}{p}\mathbb{1}\{\Delta Y_t \leq \Delta y\}\right] \tag{13}$$

where Equation 11 holds by Conditional Distributional Difference in Differences Assumption. Equation 12 holds by replacing $P(\cdot)$ with $\mathrm{E}(\mathbb{1}\{\cdot\})$ 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 13 simply applies the Law of Iterated Expectations to conclude the proof. □

B.2 Proof of the results in Example 1

The nonseparable model $Y_{it} = q(U_{it}, X_i, D_{it}) + C_i$ can be equivalently written in terms of potential outcomes:

$$\begin{aligned} Y_{1it} &= q_1(U_{it}, X_i) + C_i \\ Y_{0it} &= q_0(U_{it}, X_i) + C_i \end{aligned}$$

Unconditional Mean Difference in Differences Holds

$$\begin{aligned} \mathrm{E}[Y_{0t}|D = d] &= \int q_0(u, x) + c \ \mathrm{dF}_{U_t, X, C|D=d}(u, x, c) \\ &= \int q_0(u, x) + c \ \mathrm{dF}_{U_t} \ \mathrm{d}F_{X,C|D=d}(u, x, c) \\ &= \int q_0(u, x) + c \ \mathrm{dF}_{U_{t-1}} \ \mathrm{d}F_{X,C|D=d}(u, x, c) \\ &= \int q_0(u, x) + c \ \mathrm{dF}_{U_{t-1}, X, C|D=d}(u, x, c) \\ &= \mathrm{E}[Y_{0t-1}|D = d] \end{aligned}$$

which implies that for the treated group and untreated group the average change in untreated potential outcomes is 0.

Conditional Difference in Differences Holds

$$\begin{aligned} \mathrm{P}(\Delta Y_{0t} \leq \Delta | X = x, D = 1) &= \int \mathbb{1}\{q_0(u, x) - q_0(\tilde{u}, x) \leq \Delta\} \ \mathrm{d}F_{U_t, U_{t-1}|X,D=1}(u, \tilde{u}) \\ &= \int \mathbb{1}\{q_0(u, x) - q_0(\tilde{u}, x) \leq \Delta\} \ \mathrm{d}F_{U_t, U_{t-1}|X,D=0}(u, \tilde{u}) \\ &= \mathrm{P}(\Delta Y_{0t} \leq \Delta | X = x, D = 0) \end{aligned}$$

where the second equality holds because $(U_t, U_{t-1}) \perp\!\!\!\perp (X, D)$.

Unconditional Distributional Difference in Differences Does Not Hold

$$\begin{aligned} \mathrm{P}(\Delta Y_{0t} \leq \Delta | D = 1) &= \mathrm{E}[\mathrm{P}(\Delta Y_{0t} \leq \Delta | X, D = 1) | D = 1] \\ &= \mathrm{E}[\mathrm{P}(\Delta Y_{0t} \leq \Delta | X, D = 0) | D = 1] \end{aligned}$$

where the second equality holds by the result for the Conditional Distributional Difference in Differences Assumption holding. The last quantity is, in general, not equal to $\mathrm{P}(\Delta Y_{0t} \leq \Delta | D = 0)$ because the distribution of X can be different across the two groups.

Unconditional Copula Stability Holds

$$\begin{aligned}\mathrm{P}(\Delta Y_{0t} \leq \Delta, Y_{0t-1} \leq y | D = 1) &= \mathrm{P}(q_0(U_{it}, X_i) - q_0(U_{it-1}, X_i) \leq \Delta, q_0(U_{it-1}, X_i) \leq y | D = 1) \\ &= \mathrm{P}(q_0(U_{it-1}, X_i) - q_0(U_{it-2}, X_i) \leq \Delta, q_0(U_{it-2}, X_i) \leq y | D = 1) \\ &= \mathrm{P}(\Delta Y_{0t-1} \leq \Delta, Y_{0t-2} \leq y | D = 1)\end{aligned}$$

which implies that the CSA holds.

B.3 Asymptotic Normality

In this section, we derive the asymptotic distribution of our estimator of the QTT. First, we introduce some notation. First, to conserve on notation, let $\mathrm{F}_{\Delta t} = \mathrm{F}_{\Delta Y_t | D_t = 0}$, $\mathrm{F}_{\Delta t-1} = \mathrm{F}_{\Delta Y_{t-1} | D_t = 1}$, $\mathrm{F}_{Y_{t-1}} = \mathrm{F}_{Y_{t-1} | D_t = 1}$, and $\mathrm{F}_{Y_{t-2}} = \mathrm{F}_{Y_{t-2} | D_t = 1}$. Let

$$\phi_n(F) = \frac{1}{n_T} \sum_{i \in \mathcal{T}} \mathbb{1}\{\mathrm{F}^{-1}_{\Delta t}(\mathrm{F}_{\Delta t-1}(\Delta Y_{it-1})) \leq y - \mathrm{F}^{-1}_{Y_{t-1}}(\mathrm{F}_{Y_{t-2}}(Y_{it-2}))\}$$

and

$$\phi_0(F) = \mathrm{E}\left[\mathbb{1}\{\mathrm{F}^{-1}_{\Delta t}(\mathrm{F}_{\Delta t-1}(\Delta Y_{t-1})) \leq y - \mathrm{F}^{-1}_{Y_{t-1}}(\mathrm{F}_{Y_{t-2}}(Y_{t-2}))\} \middle| D_t = 1\right]$$

Let $F_0 = (F_{10}, F_{20}, F_{30}, F_{40})$ where F_{j0}, for $j = 1, \ldots, 4$, are distribution functions; we assume that F_{10} and F_{20} have common, compact support $\mathcal{U} \subset \mathbb{R}$ and that F_{30} and F_{40} have common, compact support $\mathcal{V} \subset \mathbb{R}$. We also suppose that each F_{j0} has a density function f_{j0} that are uniformly bounded away from 0 and ∞ on their supports. Let (U_2, V_4) be two random variables on $\mathcal{U} \times \mathcal{V}$ with joint distribution F_{U_2,V_4}. We assume that $U_2 \sim F_{20}$ and that $V_4 \sim F_{40}$ and that the conditional distribution $F_{U_2|V_4}$ has a continuous density function $f_{U_2|V_4}$ that is uniformly bounded from 0 and ∞. As a first step, we establish the Hadamard Differentiability of $\phi_0(F)$. We do this in several steps. First, we use the following result due to Callaway, Li, and Oka, (2016)

Lemma B.3. *Let $\mathbb{D} = C(\mathcal{V})^2$ and define the map $\Psi : \mathbb{D}_\Psi \subset \mathbb{D} \mapsto l^\infty(\mathcal{V})$ as*

$$\Psi(F) \equiv F_3^{-1} \circ F_4$$

where $\mathbb{D}_\Psi \equiv \mathbb{E} \times \mathbb{E}$ where $\mathbb{E}$ is the set of all distribution functions with strictly positive, bounded densities. Then, the map Ψ is Hadamard Differentiable at (F_{30}, F_{40}) tangentially to $\mathbb{D}$ with derivative at (F_{30}, F_{40}) in $\psi \equiv (\psi_1, \psi_2) \in \mathbb{D}$

$$\Psi'_{(F_{30},F_{40})}(\psi) = \frac{\gamma_2 - \gamma_1 \circ F_{30}^{-1} \circ F_{40}}{f_{30} \circ F_{30}^{-1} \circ F_{40}}$$

Lemma B.4. *Let $\mathbb{A} = C(\mathcal{U}) \times l^{\infty}(\mathcal{V})$. Define the map $\Lambda : \mathbb{A}_{\Lambda} \mapsto \mathbb{E}$ with $\mathbb{A}_{\Lambda} \equiv \mathbb{E} \times \mathbb{D}_{\Psi}$ where $\mathbb{D}_{\Psi}$ is given in Lemma B.3, given by*

$$\Lambda(\Gamma)(y) = \Gamma_1(y - \Gamma_2)$$

Then, the map Λ is Hadamard differentiable at $(F_{10}, F_{30}^{-1} \circ F_{40})$ tangentially to $\mathbb{A}$ with derivative in $\alpha \equiv (\alpha_1, \alpha_2) \in \mathbb{A}$ given by

$$\Lambda'_{(F_{10}, F_{30}^{-1} \circ F_{40})}(\alpha)(y) = \alpha_1 \circ F_{30}^{-1} \circ F_{40} + F_{10}(y - \alpha_2)$$

Proof. Let $\Lambda_1 : \mathbb{A}_{\Lambda} \mapsto \mathbb{A}_{\Lambda}$ given by $\Lambda_1(\Xi) = (\Xi_1, \cdot - \Xi_2)$. Lemma 3.9.25 of Van Der Vaart and Wellner, (1996) implies that the map Λ_1 is Hadamard differentiable at Ξ tangentially to $\mathbb{A}$ with derivative in $\xi = (\xi_1, \xi_2) \in \mathbb{A}$ given by

$$\Lambda'_{1,\Xi}(\xi) = (\xi_1, -\xi_2)$$

Let $\Lambda_2 : \mathbb{A}_{\Lambda} \mapsto \mathbb{E}$ given by $\Lambda_2(\Upsilon) = \Upsilon_1 \circ \Upsilon_2$. Lemma 3.9.27 of Van Der Vaart and Wellner, (1996) implies that Λ_2 is Hadamard differentiable at Υ tangentially to $\mathbb{A}$ with derivative at Υ in $\upsilon = (\upsilon_1, \upsilon_2) \in \mathbb{A}$ given by

$$\Lambda'_{2,\Upsilon}(\upsilon) = \upsilon_1 \circ \Upsilon_2 + \Upsilon'_{1,\Upsilon_2} \circ \upsilon_2$$

By the chain rule for Hadamard differentiable maps

$$\Lambda'_{(F_{10}, F_{30}^{-1} \circ F_{40})}(\alpha) = \Lambda'_{2,(F_{10}, F_{30}^{-1} \circ F_{40})} \circ \Lambda'_{1,(F_{10}, F_{30}^{-1} \circ F_{40})}(\alpha)$$

for $\alpha \in \mathbb{A}$. □

Lemma B.5. *Let $\mathbb{B} = C(\mathcal{U})^2$. Define the map $\Phi : \mathbb{B}_{\Phi} \subset \mathbb{B} \mapsto l^{\infty}(U)$ with $\mathbb{D}_{\Phi} := \mathbb{E} \times \mathbb{D}_{\Lambda}$ given by*

$$\Phi(\Omega) = \Omega_1^{-1} \circ \Omega_2$$

Then, the map Φ is Hadamard differentiable at $(F_{20}, F_{10}(\cdot - F_{30}^{-1} \circ F_{40}))$ tangentially to $\mathbb{B}$ with derivative at $(F_{20}, F_{10}(\cdot - F_{30}^{-1} \circ F_{40}))$ in $\omega := (\omega_1, \omega_2) \in \mathbb{B}$ given by

$$\Phi'_{(F_{20}, F_{10}(\cdot - F_{30}^{-1} \circ F_{40}))}(\omega) = \frac{\omega_2 - \omega_1 \circ F_{20}^{-1} \circ F_{10} \circ (\cdot - F_{30}^{-1} \circ F_{40})}{f_{20} \circ F_{20}^{-1} \circ F_{10} \circ (\cdot - F_{30}^{-1} \circ F_{40})}$$

Proof. The proof follows by the same argument as in Lemma B.3. □

Lemma B.6. *Let $\mathbb{D} = C(\mathcal{U})^2 \times C(\mathcal{V})^2$ and let $\mathcal{Y}$ be a compact subset of $\mathbb{R}$. Let $\phi : \mathbb{D}_{\phi} \subset \mathbb{D} \mapsto l^{\infty}(\mathcal{Y})$ be given by*

$$\phi(F)(y) = \mathrm{P}(F_1^{-1}(F_2(V_2)) + F_3^{-1}(F_4(V_4)) \leq y)$$

for $F = (F_1, F_2, F_3, F_4) \in \mathbb{D}_{\phi}$ where $\mathbb{D}_{\phi} = \mathbb{E}^4$ where $\mathbb{E}$ is the set of all distribution functions with strictly positive and bounded densities. Then, the map ϕ is Hadamard Differentiable at

F_0 tangentially to $\mathbb{D}$ with derivative in $\gamma = (\gamma_1, \gamma_2, \gamma_3, \gamma_4) \in \mathbb{D}$ given by

$$\phi'_{F_0}(\gamma)(y) = \pi'_{F_{20}^{-1} \circ F_{10}(y - F_{30}^{-1} \circ F_{40})} \circ \Phi'_{(F_{20}, F_{10}(y - F_{30}^{-1} \circ F_{40}))}(\gamma_2, \Lambda'_{(F_{10}, F_{30}^{-1} \circ F_{40})}(\gamma_1, \Psi'_{(F_{30}, F_{40})}(\gamma_3, \gamma_4))$$

Proof. First, notice that

$$\begin{aligned}\phi(F)(y) &= \mathrm{P}(V_2 \leq F_2^{-1} \circ F_1(y - F_3^{-1} \circ F_4(V_4))) \\ &= \mathrm{P}(V_2 \leq \Phi(F_2, \Lambda(F_1, \Psi(F_3, F_4)(V_4)(y))))\end{aligned}$$

Define the map $\pi : \mathbb{D}_\pi \mapsto l^\infty(\mathcal{Y})$ where $\mathbb{D}_\pi$ is the set of all functions $F_2^{-1}(F_1(\cdot - F_3^{-1}(F_4)))$ for $(F_1, F_2^{-1}, F_3^{-1}, F_4) \in \mathbb{E} \times \mathbb{E}^- \times \mathbb{E}^- \times \mathbb{E}$ as

$$\pi(\chi)(y) = \int F_{V_2|V_4}(\chi(v_4)(y)|v_4) \, \mathrm{d}F_{V_4}(v_4)$$

Then, for $F \in \mathbb{D}$ and $y \in \mathcal{Y}$, $\phi = \pi \circ \Phi \circ \Lambda \circ \Psi$

Using the same arguments as in Callaway, Li, and Oka, (2016, Lemma A2), π is Hadamard differentiable at $\chi \in \mathbb{D}_\pi$ tangentially to $\mathbb{D}$ with derivative at χ in $\zeta \in \mathbb{D}$ given by

$$\pi'_\chi(\zeta)(y) = \int \zeta(v_4) f_{V_2|V_4}(\chi(v_4)|v_4) \, \mathrm{d}F_{V_4}(v_4) \tag{14}$$

By the chain rule for Hadamard differentiable functions (cf. Van Der Vaart and Wellner, (1996, Lemma 3.9.3)),

$$\phi'_{F_0}(\gamma) = \pi'_{F_{20}^{-1} \circ F_{10}(\cdot - F_{30}^{-1} \circ F_{40})} \circ \Phi'_{(F_{20}, F_{10}(\cdot - F_{30}^{-1} \circ F_{40}))}(\gamma_2, \Lambda'_{(F_{10}, F_{30}^{-1} \circ F_{40})}(\gamma_1, \Psi'_{(F_{30}, F_{40})}(\gamma_3, \gamma_4))$$

Plugging in the results from Lemmas B.3 to B.5 and Equation (14) implies

$$\phi'_{F_0}(\gamma) = \int \frac{\gamma_1 \circ F_{30}^{-1} \circ F_{40}(v_4) - F_{10}\left(\cdot - \frac{\gamma_4 - \gamma_3 \circ F_{30}^{-1} \circ F_{40}(v_4)}{f_{30} \circ F_{30}^{-1} \circ F_{40}(v_4)}\right) - \gamma_2 \circ F_{20}^{-1} \circ F_{10}(y - F_{30}^{-1} \circ F_{40}(v_4))}{f_{20} \circ F_{20}^{-1} \circ F_{10} \circ (y - F_{30}^{-1} \circ F_{40}(v_4))}$$

$$\times f_{V_2|V_4}(F_{20}^{-1} \circ F_{10}(\cdot - F_{30}^{-1} \circ F_{40}(v_4)) \, \mathrm{d}F_{V_4}(v_4)$$

□

Next, let

$$v_n(F) = \sqrt{n}(\phi_n(F) - \phi_0(F))$$

Lemma B.7.

$$\sup_{y \in \mathcal{Y}} |v_n(\hat{F})(y) - v_n(F_0)(y)| \xrightarrow{p} 0$$

Proof. Because $\mathcal{Y}$ is a compact set, we can show that $|v_n(\hat{F})(y) - v_n(\hat{F})(y)| \xrightarrow{p} 0$ for all

$y \in \mathcal{Y}$. Notice that, for any $y \in \mathcal{Y}$,

$$v_n(\hat{\mathrm{F}})(y) - v_n(\mathrm{F}_0)(y) = \sqrt{n}(\phi_n(\hat{\mathrm{F}}_{\Delta t}, \hat{\mathrm{F}}_{\Delta t-1}, \hat{\mathrm{F}}_{Y_{t-1}}, \hat{\mathrm{F}}_{Y_{t-2}})(y) - \phi_0(\mathrm{F}_{\Delta t}, \mathrm{F}_{\Delta t-1}, \mathrm{F}_{Y_{t-1}}, \mathrm{F}_{Y_{t-2}})(y)) \\ - \sqrt{n}(\phi_n(\mathrm{F}_{\Delta t}, \mathrm{F}_{\Delta t-1}, \mathrm{F}_{Y_{t-1}}, \mathrm{F}_{Y_{t-2}})(y) - \phi_0(\mathrm{F}_{\Delta t}, \mathrm{F}_{\Delta t-1}, \mathrm{F}_{Y_{t-1}}, \mathrm{F}_{Y_{t-2}})(y))$$

Then, adding and subtracting the following terms:

$$\phi_n(F_{\Delta t}, \hat{\mathrm{F}}_{\Delta t-1}, \hat{\mathrm{F}}_{Y_{t-1}}, \hat{\mathrm{F}}_{Y_{t-2}})(y) - \phi_0(F_{\Delta t}, \hat{\mathrm{F}}_{\Delta t-1}, \hat{\mathrm{F}}_{Y_{t-1}}, \hat{\mathrm{F}}_{Y_{t-2}})(y)$$
$$\phi_n(F_{\Delta t}, \mathrm{F}_{\Delta t-1}, \hat{\mathrm{F}}_{Y_{t-1}}, \hat{\mathrm{F}}_{Y_{t-2}})(y) - \phi_0(F_{\Delta t}, \mathrm{F}_{\Delta t-1}, \hat{\mathrm{F}}_{Y_{t-1}}, \hat{\mathrm{F}}_{Y_{t-2}})(y)$$
$$\phi_n(F_{\Delta t}, \mathrm{F}_{\Delta t-1}, \mathrm{F}_{Y_{t-1}}, \hat{\mathrm{F}}_{Y_{t-2}})(y) - \phi_0(F_{\Delta t}, \mathrm{F}_{\Delta t-1}, \mathrm{F}_{Y_{t-1}}, \hat{\mathrm{F}}_{Y_{t-2}})(y)$$

implies

$$\begin{aligned}
&v_n(\hat{\mathrm{F}})(y) - v_n(\mathrm{F}_0)(y) \\
&\quad = \sqrt{n}\Big\{\phi_n(\hat{\mathrm{F}}_{\Delta t}, \hat{\mathrm{F}}_{\Delta t-1}, \hat{\mathrm{F}}_{Y_{t-1}}, \hat{\mathrm{F}}_{Y_{t-2}})(y) - \phi_n(\mathrm{F}_{\Delta t}, \hat{\mathrm{F}}_{\Delta t-1}, \hat{\mathrm{F}}_{Y_{t-1}}, \hat{\mathrm{F}}_{Y_{t-2}})(y) \\
&\qquad\quad - \Big(\phi_0(\hat{\mathrm{F}}_{\Delta t}, \hat{\mathrm{F}}_{\Delta t-1}, \hat{\mathrm{F}}_{Y_{t-1}}, \hat{\mathrm{F}}_{Y_{t-2}})(y) - \phi_0(\mathrm{F}_{\Delta t}, \hat{\mathrm{F}}_{\Delta t-1}, \hat{\mathrm{F}}_{Y_{t-1}}, \hat{\mathrm{F}}_{Y_{t-2}})(y)\Big)\Big\} \quad (15) \\
&\quad + \sqrt{n}\Big\{\phi_n(\mathrm{F}_{\Delta t}, \hat{\mathrm{F}}_{\Delta t-1}, \hat{\mathrm{F}}_{Y_{t-1}}, \hat{\mathrm{F}}_{Y_{t-2}})(y) - \phi_n(\mathrm{F}_{\Delta t}, \mathrm{F}_{\Delta t-1}, \hat{\mathrm{F}}_{Y_{t-1}}, \hat{\mathrm{F}}_{Y_{t-2}})(y) \\
&\qquad\quad - \Big(\phi_0(\mathrm{F}_{\Delta t}, \hat{\mathrm{F}}_{\Delta t-1}, \hat{\mathrm{F}}_{Y_{t-1}}, \hat{\mathrm{F}}_{Y_{t-2}})(y) - \phi_0(\mathrm{F}_{\Delta t}, \mathrm{F}_{\Delta t-1}, \hat{\mathrm{F}}_{Y_{t-1}}, \hat{\mathrm{F}}_{Y_{t-2}})(y)\Big)\Big\} \quad (16) \\
&\quad + \sqrt{n}\Big\{\phi_n(\mathrm{F}_{\Delta t}, \mathrm{F}_{\Delta t-1}, \hat{\mathrm{F}}_{Y_{t-1}}, \hat{\mathrm{F}}_{Y_{t-2}})(y) - \phi_n(\mathrm{F}_{\Delta t}, \mathrm{F}_{\Delta t-1}, \mathrm{F}_{Y_{t-1}}, \hat{\mathrm{F}}_{Y_{t-2}})(y) \\
&\qquad\quad - \Big(\phi_0(\mathrm{F}_{\Delta t}, \mathrm{F}_{\Delta t-1}, \hat{\mathrm{F}}_{Y_{t-1}}, \hat{\mathrm{F}}_{Y_{t-2}})(y) - \phi_0(\mathrm{F}_{\Delta t}, \mathrm{F}_{\Delta t-1}, \mathrm{F}_{Y_{t-1}}, \hat{\mathrm{F}}_{Y_{t-2}})(y)\Big)\Big\} \quad (17) \\
&\quad + \sqrt{n}\Big\{\phi_n(\mathrm{F}_{\Delta t}, \mathrm{F}_{\Delta t-1}, \mathrm{F}_{Y_{t-1}}, \hat{\mathrm{F}}_{Y_{t-2}})(y) - \phi_n(\mathrm{F}_{\Delta t}, \mathrm{F}_{\Delta t-1}, \mathrm{F}_{Y_{t-1}}, \mathrm{F}_{Y_{t-2}})(y) \\
&\qquad\quad - \Big(\phi_0(\mathrm{F}_{\Delta t}, \mathrm{F}_{\Delta t-1}, \mathrm{F}_{Y_{t-1}}, \hat{\mathrm{F}}_{Y_{t-2}})(y) - \phi_0(\mathrm{F}_{\Delta t}, \mathrm{F}_{\Delta t-1}, \mathrm{F}_{Y_{t-1}}, \mathrm{F}_{Y_{t-2}})(y)\Big)\Big\} \quad (18)
\end{aligned}$$

Each of the above terms converges to 0. We show below that this holds for Equation 15 while omitting the proof for the other terms – the arguments are essentially identical for each one.

Proof.

$$\begin{aligned}
&\sqrt{n}\Big\{\phi_n(\hat{\mathrm{F}}_{\Delta t}, \hat{\mathrm{F}}_{\Delta t-1}, \hat{\mathrm{F}}_{Y_{t-1}}, \hat{\mathrm{F}}_{Y_{t-2}})(y) - \phi_n(\mathrm{F}_{\Delta t}, \hat{\mathrm{F}}_{\Delta t-1}, \hat{\mathrm{F}}_{Y_{t-1}}, \hat{\mathrm{F}}_{Y_{t-2}})(y) \\
&\qquad - \Big(\phi_0(\hat{\mathrm{F}}_{\Delta t}, \hat{\mathrm{F}}_{\Delta t-1}, \hat{\mathrm{F}}_{Y_{t-1}}, \hat{\mathrm{F}}_{Y_{t-2}})(y) - \phi_0(\mathrm{F}_{\Delta t}, \hat{\mathrm{F}}_{\Delta t-1}, \hat{\mathrm{F}}_{Y_{t-1}}, \hat{\mathrm{F}}_{Y_{t-2}})(y)\Big)\Big\} \\
&\quad = \sqrt{n}\left\{\left(\frac{1}{n}\sum_{i=1}^{n} \mathbb{1}\{\hat{F}_1^{-1}(\hat{F}_2(V_{1i})) \le y - \hat{F}_3^{-1}(\hat{F}_4(V_{2i}))\} - \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}\{F_1^{-1}(\hat{F}_2(V_1)) \le y - \hat{F}_3^{-1}(\hat{F}_4(V_{2i}))\}\right)\right. \\
&\qquad \left. - \Big(\mathrm{E}\left[\mathbb{1}\{\hat{F}_1^{-1}(\hat{F}_2(V_1)) \le y - \hat{F}_3^{-1}(\hat{F}_4(V_2))\}\right] - \mathrm{E}\left[\mathbb{1}\{F_1^{-1}(\hat{F}_2(V_1)) \le y - \hat{F}_3^{-1}(\hat{F}_4(V_2))\}\right]\Big)\right\}
\end{aligned}$$

□

To show the result for Equation 15, Lemmas B.8.A, B.8.B and B.17 show that, for any $y \in \mathcal{Y}$, $\sqrt{n}(\phi_n(\hat{F}_1, \hat{F}_2, \hat{F}_3, \hat{F}_4) - \phi_n(F_1, \hat{F}_2, \hat{F}_3, \hat{F}_4))$ and $\sqrt{n}(\phi_0(\hat{F}_1, \hat{F}_2, \hat{F}_3, \hat{F}_4) - \phi_0(F_1, \hat{F}_2, \hat{F}_3, \hat{F}_4))$ are asymptotically equivalent which implies the result. □

Lemma B.8.A. *Let* $\hat{\mu}(y) = \frac{1}{n_T}\sum_{i\in\mathcal{T}} \hat{\mathrm{F}}_{\Delta t}(y - \mathrm{F}^{-1}_{Y_{t-1}}(\mathrm{F}_{Y_{t-2}}(Y_{it-2}))) - \mathrm{F}_{\Delta t}(y - \mathrm{F}^{-1}_{Y_{t-1}}(\mathrm{F}_{Y_{t-2}}(Y_{it-2})))$. *Then, for all* $y \in \mathcal{Y}$

$$\sqrt{n}\left(\phi_n(\hat{F}_1, \hat{F}_2, \hat{F}_3, \hat{F}_4)(y) - \phi_n(F_1, \hat{F}_2, \hat{F}_3, \hat{F}_4)(y) - \hat{\mu}(y)\right) = o_p(1)$$

Proof.

$$\begin{aligned}
&\sqrt{n}\left(\phi_n(\hat{F}_1, \hat{F}_2, \hat{F}_3, \hat{F}_4)(y) - \phi_n(F_1, \hat{F}_2, \hat{F}_3, \hat{F}_4)(y) - \hat{\mu}(y)\right) \\
&\quad= \sqrt{n}\left\{\frac{1}{n}\sum_{i=1}^{n}\Big[\mathbb{1}\{\hat{F}_1^{-1}(\hat{F}_2(V_{1i})) \le y - \hat{F}_3^{-1}(\hat{F}_4(V_{2i}))\} - \mathbb{1}\{F_1^{-1}(\hat{F}_2(V_{1i})) \le y - \hat{F}_3^{-1}(\hat{F}_4(V_{2i}))\} \right. \\
&\qquad\qquad \left. - \left(\hat{F}_1(y - F_3^{-1}(F_4(V_{2i})))] - F_1(y - F_3^{-1}(F_4(V_{2i})))\right)\Big]\right\} \\
&\quad\le \sup_{v\in\mathcal{V}_2}\sqrt{n}\left|\frac{1}{n}\sum_{i=1}^{n}\Big[\mathbb{1}\{\hat{F}_1^{-1}(\hat{F}_2(V_{1i})) \le y - \hat{F}_3^{-1}(\hat{F}_4(v_2))\} - \mathbb{1}\{F_1^{-1}(\hat{F}_2(V_{1i})) \le y - \hat{F}_3^{-1}(\hat{F}_4(v_2))\} \right. \\
&\qquad\qquad \left. - \left(\hat{F}_1(y - F_3^{-1}(F_4(v_2)))] - F_1(y - F_3^{-1}(F_4(v_2)))\right)\Big]\right| \\
&\quad= \sup_{v\in\mathcal{V}_2}\sqrt{n}\left|\frac{1}{n}\sum_{i=1}^{n}\Big[\mathbb{1}\{V_{1i} \le \hat{F}_2^{-1}(\hat{F}_1(y - \hat{F}_3^{-1}(\hat{F}_4(v_2))))\} - \mathbb{1}\{V_{1i} \le \hat{F}_2^{-1}(F_1(y - \hat{F}_3^{-1}(\hat{F}_4(v_2))))\} \right. \\
&\qquad\qquad \left. - \left(\hat{F}_1(y - F_3^{-1}(F_4(v_2)))] - F_1(y - F_3^{-1}(F_4(v_2)))\right)\Big]\right| + o_p(1) \\
&\quad= \sup_{v\in\mathcal{V}_2}\sqrt{n}\left|\hat{F}_1(y - \hat{F}_3^{-1}(\hat{F}_4(v_2))) - F_1(y - \hat{F}_3^{-1}(\hat{F}_4(v_2))) \right. \\
&\qquad\qquad \left. - \left(\hat{F}_1(y - F_3^{-1}(F_4(v_2))) - F_1(y - F_3^{-1}(F_4(v_2)))\right)\right| + o_p(1) \\
&\quad= o_p(1)
\end{aligned}$$

□

Lemma B.8.B. *Let* $\mu(y) = \mathrm{E}[\hat{F}_1(y - F_3^{-1}(F_4(V_2)))] - \mathrm{E}[F_1(y - F_3^{-1}(F_4(V_2)))]$. *Then, for all* $y \in \mathcal{Y}$,

$$\sqrt{n}\left(\phi_0(\hat{F}_1, \hat{F}_2, \hat{F}_3, \hat{F}_4)(y) - \phi_0(F_1, \hat{F}_2, \hat{F}_3, \hat{F}_4)(y) - \mu(y)\right) = o_p(1)$$

Proof.

$$
\begin{aligned}
&\sqrt{n}\left(\phi_0(\hat{F}_1,\hat{F}_2,\hat{F}_3,\hat{F}_4)(y)-\phi_0(F_1,\hat{F}_2,\hat{F}_3,\hat{F}_4)(y)-\mu(y)\right)\\
&\quad=\sqrt{n}\left\{\mathrm{E}\left[\mathbb{1}\{\hat{F}_1^{-1}(\hat{F}_2(V_1))\le y-\hat{F}_3^{-1}(\hat{F}_4(V_2))\}\right]-\mathrm{E}\left[\mathbb{1}\{F_1^{-1}(\hat{F}_2(V_1))\le y-\hat{F}_3^{-1}(\hat{F}_4(V_2))\}\right]\right.\\
&\qquad\left.-\left(\mathrm{E}[\hat{F}_1(y-F_3^{-1}(F_4(V_2)))]-\mathrm{E}[F_1(y-F_3^{-1}(F_4(V_2)))]\right)\right\}\\
&\quad=\sqrt{n}\left\{\mathrm{E}\left[\mathbb{1}\{V_1\le\hat{F}_2^{-1}(\hat{F}_1(y-\hat{F}_3^{-1}(\hat{F}_4(V_2))))\}\right]-\mathrm{E}\left[\mathbb{1}\{V_1\le\hat{F}_2^{-1}(F_1(y-\hat{F}_3^{-1}(\hat{F}_4(V_2))))\}\right]\right.\\
&\qquad\left.-\left(\mathrm{E}[\hat{F}_1(y-F_3^{-1}(F_4(V_2)))]-\mathrm{E}[F_1(y-F_3^{-1}(F_4(V_2)))]\right)\right\}+o_p(1)\\
&\quad\le\sup_{v_2\in\mathcal{V}_2}|F_2(\hat{F}_2^{-1}(\hat{F}_1(y-\hat{F}_3^{-1}(\hat{F}_4(v_2)))))-F_2(\hat{F}_2^{-1}(\hat{F}_1(y-\hat{F}_3^{-1}(\hat{F}_4(v_2)))))\\
&\qquad-\left(\hat{F}_1(y-F_3^{-1}(F_4(v_2)))-F_1(y-F_3^{-1}(F_4(v_2)))\right)+o_p(1)\\
&\quad=\sup_{v_2\in\mathcal{V}_2}|\hat{F}_1(y-\hat{F}_3^{-1}(\hat{F}_4(v_2)))-\hat{F}_1(y-\hat{F}_3^{-1}(\hat{F}_4(v_2)))\\
&\qquad-\left(\hat{F}_1(y-F_3^{-1}(F_4(v_2)))-F_1(y-F_3^{-1}(F_4(v_2)))\right)+o_p(1)\\
&\quad=o_p(1)
\end{aligned}
$$

□

Proof of Proposition 2 First, notice that

$$
\begin{aligned}
\sqrt{n}(\hat{\mathrm{F}}_{Y_{0t}|D=1}(y)-\mathrm{F}_{Y_{0t}|D=1}(y))&=\sqrt{n}(\phi_n(\hat{F})-\phi_0(F_0))\\
&=\sqrt{n}(\phi_n(\hat{F})-\phi_0(\hat{F}))-\sqrt{n}(\phi_0(\hat{F})-\phi_0(F_0))\\
&=\sqrt{n}(\phi_n(F_0)-\phi_0(F_0))-\phi'_{F_0}\sqrt{n}(\hat{F}-F_0)+o_p(1)
\end{aligned}
$$

where the last equality holds by Lemmas B.6 and B.7. Then, the result holds by Proposition 1 and an application of the functional central limit theorem.

Proof of Theorem 3 Under the conditions stated in Theorem 3, the result follows from the Hadamard differentiability of the quantile map (Van Der Vaart and Wellner, 1996, Lemma 3.9.23(ii)) and by Proposition 2.

Proof of Theorem 4 The result holds because our estimate of the QTT is Donsker and by Theorem 3.6.1 in Van Der Vaart and Wellner, (1996).

Asymptotic Normality of propensity score reweighted estimator Let $F_0=(\tilde{F}_{\Delta Y_{0t}|D=1}, F_{\Delta Y_{0t}|D=1}, F_{Y_{0t-1}|D=1}, F_{Y_{0t-2}|D=1})$ and $\hat{F}=(\hat{\tilde{F}}_{\Delta Y_{0t}|D=1}, \hat{F}_{\Delta Y_{0t}|D=1}, \hat{F}_{Y_{0t-1}|D=1}, \hat{F}_{Y_{0t-2}|D=1})$.

For $W=(D,X,\Delta Y)$, let

$$
\varphi(W,\Delta)=\frac{\mathbb{1}\{\Delta Y\le\Delta|X\}}{p(1-p_0(X))}(D-p_0(X))+\frac{1-D}{p}\frac{p_0(X)}{1-p_0(X)}\mathbb{1}\{\Delta Y_t\le\Delta\}
$$

Lemma B.9. *Let $\mathcal{K} = \{\varphi(W, \Delta) | \Delta \in \Delta\mathcal{Y}\}$. $\mathcal{K}$ is a Donsker class.*

Proof. Let $\mathcal{K}_1 = \{\frac{\mathbb{1}\{\Delta Y \leq \Delta | X\}}{p(1-p_0(X))}(D - p_0(X)) | \Delta \in \Delta\mathcal{Y}\}$. $\mathcal{K}_1$ is Donsker by Donald and Hsu, (2014, Lemma A.2). Let $\mathcal{K}_2 = \{\frac{1-D}{p}\frac{p_0(X)}{1-p_0(X)}\mathbb{1}\{\Delta Y_t \leq \Delta\} | \Delta \in \Delta\mathcal{Y}\}$. $\mathcal{K}_2$ is Donsker because $\mathbb{1}\{\Delta Y_t \leq \Delta\} | \Delta \in \Delta\mathcal{Y}\}$ is Donsker, and $\frac{1-D}{p}\frac{p_0(X)}{1-p_0(X)}$ is a uniformly bounded and measurable function so that we can apply Van Der Vaart and Wellner, (1996, Example 2.10.10). Then, the result holds by Van Der Vaart and Wellner, (1996, Example 2.10.7). □

Lemma B.10. *Let $\mathrm{F}_{\Delta Y_{0t}|D=1}(\Delta, \bar{p}) = \mathrm{E}\left[\frac{1-D_t}{p}\frac{\bar{p}(X)}{1-\bar{p}(X)}\mathbb{1}\{\Delta Y_t \leq \Delta\}\right]$ denote the propensity score reweighted distribution of the change in untreated potential outcomes for the treated group for a particular propensity score $\bar{p}$. Then, the pathwise derivative $\Gamma(p_0)(\hat{p} - p_0)$ exists and is given by*

$$\Gamma(\Delta, p_0)(\hat{p} - p_0) = \mathrm{E}\left[\frac{1 - D_t}{p}\frac{\mathbb{1}\{\Delta Y_t \leq \Delta\}}{(1 - p_0(X))^2}(\hat{p}(X) - p_0(X))\right]$$

Proof.

$$\begin{aligned}
&\frac{\mathrm{F}_{\Delta Y_{0t}|D=1}(\Delta, p_0 + t(\bar{p} - p_0)) - \mathrm{F}_{\Delta Y_{0t}|D=1}(\Delta, p_0)}{t} \\
&\quad = \mathrm{E}\left[\frac{1 - D_t}{p}\mathbb{1}\{\Delta Y_t \leq \Delta\}\left(\frac{p_0(X) + t(\bar{p}(X) - p_0(X))}{1 - p_0(X) - t(\bar{p}(X) - p_0(X))} - \frac{p_0(X)}{1 - p_0(X)}\right)\right] \Big/ t \\
&\quad = \mathrm{E}\left[\frac{1 - D_t}{p}\mathbb{1}\{\Delta Y_t \leq \Delta\}\frac{(\bar{p}(X) - p_0(X))}{(1 - p_0(X))^2 - t(\bar{p}(X) - p_0(X)) + p_0(X)t(\bar{p}(X) - p_0(X))}\right] \\
&\quad \to \mathrm{E}\left[\frac{1 - D_t}{p}\mathbb{1}\{\Delta Y_t \leq \Delta\}\frac{(\bar{p}(X) - p_0(X))}{(1 - p_0(X))^2}\right] \qquad \text{as } t \to 0
\end{aligned}$$

□

Lemma B.11. *Under the Conditional Distributional Difference in Differences Assumption, the Copula Stability Assumption, Assumptions 4.1, 4.2, 5.1 and 6.1 to 6.4,*

$$|\,\mathrm{F}_{\Delta Y_{0t}|D=1}(\Delta, \hat{p}) - \mathrm{F}_{\Delta Y_{0t}|D=1}(\Delta, p_0) - \Gamma(\Delta, p_0)(\hat{p} - p_0)|_\infty = o_p(1)$$

Proof.

$$\begin{aligned}
&|\,\mathrm{F}_{\Delta Y_{0t}|D=1}(\Delta, \hat{p}) - \mathrm{F}_{\Delta Y_{0t}|D=1}(\Delta, p_0) - \Gamma(\Delta, p_0)(\hat{p} - p_0)|_\infty \\
&\qquad \leq \left|\mathrm{E}\left[\frac{1 - D_t}{p}\left(\frac{\hat{p}(X)}{1 - \hat{p}(X)} - \frac{p_0(X)}{1 - p_0(X)} - \frac{(\hat{p}(X) - p_0(X)}{(1 - p_0(X))^2}\right)\right]\right| \\
&\qquad = \left|\mathrm{E}\left[\frac{1 - D_t}{p}\left(\frac{(\hat{p}(X) - p_0(X))^2}{(1 - \hat{p}(X))(1 - p_0(X))^2}\right)\right]\right| \\
&\qquad \leq C \sup_{x \in \mathcal{X}} |\hat{p}(x) - p_0(x)|^2 \to 0
\end{aligned}$$

where the last line holds because p is bounded away from 0 and 1, $p_0(x)$ is uniformly bounded away from 1, and $\hat{p}(x)$ converges uniformly to $p_0(x)$. Then, the result holds because under

Assumptions 6.1 to 6.4, $\sup_{x\in\mathcal{X}} |\hat{p}(x) - p_0(x)| = o_p(n^{-1/4})$ (Hirano, Imbens, and Ridder, 2003). □

Lemma B.12. *Under the Conditional Distributional Difference in Differences Assumption, the Copula Stability Assumption, Assumptions 4.1, 4.2, 5.1 and 6.1 to 6.4*

$$\sup_{\Delta\in\Delta\mathcal{Y}} \left| \sqrt{n} \left((\hat{F}_{\Delta Y_{0t}|D=1}(\Delta;\hat{p}) - F_{\Delta Y_{0t}|D=1}(\Delta;p_0)) - \frac{1}{n}\sum_{i=1}^{n} \varphi(W_i,\Delta) \right) \right| = o_p(1)$$

Proof. For any $y \in \mathcal{Y}$,

$$\begin{aligned}
&\sqrt{n}(\hat{F}_{\Delta Y_{0t}|D=1}(\Delta;\hat{p}) - F_{\Delta Y_{0t}|D=1}(\Delta;p_0)) \\
&\quad= \sqrt{n}\left(\hat{F}_{\Delta Y_{0t}|D=1}(\Delta;\hat{p}) - \hat{F}_{\Delta Y_{0t}|D=1}(\Delta;p_0)\right) + \sqrt{n}(\hat{F}_{\Delta Y_{0t}|D=1}(\Delta;p_0) - F_{\Delta Y_{0t}|D=1}(\Delta;p_0)) \\
&\quad= \sqrt{n}\left(F_{\Delta Y_{0t}|D=1}(\Delta;\hat{p}) - F_{\Delta Y_{0t}|D=1}(\Delta;p_0)\right) + \sqrt{n}(\hat{F}_{\Delta Y_{0t}|D=1}(\Delta;p_0) - F_{\Delta Y_{0t}|D=1}(\Delta;p_0)) + o_p(1) \\
&\quad= \sqrt{n}\Gamma(\delta,p_0)(\hat{p}-p_0) + \sqrt{n}(\hat{F}_{\Delta Y_{0t}|D=1}(\Delta;p_0) - F_{\Delta Y_{0t}|D=1}(\Delta;p_0)) + o_p(1) \\
&\quad= \frac{1}{\sqrt{n}}\sum_{i=1}^{n} \frac{\mathrm{E}[\mathbb{1}\{\Delta Y_t \le \Delta\}|X=X_i, D_t=0]}{p(1-p_0(X_i))}(D_i - p_0(X_i)) \\
&\qquad\qquad + \sqrt{n}(\hat{F}_{\Delta Y_{0t}|D=1}(\Delta;p_0) - F_{\Delta Y_{0t}|D=1}(\Delta;p_0)) + o_p(1) \\
&\quad= \frac{1}{\sqrt{n}}\sum_{i=1}^{n} \varphi(W_i,\Delta) - F_{\Delta Y_{0t}|D=1}(\Delta;p_0) + o_p(1)
\end{aligned}$$

where the second equality holds from Vaart and Wellner, (2007) under Assumptions 6.1 to 6.4 and under Lemmas B.9 to B.11. The third equality holds Lemmas B.10 and B.11. The last equality holds under Assumptions 6.1 to 6.4 and using the results on the series logit estimator in Hirano, Imbens, and Ridder, (2003). □

Proof of Proposition 3 For the counterfactual distribution of untreated potential outcomes for the treated group,

$$\sqrt{n}(\hat{F}_{Y_{0t}|D=1}(y) - F_{Y_{0t}|D=1}(y)) = \sqrt{n}(\phi_n(F_0) - \phi_0(F_0)) + \phi'_{F_0}\sqrt{n}(\hat{F} - F_0) + o_p(1)$$

which follows from an argument similar to Lemma B.6 for the first term (where we now also use the result in Lemma B.18); Lemma B.7 continues to hold and also because of the Donsker result in Lemma B.9.

Proof of Theorem 5 The result follows under the conditions stated in the theorem, by the Hadamard differentiability of the quantile map (Van Der Vaart and Wellner, 1996, Lemma 3.9.23(ii)) and by Proposition 3.

Lemma B.13. *Under the Conditional Distributional Difference in Differences Assumption, the Copula Stability Assumption, Assumptions 4.1, 4.2, 5.1 and 6.1 to 6.4. For any* $\Delta \in$

$\Delta\mathcal{Y}_{0t|D_t=1}$,

$$\sqrt{n}\left(\hat{\mathrm{F}}^{*}_{\Delta Y_{0t}|D_t=1}(\Delta;\hat{p}^*) - \hat{\mathrm{F}}_{\Delta Y_{0t}|D_t=1}(\Delta;\hat{p})\right)$$
$$= \sqrt{n}\left(\hat{\mathrm{F}}^{*}_{\Delta Y_{0t}|D_t=1}(\Delta;p_0) - \hat{\mathrm{F}}_{\Delta Y_{0t}|D_t=1}(\Delta;p_0) + \Gamma(\Delta,\hat{p})(\hat{p}^* - \hat{p})\right) + o_p(1)$$

Proof.

$$\begin{aligned}
&\sqrt{n}\left(\hat{\mathrm{F}}^{*}_{\Delta Y_{0t}|D_t=1}(\Delta;\hat{p}^*) - \hat{\mathrm{F}}_{\Delta Y_{0t}|D_t=1}(\Delta;\hat{p})\right)\\
&\quad= \sqrt{n}\left\{\left(\hat{\mathrm{F}}^{*}_{\Delta Y_{0t}|D_t=1}(\Delta;\hat{p}^*) - \hat{\mathrm{F}}_{\Delta Y_{0t}|D_t=1}(\Delta;\hat{p}^*)\right)\right.\\
&\qquad\qquad\left. - \sqrt{n}\left(\hat{\mathrm{F}}_{\Delta Y_{0t}|D_t=1}(\Delta;p_0) - \hat{\mathrm{F}}_{\Delta Y_{0t}|D_t=1}(\Delta;p_0)\right)\right\}\\
&\qquad+ \sqrt{n}\left(\hat{\mathrm{F}}^{*}_{\Delta Y_{0t}|D_t=1}(\Delta;p_0) - \hat{\mathrm{F}}_{\Delta Y_{0t}|D_t=1}(\Delta;p_0) + \Gamma(\Delta,\hat{p})(\hat{p}^* - \hat{p})\right)\\
&\qquad+ \sqrt{n}\left\{\left(\hat{\mathrm{F}}_{\Delta Y_{0t}|D_t=1}(\Delta;\hat{p}^*) - \mathrm{F}_{\Delta Y_{0t}|D_t=1}(\Delta;\hat{p}^*)\right)\right.\\
&\qquad\qquad\left. - \sqrt{n}\left(\hat{\mathrm{F}}_{\Delta Y_{0t}|D_t=1}(\Delta;\hat{p}) - \mathrm{F}_{\Delta Y_{0t}|D_t=1}(\Delta;\hat{p})\right)\right\}\\
&\qquad+ \sqrt{n}\left(\mathrm{F}_{\Delta Y_{0t}|D_t=1}(\Delta;\hat{p}^*) - \mathrm{F}_{\Delta Y_{0t}|D_t=1}(\Delta;\hat{p}) - \Gamma(\Delta,\hat{p})(\hat{p}^* - \hat{p})\right)
\end{aligned}$$

The first, third, and fourth terms in the first equality converge uniformly to 0. These hold by Lemma B.9, by arguments similar to those in Lemma B.11 and because $\sup_{x\in\mathcal{X}}|\hat{p}^*(x) - \hat{p}(x)| = o_p(n^{-1/4})$ which holds under our conditions on the propensity score. This implies the result. □

Lemma B.14. *Let* $\hat{G}^*_X(x) = \sqrt{n}\left(\hat{\mathrm{F}}^{*}_X(x) - \hat{\mathrm{F}}_X(x)\right)$ *and let* $\tilde{G}^p_{Y_{0t}|D_t=1}(\Delta) = \sqrt{n}\left(\hat{\mathrm{F}}^{p*}_{\Delta Y_{0t}|D_t=1}(\Delta) - \hat{\mathrm{F}}^{p}_{\Delta Y_{0t}|D_t=1}(\Delta)\right)$. *Under the Conditional Distributional Difference in Differences Assumption, the Copula Stability Assumption, Assumptions 4.1, 4.2, 5.1 and 6.1 to 6.4.*

$$(\hat{G}^p_{\Delta Y_{0t}|D_t=1}, \hat{G}_{\Delta Y_{t-1}|D_t=1}, \tilde{G}^p_{Y_{0t}|D_t=1}, \hat{G}_{Y_t|D_t=1}, \hat{G}_{Y_{t-1}|D_t=1}, \hat{G}_{Y_{t-2}|D_t=1}) \rightsquigarrow_* (\mathbb{W}^p_1, \mathbb{W}^p_2, \mathbb{V}^p_0, \mathbb{V}^p_1, \mathbb{W}^p_3, \mathbb{W}^p_4)$$

where $(\mathbb{W}^p_1, \mathbb{W}^p_2, \mathbb{V}^p_0, \mathbb{V}^p_1, \mathbb{W}^p_3, \mathbb{W}^p_4)$ *is the tight Gaussian process given in Proposition 3.*

Proof. The result follows from Lemma B.13 and by Van Der Vaart and Wellner, (1996, Theorem 3.6.1). □

Proof of Theorem 6 The result from Lemma B.14, by the Hadamard Differentiability of our estimator of the QTT, and by the Delta method for the bootstrap (Van Der Vaart and Wellner, 1996, Theorem 3.9.11).

B.4 Additional Auxiliary Results

Lemma B.15. *Assume Y is continuously distributed. Then,*

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

Proof. Because Y is continuously distributed,

$$\frac{1}{n}\sum_{i=1}^{n}\left(\mathbb{1}\{\hat{F}_Y(X_i)\leq q\}-\mathbb{1}\{X_i\leq\hat{F}_Y^{-1}(q)\}\right)=\begin{cases}0 & \text{if } q\in \mathit{Range}(\hat{F}_Y)\\ -\frac{1}{n} & \text{otherwise}\end{cases}$$

which implies the result. □

Lemma B.16. *Assume Y and Z are continuously distributed. Then,*

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

Proof. $\hat{F}_Z^{-1}(\hat{F}_Y(X_i))\leq z\Leftrightarrow\hat{F}_Y(X_i)\leq\hat{F}_Z(z)$ which holds by Van der Vaart, (2000, Lemma 21.1(i)). Then, an application of Lemma B.15 implies the result. □

Lemma B.17.

$$\sqrt{n}\left\{\frac{1}{n}\sum_{i=1}^{n}\mathrm{F}_Y(Z_i)-\mathrm{E}[F_Y(Z)]-\left(\frac{1}{n}\sum_{i=1}^{n}\hat{F}_Y(Z_i)-\mathrm{E}[\hat{F}_Y(Z)]\right)\right\}=o_p(1)\tag{19}$$

Proof. The result follows since Equation (19) is equal to

$$\sqrt{n}\int_{\mathcal{Z}}\int_{\mathcal{Y}}\mathbb{1}\{y\leq z\}\;\mathrm{d}(\hat{F}_Y-F_Y)(y)\;\mathrm{d}(\hat{F}_Z-F_Z)(z)$$

which converges to 0. □

Lemma B.18.

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}\{\hat{F}_{\Delta Y_{0t}|D_t=1}(X_i)\leq q\}-\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}\{X_i\leq\hat{F}^{-1}_{\Delta Y_{0t}|D_t=1}(q)\}\right)\xrightarrow{p}0$$

Proof. This follows because

$$\left|\frac{1}{n}\sum_{i=1}^{n}\left(\mathbb{1}\{\hat{F}_{\Delta Y_{0t}|D_t=1}(X_i)\leq q\}-\mathbb{1}\{X_i\leq\hat{F}^{-1}_{\Delta Y_{0t}|D_t=1}(q)\}\right)\right|\leq\frac{C}{n}$$

where C is an arbitrary constant and the result holds because the difference is equal to 0 if $q\in \mathit{Range}(\hat{F}_{\Delta Y_{0t}|D_t=1})$ and is less than or equal to $\frac{1}{np}\times\max\left\{\frac{\hat{p}(X_i)}{1-\hat{p}(X_i)}\right\}$ which is less than or

equal to $\frac{C}{n}$ because $\hat{p}(\cdot)$ is bounded away from 0 and 1 with probability 1 and p is greater than 0. This implies the first part. The main result holds by exactly the same reasoning as Lemma B.16. □

C 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.

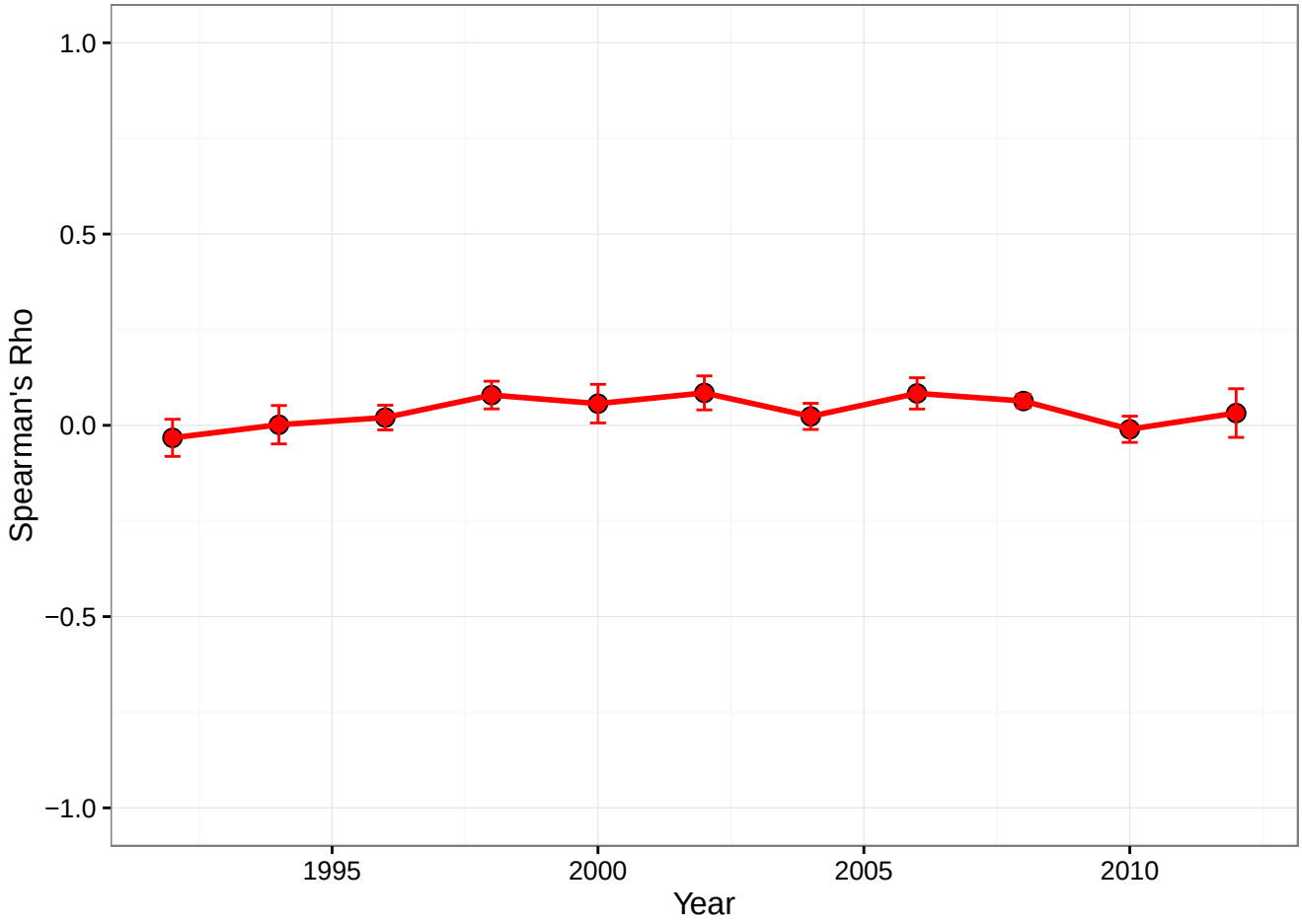
Table 2: QTT Estimates for Job Training Program

	0.7	Diff	0.8	Diff	0.9	Diff
			PanelQTT Method			
PanelQTT SL	3.21*	1.40	5.80*	3.53*	7.25*	4.05*
	(1.35)	(1.34)	(1.11)	(1.23)	(2.40)	(1.75)
PanelQTT Cov	1.46	−0.34	2.59*	0.32	2.45	−0.74
	(1.44)	(1.22)	(1.22)	(1.43)	(2.28)	(1.51)
PanelQTT UNEM	3.32*	1.51	5.80*	3.53*	7.92*	4.72*
	(1.43)	(1.37)	(1.17)	(1.24)	(2.15)	(1.54)
PanelQTT No Cov	−0.77	−2.57*	0.58	−1.69	−0.25	−3.45*
	(1.27)	(0.98)	(0.99)	(1.10)	(2.09)	(1.24)
			Conditional Independence Method			
CI SL	4.52*	2.71*	6.03*	3.76*	4.98	1.78
	(1.47)	(1.19)	(1.92)	(1.84)	(4.00)	(3.25)
CI Cov	−5.13*	−6.93*	−6.97*	−9.25*	−10.54*	−13.74*
	(1.23)	(1.14)	(1.40)	(1.48)	(2.64)	(2.02)
CI UNEM	3.45*	1.64	5.14*	2.87	4.24	1.04
	(1.40)	(1.22)	(1.54)	(1.53)	(3.22)	(2.48)
CI No Cov	−19.19*	−20.99*	−20.86*	−23.14*	−23.87*	−27.07*
	(0.89)	(0.75)	(0.92)	(1.08)	(1.92)	(1.12)
			Change in Changes			
CiC Cov	3.74*	1.94	4.32*	2.04	5.03*	1.84
	(0.88)	(1.01)	(1.02)	(1.23)	(1.54)	(1.76)
CiC UNEM	0.37	−1.44	1.84	−0.43	2.09	−1.10
	(1.31)	(1.35)	(1.43)	(1.45)	(2.02)	(1.96)
CiC No Cov	8.16*	6.36*	9.83*	7.56*	10.07*	6.87*
	(0.80)	(0.60)	(1.04)	(1.08)	(2.57)	(1.97)
			Quantile D-i-D			
QDiD Cov	2.18*	0.37	2.85*	0.58	2.45	−0.75
	(0.71)	(0.91)	(0.97)	(1.23)	(1.59)	(1.77)
QDiD UNEM	1.10	−0.70	2.66*	0.39	2.35	−0.84
	(1.13)	(1.21)	(1.26)	(1.34)	(1.87)	(1.92)
QDiD No Cov	4.21*	2.41*	4.65*	2.38*	4.90*	1.70
	(0.97)	(0.87)	(1.09)	(1.04)	(2.05)	(1.31)
Experimental	1.80		2.27*		3.20	
	(0.93)		(1.13)		(2.04)	

Notes: This table provides estimates of the QTT 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 QTT and the QTT obtained from the experimental portion of the dataset. The columns identify the method (PanelQTT, CI, CiC, or QDiD) 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 PanelQTT 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 PanelQTT model and the CI model use propensity score re-weighting techniques based on the covariate set. The CiC and QDiD 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%.

D Figures

Figure 1: NLSY Dependence between Change and Initial Level of Annual Income



This figure reports Spearman's Rho (a summary measure of the copula) for the Change in Annual Income and the Initial Annual Income using a sample of 2,283 individuals in the National Longitudinal Study of Youths who have positive income in every even year from 1990-2012. Confidence intervals are obtained using the block bootstrap.