Quantile Treatment Effects in Difference in Differences Models with Panel Data

Brantly Callaway
Department of Economics
Temple University

Tong Li
Department of Economics
Vanderbilt University

Department of Economics
DETU Working Paper 17-01
August 2017

1301 Cecil B. Moore Avenue, Philadelphia, PA 19122

http://www.cla.temple.edu/economics/faculty/detu-working-paper-series/

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

Brantly Callaway† Tong Li‡

First Version: April 2015
This Version: August 2017

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 use our method to estimate the effect of increasing the minimum wage on quantiles of local labor markets' unemployment rates and find significant heterogeneity.

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, James Heckman, Magne Mogstad, Derek Neal, John Pepper, Peter Phillips, Pedro Sant'Anna, Azeem Shaikh, Youngki Shin, Steve Stern, Pierre Perron, Ed Vytlacil, Kaspar Wuthrich, and participants in seminars at Boston University, Korea University, Lancaster University, National Sun Yat-Sen University, the National University of Singapore, Seoul National University, 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, at the 11th World Congress of the Econometric Society, at the 2017 China Meeting of the Econometric Society held in Wuhan, China, June 2017, and at the 2017 International Conference on Challenges and Perspectives of Econometrics on Data Analysis held in Hsinchu, Taiwan, June 2017 for their comments and suggestions. Li acknowledges gratefully the hospitality and support of Becker Friedman Institute at the University of Chicago. Code for the method proposed in the paper is available as the `panel.qtet` method in the R `qte` package.

†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 an instrumental variable (Abadie, Angrist, and Imbens 2002; Chernozhukov and Hansen 2005; Carneiro and Lee 2009; Frolich and Melly 2013), or access to repeated observations over time (Athey and Imbens 2006; Bonhomme and Sauder 2011; Chernozhukov, Fernandez-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 having 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 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 analyzing the effect of increasing the minimum wage on quantiles of the unemployment rates of local labor markets. Despite the average effect of increasing the minimum wage on the unemployment rate being close to 0, using our method, we find that the average effect masks substantial heterogeneity. The difference between the 10th percentile of unemployment among counties that had higher minimum wages and the 10th percentile of counterfactual unemployment had they not had higher minimum wages is negative. However, the effect is quite different elsewhere in the distribution. At the median and upper quantiles, the effect is positive. As long as counties do not change their ranks (or at least do not change their ranks too much) in the distribution of unemployment rates due to the increase in the minimum wage, these results indicate that counties with tight labor markets experienced decreases in the unemployment rate following the minimum wage increase while counties with higher unemployment rates experienced more unemployment due to the increase in the minimum wage. We find similar results using alternative methods such as Quantile Difference in Differences and Change in Changes (Athey and Imbens 2006).

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:

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

Because no one is treated in previous periods, untreated potential outcomes are observed for both the treated group and untreated group.[1] That is,

$$Y_{t-1} = Y_{0t-1} \text{ and } Y_{t-2} = Y_{0t-2}$$

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

[1] To clearly distinguish between treated and untreated potential outcomes, we use a potential outcomes notation where Y_{1t}, Y_{0t-1}, Y_{0t-2} are observed outcomes for the treated group (but Y_{0t} is not an observed outcome for the treated group) and Y_{0t}, Y_{0t-1}, and Y_{0t-2} are observed outcomes for the untreated group.

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.[2] Researchers interested in program evaluation may be interested in other quantiles as well. For example, researchers studying a job training program may be interested in the effect of the program 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)[3] 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.

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

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

3 Identification

Let $\Delta Y_{0t} = Y_{0t} - Y_{0t-1}$ denote the time difference in untreated potential outcomes. 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.

3.1 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.[4] 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

[4] 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

$$\mathrm{F}_{\Delta Y_{0t}, Y_{0t-1}|D_t=1}(\delta, y') = C_{\Delta Y_{0t}, Y_{0t-1}|D_t=1}\left(F_{\Delta Y_{0t}|D_t=1}(\delta), F_{Y_{0t-1}|D_t=1}(y')\right)$$

where $C_{\Delta Y_{0t}, Y_{0t-1}|D_t=1}(\cdot, \cdot)$ is a copula function.[5] 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 $F_{\Delta Y_{0t}|D_t=1}$ and $F_{Y_{0t-1}|D_t=1}$ is the same as the dependence between the distributions $F_{\Delta Y_{0t-1}|D_t=1}$ and $F_{Y_{0t-2}|D_t=1}$. 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.

[5] 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 3.2. *Let $\Delta\mathcal{Y}_{t|D_t=0}$ denote the support of the change in untreated outcomes for the untreated group. Let $\Delta\mathcal{Y}_{t-1|D_t=1}$, $\mathcal{Y}_{t-1|D_t=1}$, and $\mathcal{Y}_{t-2|D_t=1}$ denote the support of the change in untreated outcomes for the treated group in period $(t-1)$, the support of untreated outcomes for the treated group in period $(t-1)$, and the support of untreated outcomes for the treated 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 3.3. *The observed data $(Y_{dt,i}, Y_{t-1,i}, Y_{t-2,i}, X_i, D_{it})$ are independently and identically distributed.*

Assumption 3.2 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. Compactness is not needed for identification, but we use it for inference later in the paper. Assumption 3.3 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 3.3 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 3.2 and 3.3*

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

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 $\mathrm{P}(Y_{0t} \leq y|D_t = 1) = \mathrm{E}[\mathbb{1}\{\Delta Y_{0t} + Y_{0t-1} \leq y\}|D_t = 1]$[6] 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.

The following example shows what additional conditions need to be satisfied for our model to be valid in a standard DID setup.

Example 1. *Consider the following baseline model for Mean DID.*

$$Y_{0it} = \theta_t + C_i + v_{it}$$

where θ_t is a time fixed effect that is common for the treated and untreated groups, C_i is individual heterogeneity that may be distributed differently across the treated group and untreated group, and v_{it} are time varying unobservables.[7] *For Mean DID to identify the ATT, it must be the case that $E[\Delta v_{it}|D_{it} = 1] = E[\Delta v_{it}|D_{it} = 0]$. Sufficient conditions for the assumptions in our model to hold are (i) $\Delta v_{it} \perp\!\!\!\perp D_{it}$ and (ii) $(v_{it}, v_{it-1}|C_i, D_{it} = 1)$ and*

[6] 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.

[7] One other thing to note in this model is that it does not restrict how treated outcomes are generated at all which is standard in the Mean DID case and holds in our case as well.

$(v_{it-1}, v_{it-2}|C_i, D_t = 1)$ *follow the same distribution.*

Condition (i) just strengthens Mean DID to Distributional DID. Condition (ii) implies that the Copula Stability Assumption will hold. Notice that it allows for serial correlation in the time varying unobservables, and it will hold automatically if the time varying unobservables are iid.

4 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.[8] This section extends the previous results to the case where a Conditional DDID assumption holds.

Conditional Distributional Difference in Differences Assumption.

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

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 2. *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*

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

- *The Unconditional Copula Stability Assumption holds*

Example 2 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 2 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.[9]

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 4.1. *$p \equiv P(D_t = 1) > 0$ 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.

[9] 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 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.

Theorem 2. *Under the Conditional Distributional Difference in Differences Assumption, the Copula Stability Assumption, and Assumptions 3.2, 3.3 and 4.1*

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

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 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 $\dfrac{p(X)}{1-p(X)}$ 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.

5 Estimation

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

We estimate

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

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

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

We estimate counterfactual quantiles by

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

where

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

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

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

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.

5.1 Inference

This section considers the asymptotic properties of the estimator. First, it focuses on the case with no covariates and then extends the results to the case where the Distributional

Difference in Differences Assumption holds conditional on covariates. The proofs for each of the results in this section are given in the Appendix.

5.1.1 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 3.2 and 3.3*

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

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 3.2 and 3.3*

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

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

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

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 for estimating the counterfactual distribution of untreated potential outcomes for the treated group 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 3.2 and 3.3*

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

where $(\bar{\mathbb{G}}_0(\tau), \bar{\mathbb{G}}_1(\tau))$ *is a stochastic process in the metric space* $(l^\infty(\mathcal{T}))^2$ *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* $b = 1, \dots, B$,

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

2. *Compute*

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

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

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

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

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 3.2 and 3.3,*

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

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 (Gine 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)

5.1.2 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 5.1. $\mathrm{E}[\mathbb{1}\{\Delta Y_{0t} \leq y\}|X, D_t = 0]$ *is continuously differentiable for all* $x \in \mathcal{X}$.

Assumption 5.2. *(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 5.3. *(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 5.4. *(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 5.1 to 5.4 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 5.2 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 5.3(i) is a standard assumption on differentiability of the propensity score and guarantees the existence of ν that satisfies the conditions of Assumption 5.4. Assumption 5.3(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* $\tilde{G}^p_{Y_{0t}|D_t=1}(y) = \sqrt{n}\left(\tilde{\mathrm{F}}^p_{Y_{0t}|D_t=1}(y) - \mathrm{F}_{Y_{0t}|D_t=1}(y)\right)$*. Under the Conditional Distributional Difference in Differences Assumption, the Copula Stability Assumption, Assumptions 3.2, 3.3, 4.1 and 5.1 to 5.4*

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

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 3.2, 3.3, 4.1 and 5.1 to 5.4*

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

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*

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

and

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

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 3.2, 3.3, 4.1 and 5.1 to 5.4*

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

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 3.2, 3.3, 4.1 and 5.1 to 5.4,*

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

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

6 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 $T \to \infty$.[10] 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 assigned to treatment. See Frolich 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, Fernandez-Val, Hahn, and Newey 2013). The most similar of these is Chernozhukov, Fernandez-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, Fernandez-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.

[10]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.

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.

7 Application

In this section, we use our method to study the effect of increasing the minimum wage on county-level unemployment rates. There is a wide body of research that studies the effect of the minimum wage on employment exploiting policy level changes across states (for example, Neumark and Wascher (1992) and Dube, Lester, and Reich (2010), among many others). Like most of the literature, we use variation in state-level changes in the minimum wage. Also, we suppose that there may be time invariant differences in the unemployement rate across counties that cannot be accounted for by observable differences in county characteristics. This implies that a DID approach should be used and is in line with much of the literature on minimum wage increases.

The aim of this section is different from most research on the effect of increasing the minimum wage. The literature almost exclusively looks at the average effect, or the coefficient in a linear regression model, of increasing the minimum wage on employment for teenagers, restaurant workers, or some other subgroup. Instead, by looking at the QTE, we examine how the effect of increasing the minimum wage varies by the strength of a county's local labor market. In other words, we ask the question: How does the effect of increasing the minimum wage differ across counties that would have had relatively high (or low) unemployment rates in the absence of the change in minimum wage policy? This goal is also different from trying to understand the effect of minimum wage increases at different parts of the individual income distribution as in Dube (2017).

Unlike most of the literature on minimum wages, instead of using a long panel of counties, states, and many changes in minimum wage policy across states; we focus on a particular period where the federal minimum wage was flat while there is variation in state minimum wages. The U.S. federal minimum wage increased from \$4.25 to \$5.15 between 1996 and 1997. It did not increase again until the Fair Minimum Wage Act was proposed on January 5, 2007 and enacted on May 25, 2007. The Act increased the federal minimum wage to \$5.85

on July 24, 2007 and increased the minimum wage in two more increments, settling at $7.25 in July of 2009.

In 2006, there were 33 states for whom the federal minimum wage was the binding minimum wage in the state. Of these, we drop two states – New Hampshire and Pennsylvania – because they are located in the Northern census region; census region is known to be an important control in the minimum wage literature (Dube, Lester, and Reich 2010) and almost all states in the Northern census region had minimum wages higher than the federal minimum wage by 2006. Of the remaining states, 11 increased their minimum wage by the first quarter of 2007 while 20 did not increase their minimum wage until the federal minimum wage increased in July of 2007. [11]

County level unemployment rates are the outcome variable. We obtain these from the Local Area Unemployment Statistics Database from the Bureau of Labor Statistics. Unemployment rates are available monthly and we use unemployment rates in February as the outcome variable. We choose February instead of January because it does not overlap with the holidays and choose it over later months because it is further away from the federal minimum wage change in July. We also merge in county characteristics from the 2000 County Data Book. In our application, these include 2000 county population and 1997 county median income. We collected data for each year from 2000-2007. Our method requires three periods of data, but the earlier periods allow us to pre-test our model in earlier periods.

Table 1 provides summary statistics. From 2005-2007, the level of unemployment rates is higher for treated counties than for untreated counties. The gap narrows from 2005 to 2006, the period before any counties have increased minimum wages, and then expands again from 2006 to 2007; this may provide some suggestive evidence that the minimum wage is inreasing unemployment rates on average. Counties that are treated are also different from untreated counties in terms of their observable characteristics. Treated counties are more likely to be in the West and North-Central regions while untreated counties are more likely to be in the South. Median incomes are very similar (though statistically different) across treated and untreated counties. And treated counties tend to be more populated; log population of 10.34 for treated counties is almost 31,000 while log population of 9.91 for untreated counties is just over 20,000.

The main results from using our method are presented in Figure 1. The upper panel provides estimates without conditioning on covariates. The lower panel provides estimates that

[11] The states that increased their minimum wage were: Arizona, Arkansas, Colorado, Maryland, Michigan, Montana, Nevada, North Carolina, Ohio, and West Virginia. The states that did not increase their minimum wage were: Alabama, Georgia, Idaha, Indiana, Iowa, Kansas, Kentucky, Louisiana, Mississippi, Nebraska, New Mexico, North Dakota, Oklahoma, South Carolina, South Dakota, Tennessee, Texas, Utah, Virginia, and Wyoming

condition on county characteristics; the specification for the propensity score interacts region with quadratic terms in log median income and log population as well as their interaction. The results are very similar whether or not covariates are included.

On average, we find that the effect of increasing the minimum wage has a small positive effect on the unemployment rate. Both with and without covariates, we estimate that increasing the minimum wage increases the unemployment rate by 0.12 percentage points. Without covariates, the effect is statistically significant. With covariates, the effect is not statistically significant. However, there is much heterogeneity. At the low end of the unemployment rate distribution, the effect of increasing the minimum wage on the unemployment rate appears to be negative. For example, at the 10th percentile, the unemployment rate is estimated to be 0.44 (p-value: 0.000) percentage points lower following the minimum wage increase than it would have been without the minimum wage increase (with covariates the estimate is 0.45 (p-value: 0.008)). However, in the middle and upper parts of the unemployment rate distibution, increasing the minimum wage appears to increase unemployment. The difference between the medians of unemployment rates in the presence or absence of the minimum wage increase is 0.31 (p-value: 0.000) percentage points (with covariates the estimate is 0.32 (p-value: 0.029)). The estimated difference between the 90th percentiles is 0.36 (p-value: 0.029) percentage points (with covariates the estimate is 0.27 (p-value: 0.216).

For comparison, Figure 3 plots bounds on the QTT when no assumption is made about the copula between the change in untreated potential outcomes and the initial level of untreated potential outcomes for the treated group as in Fan and Yu (2012). These bounds are very wide. For example, the difference between the median unemployment rate for treated counties and their counterfactual unemployment rate is bounded between -1.01 and 1.41.

Neither our Distributional Difference in Differences Assumption nor the Copula Stability Assumption are directly testable, but, like existing Difference in Differences methods, our assumptions can be pre-tested when additional pre-treatment periods are available. The simplest way to implement a pre-test is to estimate the model in the period (or periods) before treatment and test that the QTT is 0 for all values of τ. Also, because our Copula Stability Assumption is new, we provide an additional test for only the CSA. The idea of this test is to compute Kendall's Tau (a standard dependence measure that depends only on the copula (see Nelsen (2007))) in each pre-treatment year and test whether or not it changes over time. We perform both of these tests on the minimum wage data next.

Figure 2 plots Kendall's Tau for the change in unemployment rates and the initial level of unemployment rates for treated counties from 2001 to 2006. Kendall's Tau varies very little over this period and is always somewhat less than 0 indicating slight negative dependence between the change and initial level of unemployment. A Wald test fails to reject the equality of Kendall's Tau in all periods (p-value: 0.524). Second, we compute QTTs in each pre-treatment period from 2002 to 2006. In these periods, the QTTs should be equal to 0 everywhere. These are available in Supplementary Appendix Figure 2 and our method tends to perform very well in the earlier periods. Finally, as an additional robustness check, we compute QTTs using the Change in Changes method with and without covariates and with the Quantile Difference in Differences method (these are available in Supplementary Appendix Figure 1). These other methods show very similar patterns as our main results.

Taken together, these results suggest that there is a great deal of heterogeneity of the effect of increasing the minimum wage across local labor markets. If we impose the additional assumption that counties maintain their rank in the distribution of unemployment when the minimimum wage increases, the results indicate that counties with tight labor markets experience decreases in unemployment while counties that with high unemployment see fairly large increases in unemployment. Even in the absence of such an assumption, our results indicate that increasing the minimum wage can have negative consequences for some loal labor markets although the average effect may be fairly small.

8 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 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, Stephane 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] 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.

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

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

[11] 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.

[12] 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.

[13] 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.

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

[15] 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.

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

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

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

[19] 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.

[20] Dube, Arindrajit. "Minimum wages and the distribution of family incomes". Working Paper. 2017.

[21] Dube, Arindrajit, T William Lester, and Michael Reich. "Minimum wage effects across state borders: estimates using contiguous counties". *The review of economics and statistics* 92.4 (2010), pp. 945–964.

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

[32] 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.

[33] 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.

[34] 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.

[35] 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.

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

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

[38] 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.

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

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

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

[42] 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.

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

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

[45] 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.

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

[47] Neumark, David and William Wascher. "Employment effects of minimum and subminimum wages: panel data on state minimum wage laws". *Industrial & Labor Relations Review* 46.1 (1992), pp. 55–81.

[48] 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.

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

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

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

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

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

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

[55] Wuthrich, 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 3.3*

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

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

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

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); Wuthrich (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 $f_t(\cdot,\cdot) = f_{\Delta Y_{0t}, Y_{0t-1}|D_t=1}(\cdot,\cdot)$ be the joint pdf of the change in untreated potential outcome and the initial untreated potential outcome for the treated group, and let $f_{t-1}(\cdot,\cdot) = f_{\Delta Y_{0t-1}, Y_{0t-2}|D_t=1}(\cdot,\cdot)$ be the joint pdf in the previous period. Similarly, let $c_t(\cdot,\cdot) = c_{\Delta Y_{0t}, Y_{0t-1}|D_t=1}(\cdot,\cdot)$ and $c_{t-1}(\cdot,\cdot) = c_{\Delta Y_{0t-1}, Y_{0t-2}|D_t=1}(\cdot,\cdot)$ 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. And, finally, let $\Delta\mathcal{Y} = \Delta\mathcal{Y}_{t|D_t=1}$ (the support of the change in untreated potential outcomes for the treated group) and $\mathcal{Y} = \mathcal{Y}_{t-1|D_t=1}$ (the support of outcomes for the treated group in period $t-1$). Then,

$$
\begin{aligned}
&P(Y_{0t} \le y | D_t = 1) \\
&\qquad = P(\Delta Y_{0t} + Y_{0t-1} \le y | D_t = 1) \\
&\qquad = \mathrm{E}[\mathbb{1}\{\Delta Y_{0t} \le y - Y_{0t-1}\} | D_t = 1] \\
&\qquad = \int_{\mathcal{Y}} \int_{\Delta\mathcal{Y}} \mathbb{1}\{\delta \le y - y'\} f_t(\delta, y') \, \mathrm{d}\delta \, \mathrm{d}y' \\
&\qquad = \int_{\mathcal{Y}} \int_{\Delta\mathcal{Y}} \mathbb{1}\{\delta \le y - y'\} c_t(F_{\Delta Y_{0t}|D_t=1}(\delta), F_{Y_{0t-1}|D_t=1}(y')) f_{\Delta Y_{0t}|D_t=1}(\delta) f_{Y_{0t-1}|D_t=1}(y') \, \mathrm{d}\delta \, \mathrm{d}y' && (5) \\
&\qquad = \int_{\mathcal{Y}} \int_{\Delta\mathcal{Y}} \mathbb{1}\{\delta \le y - y'\} c_{t-1}(F_{\Delta Y_{0t}|D_t=1}(\delta), F_{Y_{0t-1}|D_t=1}(y')) f_{\Delta Y_{0t}|D_t=1}(\delta) f_{Y_{0t-1}|D_t=1}(y') \, \mathrm{d}\delta \, \mathrm{d}y' && (6) \\
&\qquad = \int_{\mathcal{Y}} \int_{\Delta\mathcal{Y}} \mathbb{1}\{\delta \le y - y'\} f_{t-1}\left(F^{-1}_{\Delta Y_{0t-1}|D_t=1}(F_{\Delta Y_{0t}|D_t=1}(\delta)), F^{-1}_{Y_{0t-2}|D_t=1}(F_{Y_{0t-1}|D_t=1}(y'))\right) \\
&\qquad\quad \times \frac{f_{\Delta Y_{0t}|D_t=1}(\delta)}{f_{\Delta Y_{0t-1}|D_t=1}(F^{-1}_{\Delta Y_{0t-1}|D_t=1}(F_{\Delta Y_{0t}|D_t=1}(\delta)))} \times \frac{f_{Y_{0t-1}|D_t=1}(y')}{f_{Y_{0t-2}|D_t=1}(F^{-1}_{Y_{0t-2}|D_t=1}(F_{Y_{0t-1}|D_t=1}(y')))} \, \mathrm{d}\delta \, \mathrm{d}y' && (7)
\end{aligned}
$$

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: $u = F^{-1}_{\Delta Y_{0t-1}|D_t=1}(F_{\Delta Y_{0t}|D_t=1}(\delta))$ and $v = F^{-1}_{Y_{0t-2}|D_t=1}(F_{Y_{0t-1}|D_t=1}(y'))$. This implies the following:

1. $\delta = F^{-1}_{\Delta Y_{0t}|D_t=1}(F_{\Delta Y_{0t-1}|D_t=1}(u))$
2. $y' = F^{-1}_{Y_{0t-1}|D_t=1}(F_{Y_{0t-2}|D_t=1}(v))$
3. $\mathrm{d}\delta = \dfrac{f_{\Delta Y_{0t-1}|D_t=1}(u)}{f_{\Delta Y_{0t}|D_t=1}(F^{-1}_{\Delta Y_{0t}|D_t=1}(F_{\Delta Y_{0t-1}|D_t=1}(u)))} \mathrm{d}u$
4. $\mathrm{d}y' = \dfrac{f_{Y_{0t-2}|D_t=1}(v)}{f_{Y_{0t-1}|D_t=1}(F^{-1}_{Y_{0t-1}|D_t=1}(F_{Y_{0t-2}|D_t=1}(v)))} \mathrm{d}v$

Plugging in (1)-(4) in Equation 7 and noticing that the substitutions for $\mathrm{d}\delta$ and $\mathrm{d}y'$ cancel out the fractional terms in the third and fourth lines of Equation 7 implies

$$\begin{aligned}
\text{Equation 7} = &\int_{\mathcal{Y}_{t-2|D_t=1}} \int_{\Delta\mathcal{Y}_{t-1|D_t=1}} \mathbb{1}\{F^{-1}_{\Delta Y_{0t}|D_t=1}(F_{\Delta Y_{0t-1}|D_t=1}(u)) \leq y - F^{-1}_{Y_{0t-1}|D_t=1}(F_{Y_{0t-2}|D_t=1}(v))\} \\
&\times \quad f_{t-1}(u,v)\,\mathrm{d}u\mathrm{d}v && (8)\\
= &\mathrm{E}\left[\mathbb{1}\{F^{-1}_{\Delta Y_{0t}|D_t=1}(F_{\Delta Y_{0t-1}|D_t=1}(\Delta Y_{0t-1})) \leq y - F^{-1}_{Y_{0t-1}|D_t=1}(F_{Y_{0t-2}|D_t=1}(Y_{0t-2}))\}|D_t = 1\right] && (9)\\
= &\mathrm{E}\left[\mathbb{1}\{F^{-1}_{\Delta Y_{0t}|D_t=0}(F_{\Delta Y_{0t-1}|D_t=1}(\Delta Y_{0t-1})) \leq y - F^{-1}_{Y_{0t-1}|D_t=1}(F_{Y_{0t-2}|D_t=1}(Y_{0t-2}))\}|D_t = 1\right] && (10)
\end{aligned}$$

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,

$$\begin{aligned}
\mathrm{P}(\Delta Y_{0t} \leq \delta | D_t = 1) &= \frac{\mathrm{P}(\Delta Y_{0t} \leq \delta, D_t = 1)}{p} \\
&= \mathrm{E}\left[\frac{\mathrm{P}(\Delta Y_{0t} \leq \delta, D_t = 1|X)}{p}\right] \\
&= \mathrm{E}\left[\frac{p(X)}{p}\,\mathrm{P}(\Delta Y_{0t} \leq \delta | X, D_t = 1)\right] \\
&= \mathrm{E}\left[\frac{p(X)}{p}\,\mathrm{P}(\Delta Y_{0t} \leq \delta | X, D_t = 0)\right] && (11)\\
&= \mathrm{E}\left[\frac{p(X)}{p}\,\mathrm{E}[(1-D_t)\mathbb{1}\{\Delta Y_t \leq \delta)\}|X, D_t = 0]\right] && (12)\\
&= \mathrm{E}\left[\frac{p(X)}{p(1-p(X))}\,\mathrm{E}[(1-D_t)\mathbb{1}\{\Delta Y_t \leq \delta)\}|X]\right] \\
&= \mathrm{E}\left[\frac{1-D_t}{1-p(X)}\frac{p(X)}{p}\mathbb{1}\{\Delta Y_t \leq \delta\}\right] && (13)
\end{aligned}$$

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. $\square$

B.2 Proof of the results in Example 1

For the first part, notice that $\Delta Y_{0it} = \theta_t - \theta_{t-1} + \Delta v_{it}$. This has the same distribution for the treated group and untreated group under Condition (i).

For the second part, first note that,

$$\begin{aligned}
F_{\Delta Y_{0t}|D_t=1}(\delta) &= P(\Delta Y_{0t} \leq \delta | D_t = 1) \\
&= P(\Delta v_{it} \leq \delta - (\theta_t - \theta_{t-1}) | D_t = 1) \\
&= P(\Delta v_{it-1} \leq \delta - (\theta_t - \theta_{t-1}) | D_t = 1) \\
&= P(\Delta Y_{0t-1} \leq \delta - (\theta_t - \theta_{t-2}) | D_t = 1) \\
&= F_{\Delta Y_{0t-1}|D_t=1}(\delta - (\theta_t - \theta_{t-2}) | D_t = 1)
\end{aligned}$$

where the third equality holds by Condition (ii) and the Law of Iterated Expectations. Similarly,

$$\begin{aligned}
F_{Y_{0t-1}|D_t=1}(y) &= P(Y_{0t-1} \leq y | D_t = 1) \\
&= P(C_i + v_{it-1} \leq y - \theta_{t-1} | D_t = 1) \\
&= P(C_i + v_{it-2} \leq y - \theta_{t-1} | D_t = 1) \\
&= P(Y_{0t-2} \leq y - (\theta_{t-1} - \theta_{t-2}) | D_t = 1) \\
&= F_{Y_{0t-2}|D_t=1}(y - (\theta_{t-1} - \theta_{t-2}))
\end{aligned}$$

where the third equality holds by Condition (ii). Finally, consider

$$\begin{aligned}
&C_{\Delta Y_{0t}, Y_{0t-1}|D_t=1}(u, v) \\
&\quad = P(F_{\Delta Y_{0t}|D_t=1}(\Delta Y_{0t}) \leq u, F_{Y_{0t-1}|D_t=1}(Y_{0t-1}) \leq v | D_t = 1) \\
&\quad = P(F_{\Delta Y_{0t-1}|D_t=1}(\Delta Y_{0t} - (\theta_t - \theta_{t-2})) \leq u, F_{Y_{0t-2}|D_t=1}(Y_{0t-1} - (\theta_{t-1} - \theta_{t-2})) \leq v | D_t = 1) \\
&\quad = P(F_{\Delta Y_{0t-1}|D_t=1}(\Delta v_t - (\theta_{t-1} - \theta_{t-2})) \leq u, F_{Y_{0t-2}|D_t=1}(C + v_{t-1} + \theta_{t-2}) \leq v | D_t = 1) \\
&\quad = P(F_{\Delta Y_{0t-1}|D_t=1}(\Delta v_{t-1} - (\theta_{t-1} - \theta_{t-2})) \leq u, F_{Y_{0t-2}|D_t=1}(C + v_{t-2} + \theta_{t-2}) \leq v | D_t = 1) \\
&\quad = P(F_{\Delta Y_{0t-1}|D_t=1}(\Delta Y_{0t-1}) \leq u, F_{Y_{0t-2}|D_t=1}(Y_{0t-2}) \leq v | D_t = 1) \\
&\quad = C_{\Delta Y_{0t-1}, Y_{0t-2}|D_t=1}(u, v)
\end{aligned}$$

which proves the result. Condition (ii) implies that the joint distribution of (v_{it}, v_{it-1}, C_i) is the same as the joint distribution of $(v_{it-1}, v_{it-2}, C_i)$ which implies the result in the fourth equality.

B.3 Proof of the results in Example 2

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.4 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

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

and

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

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 $\Lambda : \mathbb{A}_\Lambda \mapsto \mathbb{E}$ where $\mathbb{E}$ is the set of all distribution functions with strictly positive, bounded densities and 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

$$\begin{aligned} \phi'_{F_0}(\gamma) = \int & \frac{\gamma_1 \circ F_{30}^{-1} \circ F_{40}(v_4) - F_{10}\left(\cdot - \frac{\gamma_4 - \gamma_3 \circ F_{30}^{-1} \circ F_{40}(v_4)}{f_{30} \circ F_{30}^{-1} \circ F_{40}(v_4)}\right) - \gamma_2 \circ F_{20}^{-1} \circ F_{10}(y - F_{30}^{-1} \circ F_{40}(v_4))}{f_{20} \circ F_{20}^{-1} \circ F_{10} \circ (y - F_{30}^{-1} \circ F_{40}(v_4))} \\ & \times f_{V_2|V_4}(F_{20}^{-1} \circ F_{10}(\cdot - F_{30}^{-1} \circ F_{40}(v_4)) \, \mathrm{d}F_{V_4}(v_4) \end{aligned}$$

□

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 $y \in \mathcal{Y}$,

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

Then, adding and subtracting the following terms:

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

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\} \qquad (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\} \qquad (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\} \qquad (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\} \qquad (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.

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

$\square$

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. $\square$

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.

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

$\square$

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.

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

□

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

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

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

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

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

□

Lemma B.11. *Under the Conditional Distributional Difference in Differences Assumption, the Copula Stability Assumption, Assumptions 3.2, 3.3, 4.1 and 5.1 to 5.4,*

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

Proof.

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

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 5.1 to 5.4, $\sup_{x\in\mathcal{X}} |\hat{p}(x) - p_0(x)| = o_p(n^{-1/4})$ (Hirano, Imbens, and Ridder 2003). $\square$

Lemma B.12. *Under the Conditional Distributional Difference in Differences Assumption, the Copula Stability Assumption, Assumptions 3.2, 3.3, 4.1 and 5.1 to 5.4*

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

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

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

where the second equality holds from Vaart and Wellner (2007) under Assumptions 5.1 to 5.4 and under Lemmas B.9 to B.11. The third equality holds under Lemmas B.10 and B.11. The last two equalities holds under Assumptions 5.1 to 5.4 and using the results on the series logit estimator in Hirano, Imbens, and Ridder (2003) and the result follows from the compactness of $\Delta\mathcal{Y}$. $\square$

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

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

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 3.2, 3.3, 4.1 and 5.1 to 5.4. 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.

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

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 3.2, 3.3, 4.1 and 5.1 to 5.4.*

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

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 holds by 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.5 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.

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

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. $\square$

C Tables

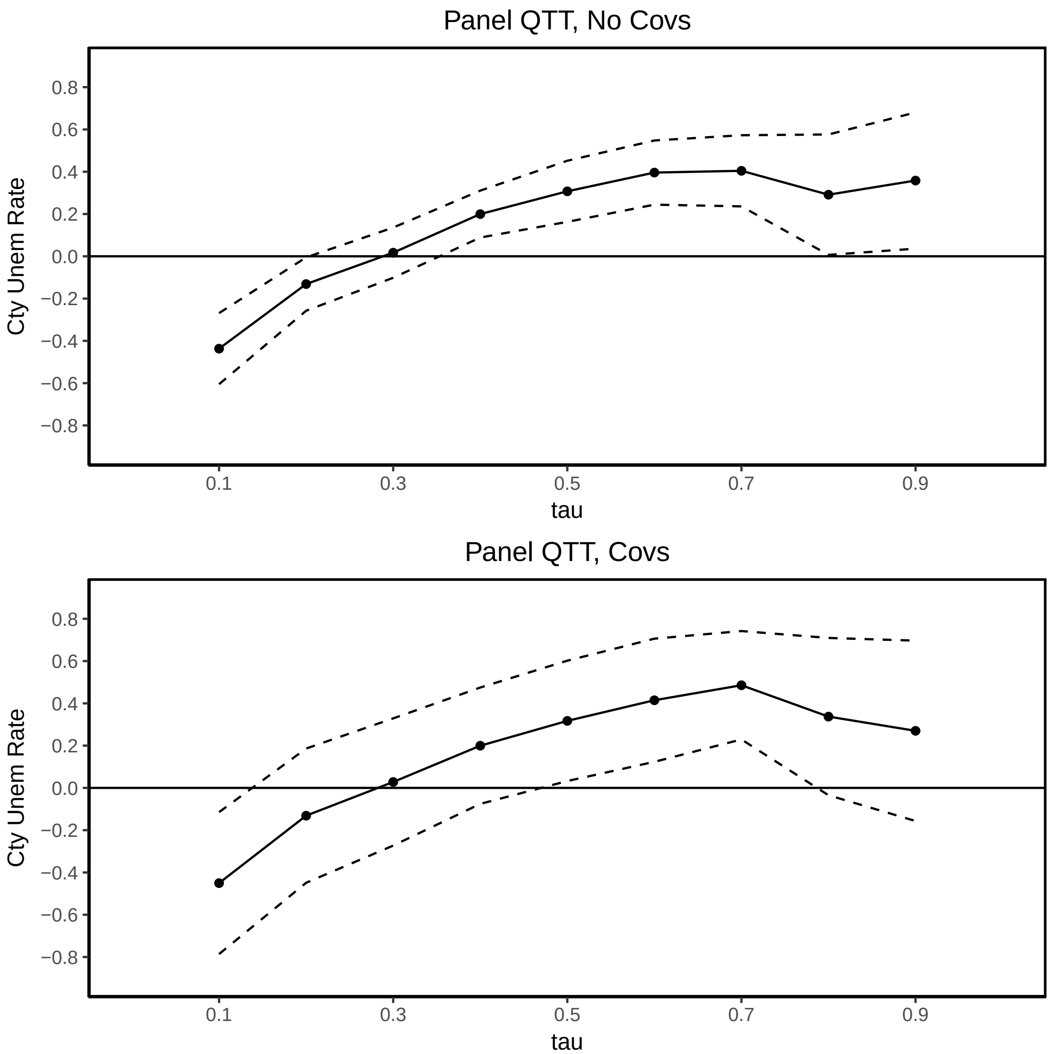
Table 1: Summary Statistics

	Treated Counties	Untreated Counties	Diff	P-val on Diff
Unem. Rate 2007	6.10	5.07	1.028	0.00
Unem. Rate 2006	6.25	5.34	0.904	0.00
Unem. Rate 2005	7.09	6.10	0.984	0.00
South	0.37	0.64	-0.274	0.00
North-Central	0.42	0.28	0.135	0.00
West	0.21	0.07	0.14	0.00
Log Med. Inc.	10.35	10.32	0.033	0.00
Log Pop.	10.34	9.91	0.437	0.00

Notes: Summary statistics for counties by whether or not their minimum wage increased in Q1 of 2007 (treated) or not (untreated). Unemployment rates are calculated using February unemployment and labor force estimates from the Local Area Unemployment Database. Median income is the county's median income from 1997 and comes from the 2000 County Data Book. Population is the county's population in 2000 and comes from the 2000 County Data Book.
Sources: Local Area Unemployment Statistics Database from the BLS and 2000 County Data Book

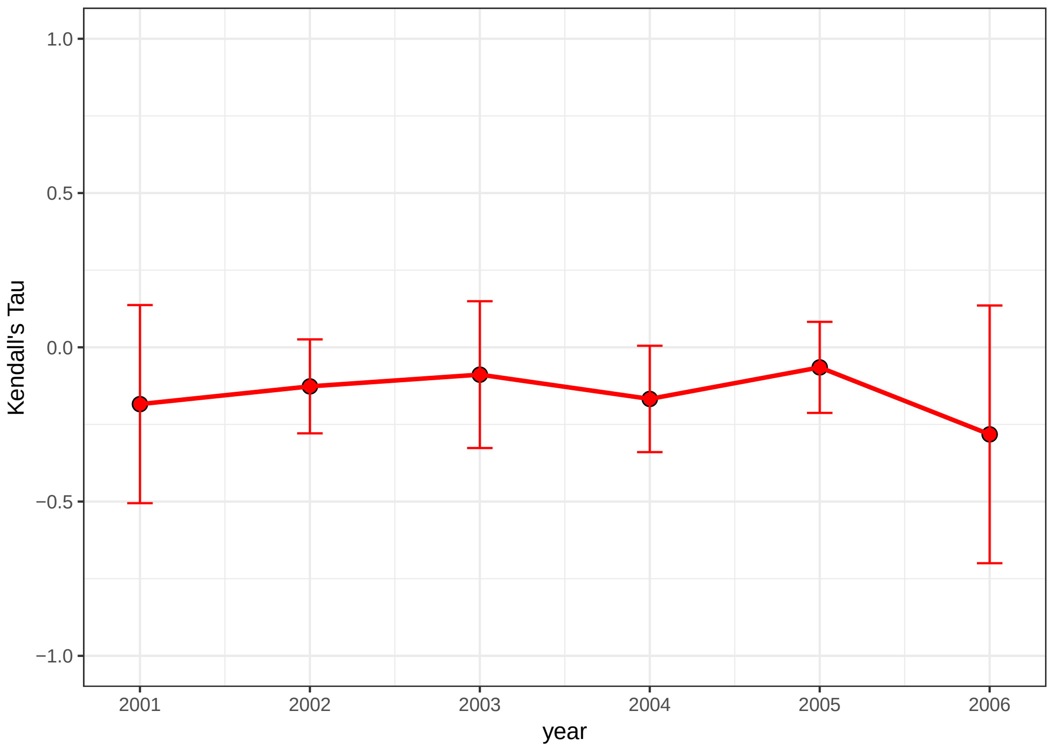
D Figures

Figure 1: QTT estimates of the effect of increasing the minimum wage on county-level unemployment rates



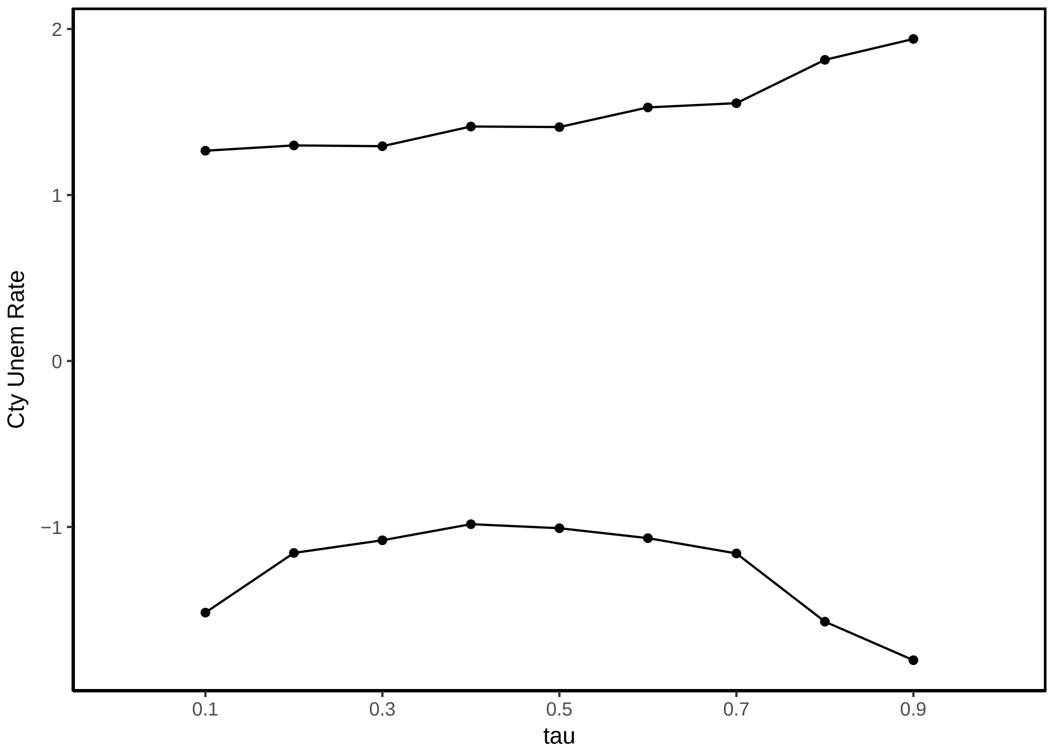
Notes: The top panel provides estimates of the QTT using the no-covariates version of the method proposed in the current paper. The lower panel provides QTT estimates when the DDID assumption holds only after conditioning on covariates. Standard errors are computed using the bootstrap with 100 iterations.
Sources: Local Area Unemployment Statistics Database from the BLS and 2000 County Data Book

Figure 2: Kendall's Tau estimates for treated counties by year



Notes: The figure estimates Kendall's Tau for states that increased their minimum wages in the first quarter of 2007. Standard errors are computed using the block bootstrap with 100 iterations.
Sources: Local Area Unemployment Statistics Database from the BLS

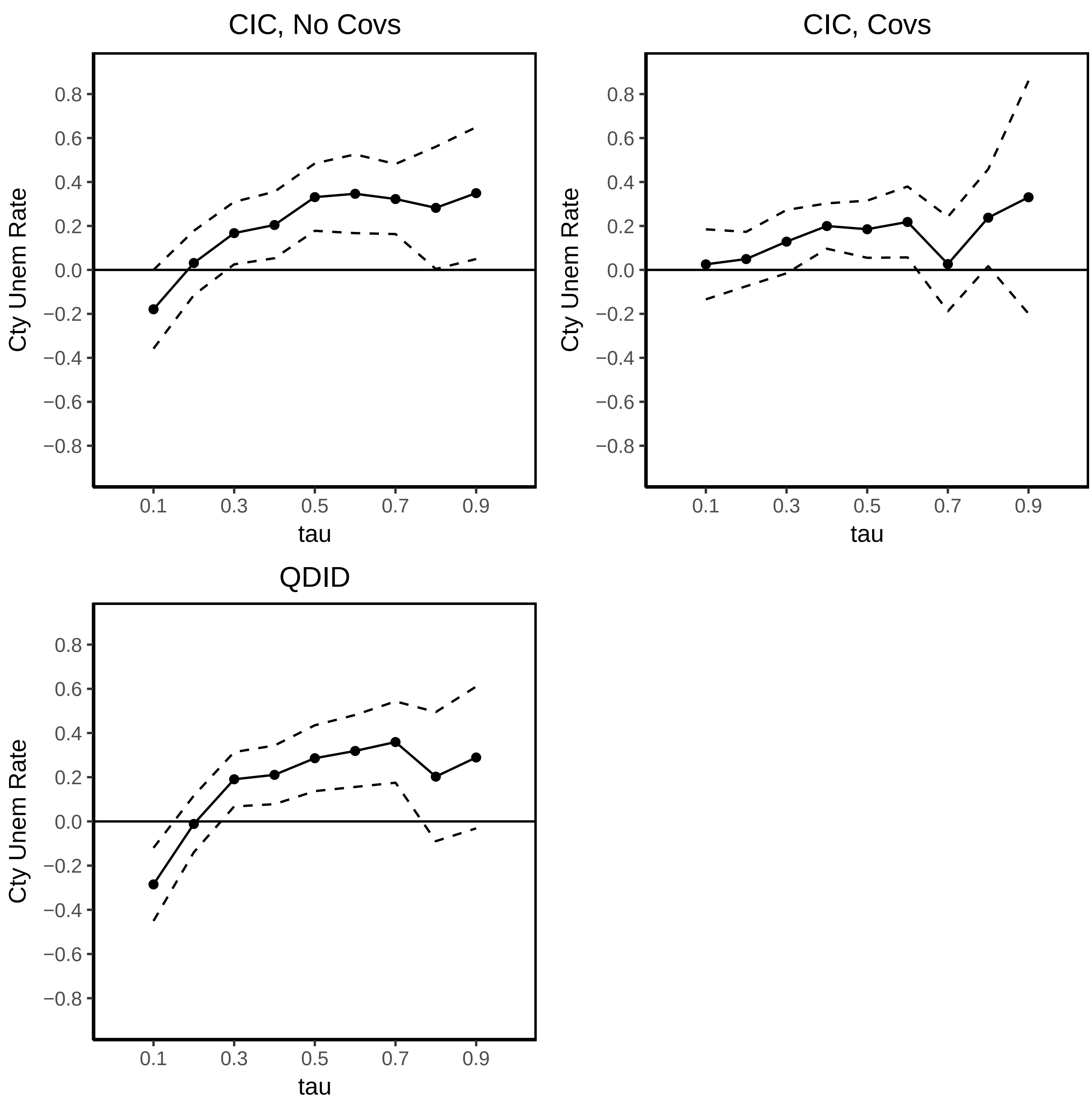
Figure 3: Bounds for QTT with Unknown Copula



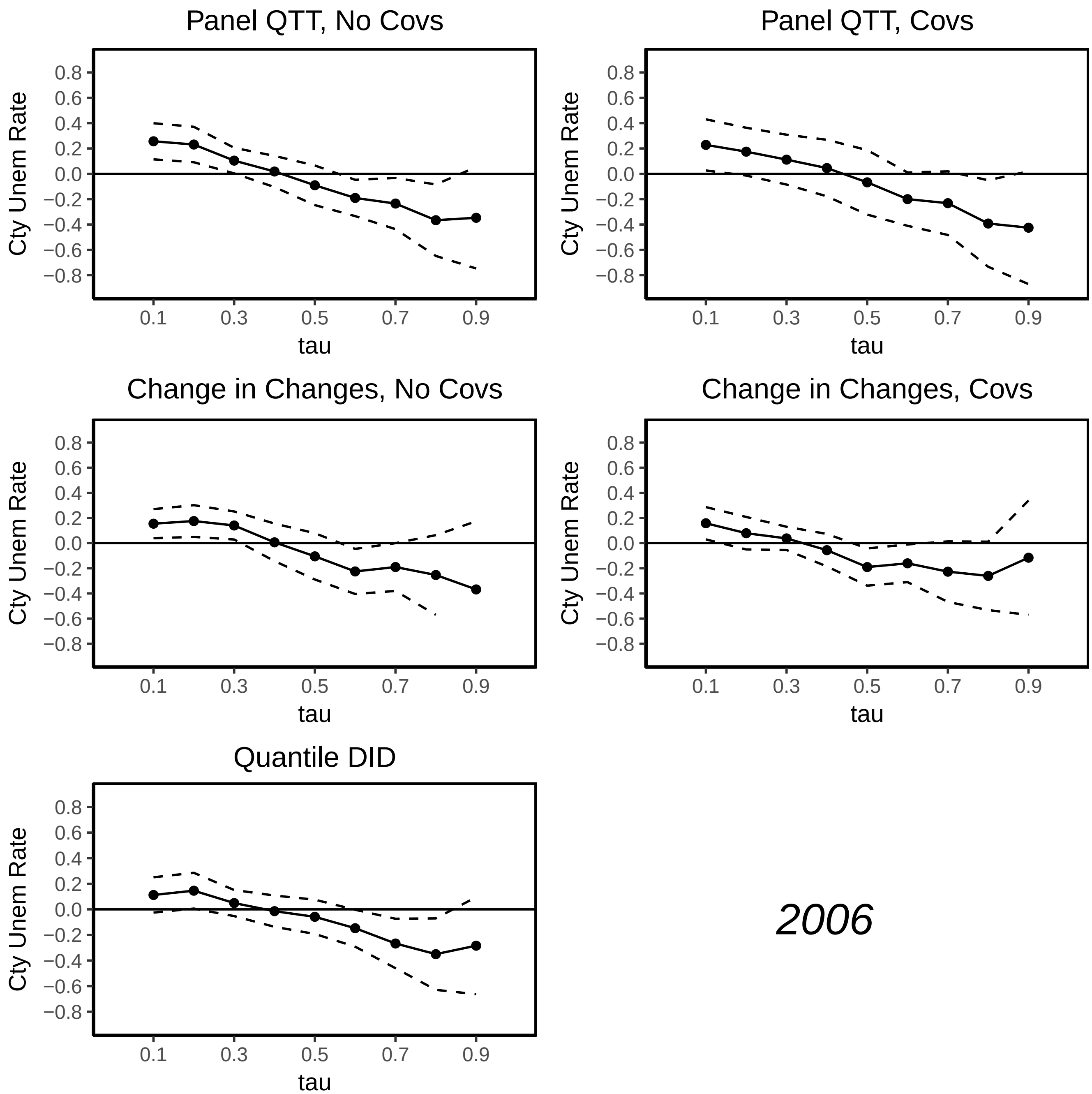
Notes: The figure shows bounds on QTTs when the copula between the change in untreated potential outcomes and the initial level of untreated potential outcomes for the treated group is treated as being completely unknown. The results are obtained using the authors' implementation of the method in Fan and Yu (2012). The figure displays point estimates of the bounds and does not include standard errors or any uncertainty due to sampling.
Sources: Local Area Unemployment Statistics Database from the BLS

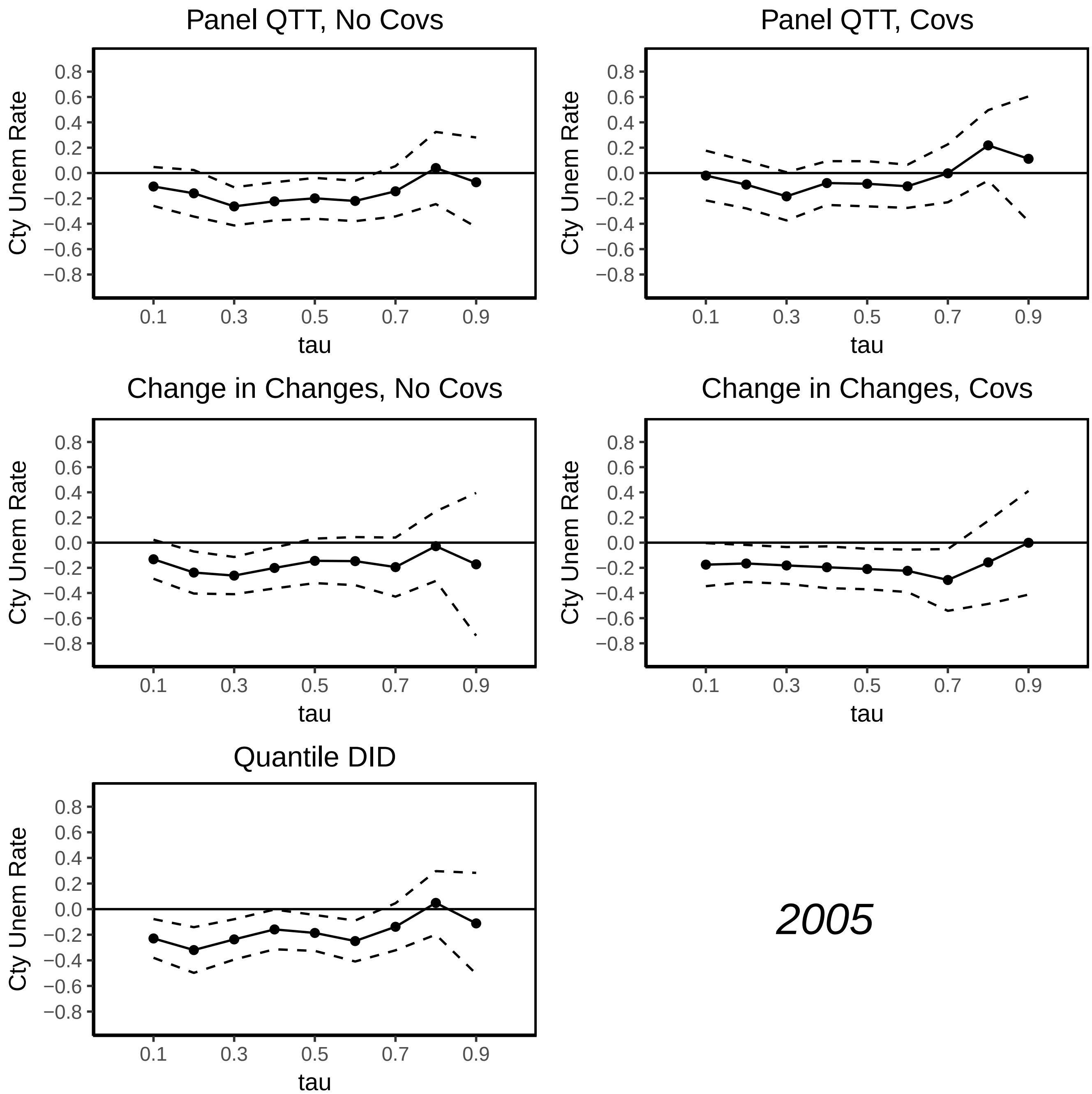
E Supplementary Figures

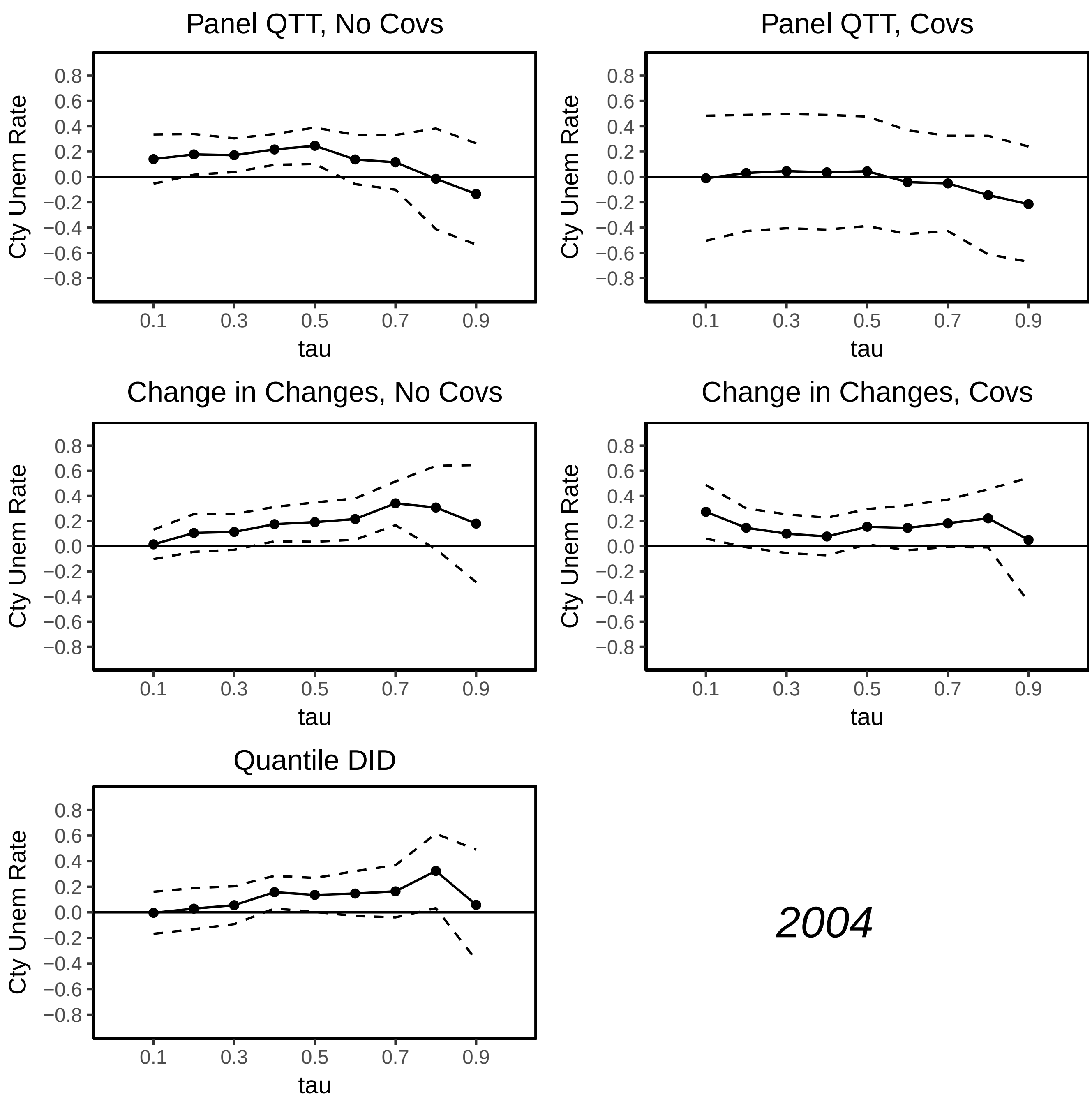
E.1 Change in Changes and Quantile Difference in Differences Estimates for 2007

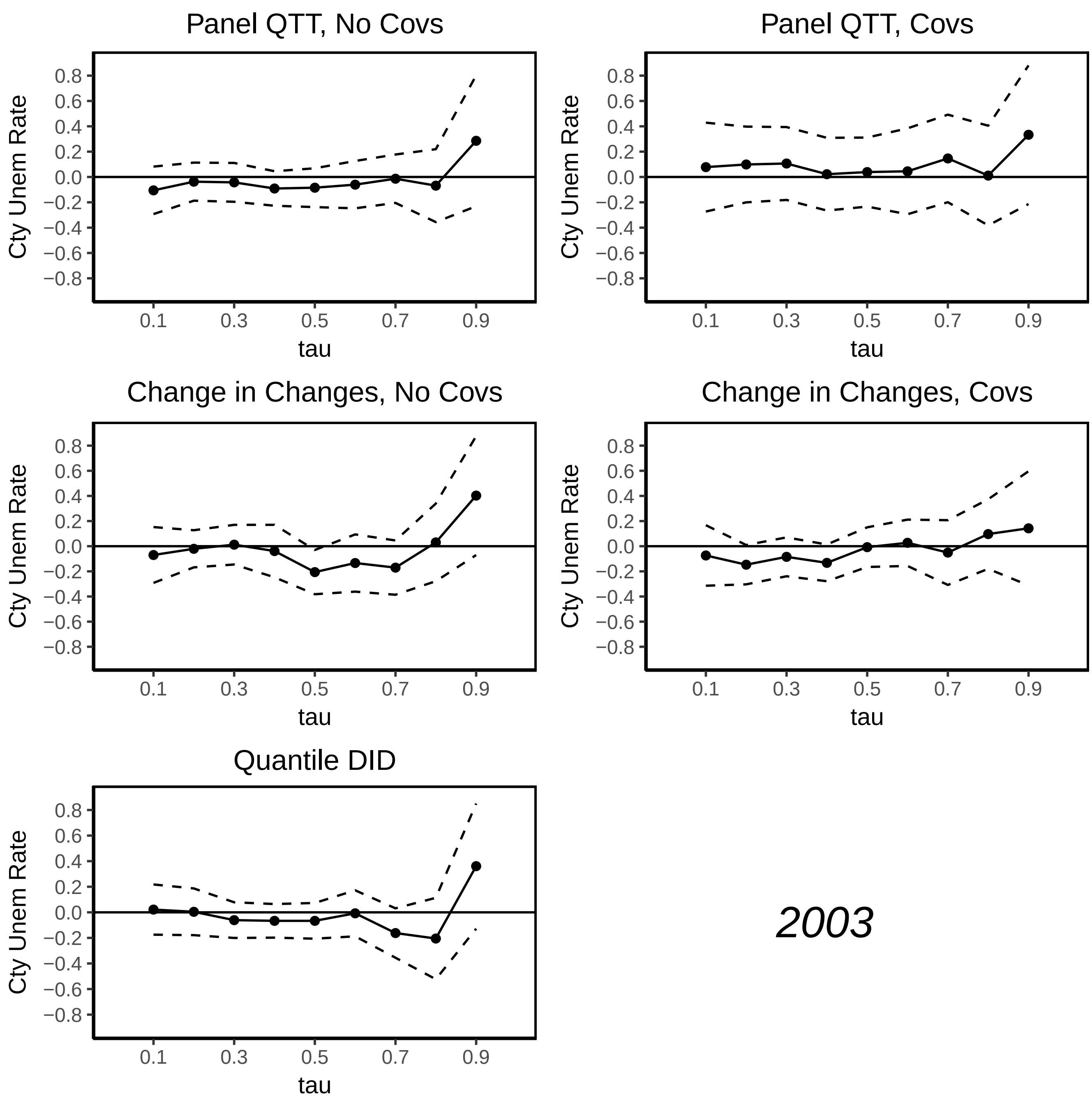


E.2 Pre-Treatment QTT Estimates

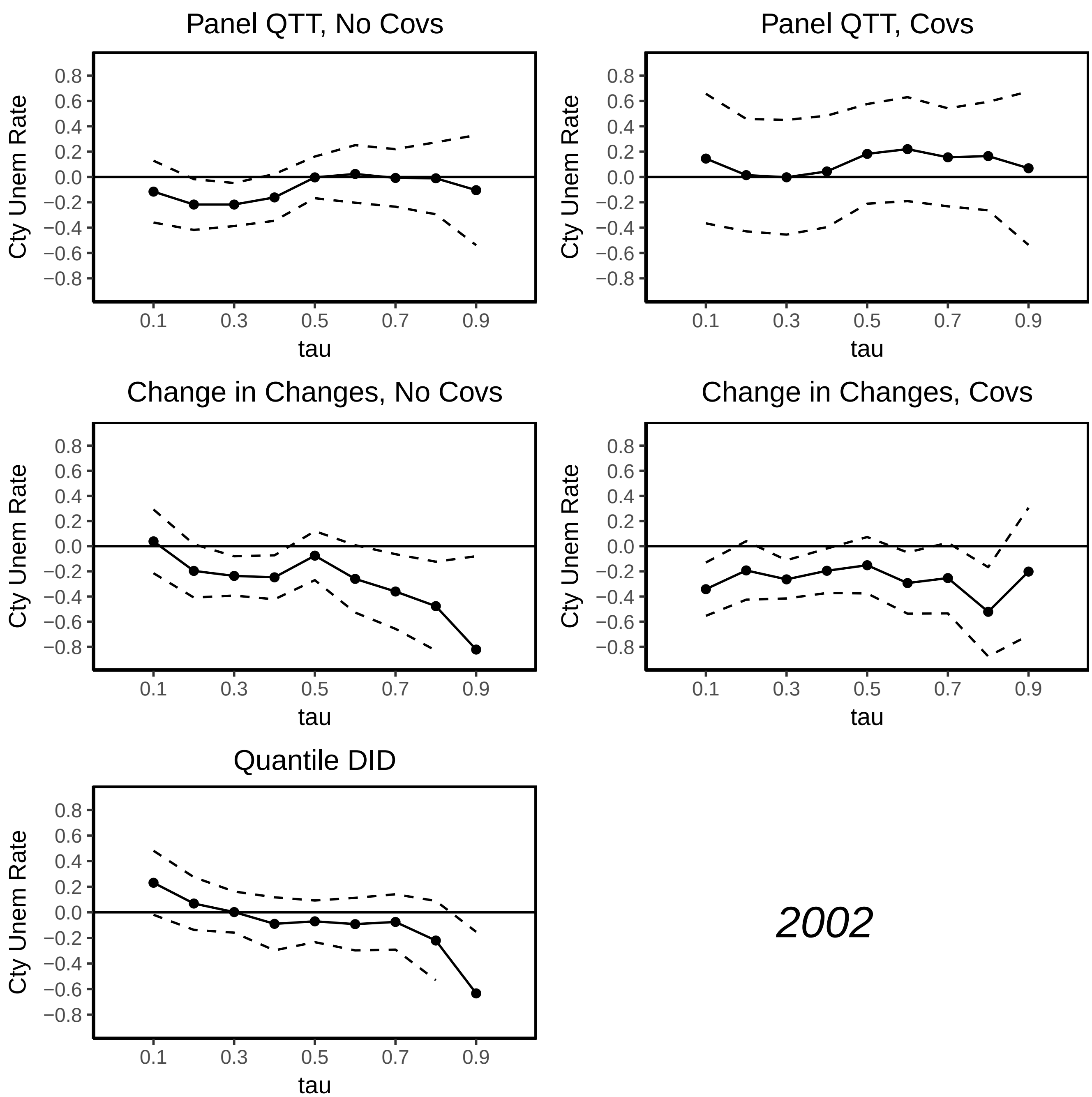



Panel QTT, No Covs
Panel QTT, Covs
Change in Changes, No Covs
Change in Changes, Covs
Quantile DID
Cty Unem Rate
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
−0.6
−0.8
0.1
0.3
0.5
0.7
0.9
tau
2005


Panel QTT, No Covs
Panel QTT, Covs
Change in Changes, No Covs
Change in Changes, Covs
Quantile DID
Cty Unem Rate
tau
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
−0.6
−0.8
0.1
0.3
0.5
0.7
0.9
2004



Panel QTT, No Covs
Panel QTT, Covs
Change in Changes, No Covs
Change in Changes, Covs
Quantile DID
Cty Unem Rate
tau
0.1
0.3
0.5
0.7
0.9
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
−0.6
−0.8
2003


Panel QTT, No Covs
Panel QTT, Covs
Change in Changes, No Covs
Change in Changes, Covs
Quantile DID
Cty Unem Rate
tau
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
−0.6
−0.8
0.1
0.3
0.5
0.7
0.9
2002