Essays on Distributional Treatment Effects with Panel Data

By

Brantly Mercer Callaway, IV

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in

Economics

May, 2016

Nashville, Tennessee

Approved:

Professor Tong Li

Professor William Collins

Professor Federico Gutierrez

Professor Atsushi Inoue

To Sharon

ACKNOWLEDGMENTS

There are many people who have been very important in supporting me academically and personally while I have been in graduate school. I am particularly grateful to members of my dissertation committee. Professors Federico Gutierrez and Atsushi Inoue have been accessible and provided me with much valuable feedback on my work. Professor William Collins taught me Labor Economics and many of the econometrics issues considered in my dissertation are related to problems I first encountered working with him. I also thank Professors Yanqin Fan, Tatsushi Oka, and Pedro Sant'Anna from whom I have learned much about econometrics. I have also benefitted from being the student of many great teachers along the way: Jack Bamford, Vivek Ghosal, and Karl Mealor. I am grateful for financial support from Tong Li and from the Noel Dissertation Fellowship.

It would be difficult to overstate how fortunate I have been to work with my advisor, Tong Li. Although I did not start graduate school planning to study econometrics, working on my dissertation has been a challenging and rewarding experience – I owe this in large part to Professor Li. I am especially thankful for his wisdom and guidance over the years. He has given me the freedom to grow as a researcher from making my own mistakes, but he has always been able to see the big picture. More than once, he has made a suggestion to me that I put out of mind for several months before realizing how useful it would be. I am also grateful for his accessibility, generosity, and friendship.

I am thankful for the support of my family: Mom, Dad, Witt, Paul, and Haley – being proud of me and encouraging me has (always) helped me more than you know. I am also grateful to my grandparents for our many enlightening discussions about the economy. Many thanks to my in-laws Mark and Valerie Pamperin – I appreciate all the patience with me and babysitting over the past years. And I thank my son, Little Brantly, for his willingness to read economics papers (instead of children's books) before bed – we will switch to your books soon. And to our new baby, thank you for listening to me practice my job market and dissertation presentations over the past seven months – I am really hoping that you will arrive in time to see graduation.

And most of all I am thankful to my wife, Sharon. She has been a constant source of love, support, encouragement, and comfort in times of doubt. She has made many sacrifices for me over these years of graduate school without ever counting them as sacrifices. Our time in Nashville has truly been the time of my life.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

Quantile Treatment Effects in Difference in Differences Models with Panel Data[1]

1.1 Introduction

Although most research using program evaluation techniques focuses on estimating the average effect of participating in a program or treatment, in some cases a researcher may be interested in understanding the distributional impacts of treatment participation. For example, for two labor market policies with the same mean impact, policymakers are likely to prefer a policy that tends to increase income in the lower tail of the income distribution to one that tends to increase income in the middle or upper tail of the income distribution. In contrast to the standard linear model, the treatment effects literature explicitly recognizes that the effect of treatment can be heterogeneous across different individuals (Heckman and Robb, 1985; Heckman, Smith, and Clements, 1997). Recently, many methods have been developed that identify distributional treatment effect parameters under common identifying assumptions such as selection on observables (Firpo, 2007), access to a an instrumental variable (Abadie, Angrist, and Imbens, 2002; Chernozhukov and Hansen, 2005; Carneiro and Lee, 2009; Frölich and Melly, 2013), or access to repeated observations over time (Athey and Imbens, 2006; Bonhomme and Sauder, 2011; Chernozhukov, Fernández-Val, Hahn, and Newey, 2013). This chapter 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; Meyer, Viscusi, and Durbin, 1995). The intuition underlying the Difference in Differences approach is that, even after possibly controlling for

[1]This chapter is co-authored with Tong Li (Vanderbilt University)

some covariates, treated agents and untreated agents may still differ from each other in unobserved ways that affect the outcome of interest. These differences render cross-sectional comparisons between individuals with the same covariates unable to identify the true effect of treatment. However, if the effect of these unobserved differences on outcomes is constant over time (this is the so-called "parallel trends" assumption), then the researcher can use the difference between the *change* in outcomes for the treated group and the untreated group (rather than differences in the *level* of outcomes) to identify the ATT.

The first contribution of the current chapter 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. Under this assumption, we are able to identify the entire counterfactual distribution of untreated potential outcomes for the treated group and all of its quantiles.

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. This stands in contrast to other available methods for identifying the QTT such as Quantile Difference in Differences and Change in Changes (Athey and Imbens, 2006). Like the Distributional Difference in Differences Assumption used in this chapter, those models exploit having access to (i) both a treated and control group and (ii) observations at different points in time; however, using these assumptions requires at least somewhat different intuition regarding whether or not they are appropriate. The assumptions used in this chapter neither imply or are implied by the assumptions in those models. But the key distinction is that to employ the assumptions requires familiar reasoning from applied researchers in the case of our model and at least somewhat different reasoning in the case of existing models. On the other hand, Quantile Difference in Differences and Change in Changes models are both available when the researcher has only two periods of data that can be repeated cross sections or panel. To use our method requires at least three periods of panel data.

Although applying a Mean Difference in 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 is able to exploit 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 chapter, the QTT is known to be partially identified (Fan and Yu, 2012) without further assumptions. In practice, these bounds may 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 metioned above, does not change over time. For example, if the change in untreated potential outcomes for the treated group is independent of the initial level of untreated potential outcomes for the treated group, the Copula Stability Assumption says that they will continue to be independent in the next period. 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 this chapter is to extend the results to the case where the identifying assumptions hold conditional on covariates. There are many cases where observed characteristics may affect the path of the untreated outcomes. In this case, if the distribution of characteristics differs between the treated and untreated groups, then the unconditional "parallel trends" assumption is necessarily violated. One example of this phenomenom is the so-called Ashenfelter's dip (Ashenfelter, 1978) where individuals entering a job training program are likely to have experienced a negative transitory shock to wages. Because the shock is transitory, a job training participant's wages are likely to recover even in the absence of job training which implies that using an unconditional Difference in Differences assumption will tend to overstate the effect of the job training program. Conditioning on lags of wages or unemployment histories could help alleviate this problem (Heckman, Ichimura, Smith, and Todd, 1998; Heckman and Smith, 1999; Abadie, 2005). Additionally, if other background characteristics such as education or experience are distributed differently across the treated and untreated groups and the path of wages in the absence of treatment differs by these background characteristics, then an unconditional Difference in Differences assumption will be violated, but a conditional Difference in Differences assumption will be valid.

We also show that a Conditional Copula Stability Assumption holds in a general model of the type that is compatible with the Conditional Distributional Difference in Differences Assumption. Estimation under the Conditional Distributional Difference in Differences

Assumption and the Conditional Copula Stability Assumption is challenging as it involves the nonparametric estimation of several conditional distribution functions and conditional quantile functions for the conditional QTT.[2] To obtain the unconditional QTT additionally requires integrating out the covariates in the identified counterfactual distribution of untreated potential outcomes for the treated group before inverting for the QTT. Under a somewhat stronger set of assumptions – a combination of the Conditional Distributional Difference in Differences Assumption and the Unconditional Copula Stability Assumption – we develop very simple estimators based on a propensity score re-weigthing approach (Hirano, Imbens, and Ridder, 2003; Abadie, 2005; Firpo, 2007). We provide a set of sufficient conditions for this stronger set of assumptions to hold. This combination of assumptions may provide the right balance between generality of assumptions and computational simplicity for much applied work. We derive $\sqrt{n}$-consistency and asymptotic normality for estimation under these assumptions and when the propensity score is estimated parametrically or nonparametrically in a first step.

Having simple identification results when identification holds conditional on some covariates stands in contrast to the existing methods for estimating QTTs. The methods are either (i) unavailable or at least computationally challenging when the researcher desires to make the identifying assumptions conditional on covariates or (ii) require strong parametric assumptions on the relationship between the covariates and outcomes. Because the ATT can be obtained by integrating the QTT and is available under weaker assumptions, a researcher's primary interest in studying the QTT is likely to be in the shape of the QTT rather than the location of the QTT. In this regard, the parametric assumptions required

[2] We primarily focus on the unconditional QTT rather than the conditional QTT though the latter is identified under the current setup. The interpretation of conditional and unconditional quantiles is different as observations at, for example, the lower part of a conditional distribution may or may not be in the lower part of the unconditional distribution. In the job training example in this chapter, if policymakers are most concerned with the impact of job training on individuals in the lower part of the unconditional income distribution, then the unconditional QTT is an appropriate parameter for evaluating the program. In our setup, we are also able to estimate the unconditional QTT at the parametric rate without functional form assumptions, but the conditional QTT could only be estimated at a slower, nonparametric rate. See Firpo, Fortin, and Lemieux (2009) and Frölich and Melly (2013) for more discussion of these issues.

by other methods to accommodate covariates are troubling because nonlinearities or misspecification of the parametric model could easily be confused with the shape of the 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 chapter does not appear to be available for existing methods.

Based on our identification results, estimation of the QTT is straightforward and computationally fast. The estimate of the QTT is consistent and $\sqrt{n}$-asymptotically normal. 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. We discuss parametric, semiparametric, and nonparametric estimation of the propensity score which allows for some flexibility for applied researchers in choosing how to implement the method. We show that under standard conditions the speed of convergence of our estimate of the QTT is not affected by the method chosen for the first stage estimation of the propensity score.

It should be noted that the quantile treatment effects studied in this chapter do not correspond to the distribution or quantile of the treatment effect itself. Because treated and untreated outcomes are never simultaneously observed for any individual, the distribution of the treatment effect is not directly identified. For the QTT, the distribution of treated outcomes for the treated group is compared to the counterfactual distribution of untreated outcomes for the treated group. Even when this counterfactual distribution is identified, unless there is some additional assumption on the dependence between these two distributions (Heckman, Smith, and Clements, 1997; Fan and Park, 2009) or some additional structure placed on the individual's decision on whether or not to be treated (Carneiro, Hansen, and Heckman, 2003; Abbring and Heckman, 2007) the distribution of the treatment effect is not identified. In some cases, knowledge of the quantile treatment effect provides all the information needed to evaluate a program. For example, for social welfare evaluations that do not depend on the identity of the individual – the anonymity condition – quantile treatment effects provide a complete summary of the welfare effects of a policy (Sen, 1997; Carneiro, Hansen, and Heckman, 2001). On the other hand, parameters that depend on the joint distribution of treated and untreated potential outcomes such as the fraction of the population that benefits from treatment are not identified.

We conclude the chapter by comparing the performance of our method with alternative estimators of the QTT: the Quantile Difference in Differences model, the Change in Changes model, and a model based on selection on observables (Firpo, 2007) in an application to estimating the QTT of participating in a job training program using a well known dataset from LaLonde (1986). This dataset contains an experimental component where individuals were randomly assigned to a job training program and an observational component from the Panel Study of Income Dynamics (PSID). It has been used extensively in the literature to measure how well various observational econometric techniques perform in estimating various treatment effect parameters.

Our method is also related to the work on quantile regression with panel data (Koenker, 2004; Abrevaya and Dahl, 2008; Lamarche, 2010; Canay, 2011; Rosen, 2012; Chen, 2015) though our method is distinct in several ways. First, because we do not impose a parametric model, our method allows for the effet of treatment to vary across indiviudals with different covariates in an unspecified way. Second, our method is consistent under fixed-T asymptotics while the papers mentioned above generally require $T \to \infty$.[3] Third, we focus on an unconditional QTT whereas the quantile treatment effects identified by these models are conditional. Moreover, even the conditional quantiles identified using our method are subtly different from the conditional quantiles using panel quantile regression.[4] The difference

[3]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.

[4]Once again, the exception is Abrevaya and Dahl (2008) whose conditional quantiles should be interpreted

is that those conditional quantiles are conditional on the covariates X and the fixed effect; the conditional quantiles in the current chapter are conditional only on the covariates. Finally, of course our method only applies to the case where the researcher is interested in the effect of a binary treatment; quantile regression methods can be employed in cases where one is interested in the effect of a continuous variable on the conditional quantile whereas our method is not available in this case.

Because we focus on nonparametric identifying assumptions, the current chapter is also related to the literature on nonseparable panel data models (Altonji and Matzkin, 2005; Evdokimov, 2010; Bester and Hansen, 2012; Graham and Powell, 2012; Hoderlein and White, 2012; Chernozhukov, Fernández-Val, Hahn, and Newey, 2013). The most similar of these is Chernozhukov, Fernández-Val, Hahn, and Newey (2013) which considers a nonseparable model and, similarly to our case, obtains point identification for observations that are observed in both treated and untreated states. In some ways, the current chapter is more general as (i) we allow for the time trend to be an unrestricted function of the observed covariates that can change over time and (ii) we allow for conditioning on both discrete and continuous regressors. In other ways, their model is more general than ours as it allows for non-continuous outcomes and they also derive bounds on the treatment effect in a dynamic model.

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

The outline of the chapter is as follows. Section 2 provides some background on the notation and setup most commonly used in the treatment effects literature and discusses the various distributional treatment effect parameters estimated in this chapter. Section 3

in the same manner as our conditional quantiles.

provides our main identification result in the case where the Distributional Difference in Differences assumption holds with no covariates. Section 4 extends this result to the case with covariates and provides a propensity score re-weighting procedure to make estimation more feasible. Section 5 details our estimation strategy and the asymptotic properties of our estimation procedure. Section 6 is the empirical example using the job training data. Section 7 concludes.

1.2 Background

This section begins by covering some background, notation, and issues in the treatment effects literature. It then discusses the most commonly estimated treatment effects parameters paying particular attention to distributional treatment effect parameters. Finally, we introduce some background on Difference in Differences: (i) the most common parameters estimated using a Difference in Differences assumption and (ii) the reason why a similar assumption only leads to partial identification of distributional treatment effects.

1.2.1 Treatment Effects Setup

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

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

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

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

1.2.2 Common Treatment Effect Parameters and Identifying Assumptions

The most commonly estimated treatment effect parameters are the Average Treatment Effect (ATE) and the Average Treatment Effect on the Treated (ATT).[5] The unconditional on covariates versions of these are given below:

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

It is also common to estimate versions of ATE and ATT conditional on covariates X. The unconditional ATE and ATT can then be obtained by integrating out X. The parameters

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

provide a summary measure of the average effect of treatment for a random individual in the population (ATE) or for an individual from the subgroup of the population that is treated (ATT).

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

1.2.3 Quantiles and Quantile Treatment Effects

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

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

$$x_\tau = G_X^{-1}(\tau) \equiv \inf\{x : F_X(x) \geq \tau\} \tag{1.1}$$

When X is continuously distributed, x_τ satisfies $P(X \leq x_\tau) = \tau$. An example is the 0.5-quantile – the median.[6] Researchers interested in program evaluation may be interested in other quantiles as well. In the case of the job training program, researchers may be interested in the effect of job training on low income individuals. In this case, they may

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

study the 0.05 or 0.1-quantile. Similarly, researchers studying the effect of a policy on high earners may look at the 0.99-quantile.

Let $F_{Y_{1t}}(y)$ and $F_{Y_{0t}}(y)$ denote the distributions of Y_{1t} and Y_{0t}, respectively. Then, the Quantile Treatment Effect (QTE)[7] is defined as

$$\text{QTE}(\tau) = F^{-1}_{Y_{1t}}(\tau) - F^{-1}_{Y_{0t}}(\tau) \tag{1.2}$$

Analogously to the case of identifying the ATE, QTE is not directly identified because the researcher cannot simultaneously observe Y_{1t} and Y_{0t} for any individual. When treatment is randomized, each distribution will be identified and the quantiles can be recoverd. Similarly, selection on observables also identifies QTE because the marginal distributions of Y_{1t} and Y_{0t} are identified (Firpo, 2007).[8]

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

$$\text{QTT}(\tau) = F^{-1}_{Y_{1t}|D_t=1}(\tau) - F^{-1}_{Y_{0t}|D_t=1}(\tau) \tag{1.3}$$

The QTT is the parameter studied in this chapter. 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.

[7] The QTE was first studied by Doksum (1974) and Lehmann (1974)

[8] There are also several papers that identify versions of QTE when the researcher has an available instrument. See Abadie, Angrist, and Imbens (2002), Chernozhukov and Hansen (2005), and Frölich and Melly (2013).

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

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

ASSUMPTION 1 (Mean Difference in Differences).

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

This is the "parellel trends" assumptions common in applied research. It states that, on average, the unobserved change in untreated potential outcomes for the treated group is equal to the observed change in untreated outcomes for the untreated group. To study the QTT, Assumption 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 due to Fan and Yu (2012) strengthens Assumption 1 and this is the assumption maintained throughout the chapter.

DISTRIBUTIONAL DIFFERENCE IN DIFFERENCES ASSUMPTION.

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

The Distributional Difference in Differences Assumption says that the distribution of the change in potential untreated outcomes does not depend on whether or not the individual belongs to the treatment or the control group. Intuitively, it generalizes the idea of "parallel trends" holding on average to the entire distribution. In applied work, the validity of using a Difference in Differences approach to estimate the ATT hinges on whether the unobserved trend for the treated group can be replaced with the observed trend for the untreated group. This is exactly the same sort of thought experiment that needs to be satisfied for the Distributional Difference in Differences Assumption to hold. Being able to invoke a standard assumption to identify the 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.[9] The resulting bounds are given by

$$\begin{aligned} F_{Y_{0t}|D_t=1}(s) &\leq 1+\min\left[\inf_y F_{(Y_{0t}-Y_{0t-1})|D_t=1}(y)+F_{Y_{0t-1}|D_t=1}(s-y)-1,0\right] \\ F_{Y_{0t}|D_t=1}(s) &\geq \max\left[\sup_y F_{(Y_{0t}-Y_{0t-1})|D_t=1}(y)+F_{Y_{0t-1}|D_t=1}(s-y)-1,0\right] \end{aligned} \tag{1.4}$$

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

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

1.3 Main Results: Identifying QTT in Difference in Differences Models

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

The main theoretical contribution of this chapter 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 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 does not say that these distributions themselves are constant over time; instead, only the dependence between the two marginal distributions is constant over time. We discuss this assumption in more detail and show how it can be used to point identify the 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 $F_{Y_{0t}|D_t=1}(y_{0t})$ (the distribution of interest) could be derived from it. The marginal distributions $F_{\Delta Y_{0t}|D_t=1}(\Delta y_{0t})$ (through the Distributional Difference in Differences assumption) and $F_{Y_{0t-1}|D_t=1}(y_{0t-1})$ (from the data) are both identified. However, because observations are observed separately for untreated and treated individuals, even though each of these marginal distributions are identified from the data, the joint distribution is not identified. Since, from Sklar's Theorem (Sklar, 1959), joint distributions can be expressed as the copula function (capturing the dependence) of the two marginal distributions, the only piece of information that is missing is the copula.[10] We use the idea that the dependence is the same between period t and period $(t-1)$. With this additional information, we can show that $F_{Y_{0t}|D_t=1}(y_{0t})$ is identified.

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

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

where $C_{\Delta Y_{0t},Y_{0t-1}|D_t=1}(\cdot,\cdot)$ is a copula function.[11] Next, we state the second main assummption which replaces the unknown copula with copula for the same outcomes but in the previous period which is identified because no one is treated in the periods before t.

COPULA STABILITY ASSUMPTION.

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

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

[10] Joe (1997), Nelsen (2007), and Joe (2015) are useful references for more details on copulas.

[11] 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.

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

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

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

(a) $\Delta\mathcal{Y}_{t|D_t=0} \subseteq \Delta\mathcal{Y}_{t-1|D_t=1}$

(b) $\mathcal{Y}_{t-1|D_t=1} \subseteq \mathcal{Y}_{t-2|D_t=1}$

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

ASSUMPTION 4. *(Distribution of Y)*

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

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

$$\begin{aligned} &\mathrm{P}(Y_{0t} \leq y | D_t = 1) \\ &\quad = \mathrm{E}\left[1\{ F^{-1}_{\Delta Y_{0t}|D_t=0}(F_{\Delta Y_{0t-1}|D_t=1}(\Delta Y_{0t-1})) \leq y - F^{-1}_{Y_{0t-1}|D_t=1}(F_{Y_{0t-2}|D_t=1}(Y_{0t-2}))\} | D_t = 1 \right] \end{aligned} \tag{1.5}$$

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.

Theorem 1 is the main identification result of the chapter. 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 $\mathrm{P}(Y_{0t} \leq y | D_t = 1) = \mathrm{E}[1\{\Delta Y_{0t} + Y_{0t-1} \leq y\} | D_t = 1]$[12] 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.

It can be estimated by plugging in the sample counterparts of the terms on the right

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

hand side of Equation 1.5:

$$\frac{1}{n_T}\sum_{i\in\mathcal{T}}\left[1\{\hat{\mathrm{F}}^{-1}_{\Delta Y_{0t}|D_t=0}(\hat{\mathrm{F}}_{\Delta Y_{0t-1}|D_t=1}(\Delta Y_{0t-1,i})) \leq y - \hat{\mathrm{F}}^{-1}_{Y_{0t-1}|D_t=1}(\hat{\mathrm{F}}_{Y_{0t-2}|D_t=1}(Y_{0t-2,i}))\}\right] \tag{1.6}$$

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

Pre-Testing the Assumptions

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

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

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

1.4 Allowing for Covariates

The results in the previous section can be extended to the case where both the Distributional Difference in Differences Assumption and the Copula Stability Assumption hold conditional on covariates. In many applications, this combination of assumptions is more likely to hold than the preceding set of unconditional assumptions. First, for particular observations in the treated group, the unobserved path of untreated potential outcomes may be better approximated using observations from the control group that have similar observed characteristics. Second, the dependence between the change in untreated potential outcomes and the initial level of untreated potential outcome for the treated group may be more likely to stay the same over time for observations that have similar characteristics. For example, if the return to some observable characteristic changes over time – a prominent example would be that the return to education has increased over time – then, the Unconditional Copula Stability Assumption will not hold, but a conditional Copula Stability Assumption can continue to hold.

This is a useful contribution as existing methods for estimating the QTT do allow for the outcome distributions to depend on covariates for identification. Athey and Imbens (2006) suggest specifying a parametric model and then performing a type of residualization to recover the QTT. Though this type of procedure is likely to be feasible in applications, using a linear model is likely to be unsatisfactory for studying treatment effect heterogeneity because nonlinearities or model misspecification are likely to be confused with the shape of the QTT.

Making assumptions conditional on covariates also means that one could estimate conditional QTTs. One could obtain the unconditional QTTs, which we have been concerned with, by first integrating the conditional distributions over the observed covariates to form unconditional distributions and then inverting these unconditional distributions. Conditional QTTs could be of interest in their own right as well though nonparametric estimation will suffer from the curse of dimensionality. Finally, a researcher could be interested in the difference between QTTs for different groups defined by some subset of the observed characteristics; one example would be the QTT by gender. These could be obtained by integrating the conditional distributions over the observed covariates that are not of interest only and then inverting these distributions.

We next state the conditional versions of the key identifying assumptions.

CONDITIONAL DISTRIBUTIONAL DIFFERENCE IN DIFFERENCES ASSUMPTION.

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

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.

CONDITIONAL COPULA STABILITY ASSUMPTION.

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

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

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

and

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

which is identified, and

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

and

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

which is identified.

We show in the next section that the set of assumptions required for Theorem 2 is likely to hold in many economic models. One drawback of this estimator, however, is that it is challenging to implement. It requires nonparametric estimation of five conditional distributions or quantile functions, and then requires integrating over X for a function of four of the conditional distributions and quantile functions.

Next, we show that a somewhat stronger combination of assumptions – namely, a combination of the Conditional Distributional Difference in Differences Assumption and the unconditional Copula Stability Assumption – leads to a very simple estimator of the QTT while allowing the unobserved path of untreated outcomes for the treated group to continue to depend on the observed covariates.

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

Using a propensity score re-weighting technique also gives the researcher some flexibility in choosing the best way implement our method. The propensity score can be specified parametrically which requires strong functional form assumptions but is easy to compute and feasible in medium sized samples. At the other extreme, the propensity score could be estimated nonparametrically without invoking functional form assumptions but is more difficult to compute and may suffer from slower convergence depending on the assumptions on the smoothness of the propensity score. Finally, semiparametric methods are available such as Ichimura (1993) and Klein and Spady (1993) that offer some additional flexibility relative to parametric models and computational advantages relative to nonparametric methods.

It should be noted that interest still centers on the unconditional QTT rather than a QTT conditional on X. The role of the covariates is to make the Distributional Difference in Differences Assumption valid. One reason for this focus is that the unconditional QTT is easily interpreted while a conditional QTT may be difficult to interpret and estimate especially when X contains a large number of variables.

By invoking the Conditional Distributional Difference in Differences Assumption rather than the Distributional Difference in Differences Assumption, it is important to note that, for the purpose of identification, the only part of Theorem 1 that needs to be adjusted is the identification of $F_{\Delta Y_{0t}|D_t=1}(\Delta y)$. Under the Distributional Difference in Differences Assumption, this distribution could be replaced directly by $F_{\Delta Y_{0t}|D_t=0}(\Delta y)$; however, now we utilize a propensity score re-weighting technique to replace this distribution with another object (discussed more below). Importantly, all other objects in Theorem 1 can be handled in exactly the same way as they were previously. Particularly, the Copula Stability Assumption continues to hold without needing any adjustment such as conditioning on X. The

Copula Stability Assumption is an assumption on the dependence between $F_{Y_{0t-1}|D_t=1}(y)$ (which is observed) and $F_{\Delta Y_{0t}|D_t=1}(\Delta y)$ which we next show is identified under Conditional Distributional Difference in Differences Assumption. With these two distributions in hand, which do not depend on X, we can once again invoke the same Copula Stability Assumption to obtain identification in the same way as Theorem 1.

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

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

The first part of this assumption says that there is some positive probability that individuals are treated. The second part says that for an individual with any possible value of covariates x, there is some positive probability that he will be treated and a positive probability he will not be treated. This is a standard overlap assumption used in the treatment effects literature.

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

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

where

$$
F_{\Delta Y_{0t}|D_t=1}(y) = \mathrm{E}\left[\frac{1-D_t}{p}\frac{p(X)}{1-p(X)}1\{\Delta Y_t \leq \Delta y\}|D_t = 0\right] \tag{1.7}
$$

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 1.7. Equation 1.7 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 1.7 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 $1\{\Delta Y_t \leq \Delta y\}$ in our case is given by Y_t, ΔY_t, and $1\{Y_t \leq y\}$ in each of the other cases, respectively.

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

$$\hat{\mathrm{P}}(\Delta Y_{0t} \leq \Delta y | D_t = 1) = \frac{1}{n} \sum_{i=1}^{n} \left[\frac{(1-D_t)}{p} \frac{\hat{p}(X_i}{1-\hat{p}(X_i))} 1\{\Delta Y_{it} \leq \Delta y\} \right] \tag{1.8}$$

This analog of the distribution of the change in untreated potential outcomes for the treated group can then be combined with estimates of the other distributions in Theorem 3 to estimate the QTT.

One final thing to notice in this section is that we have written the Conditional Distributional Difference in Differences Assumption in terms of time invariant covariates X, but the assumption can be extended to the case where covariates can change over time denoted X_{it} under standard assumptions. In particular, this extension would require a strict exogeneity assumption such as $\mathrm{P}(Y_{0it} \leq y | X_i, c_i) = \mathrm{P}(Y_{0it} \leq y | X_{it}, c_i)$ where $X_i = (X_{it}, X_{it-1}, X_{it-2})$ is the vector of covariates across all periods, and c_i is an individual specific fixed effect. The strict exogeneity assumption says that conditional on the individual fixed effect and current period values of the covariates, the distribution of untreated potential outcomes does not depend on the value of covariates in other periods. Under the strict exogeneity assumption, a natural version of the Conditional Distributional Difference in Differences Assumption would be $\mathrm{P}(\Delta Y_{0it} \leq \Delta | X_i, c_i, D_t = 1) = \mathrm{P}(\Delta Y_{0it} \leq \Delta | X_{it}, X_{it-1}, D_t = 1)$. Then, in the subsequent analysis, one should replace X with $\tilde{X} = (X_C, X_{it}, X_{it-1})$ where X_C are the covariates that do not change over time. The Conditional Copula Stability Assumption would need to condition on the full vector of covariates (X_t, X_{t-1}, X_{t-2}) so that the dependence between the change in untreated potential outcomes and the initial untreated potential outcome is the same for observations that have the same value of observed covariates in all three periods. Under these assumptions, the proceeding results continue to go through when time varying covariates are present that satisfy a strict exogeneity assumption. We proceed throughout the rest of the chapter, however, using the notation for time invariant X.

How to Interpret the Copula Stability Assumption

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

$$Y_{0it} = g(X_i, \nu_{it}) + h_t(X_i) + m(X_i, \varsigma_i) \tag{1.9}$$

where $g(\cdot)$ is a nonseparable function of observable individual-specific covariates X_i and unobservables (ς_i, ν_{it}) of which ς_i is a vector of time invariant unobservables and ν_{it} is a vector of time varying unobservables; $h_t(\cdot)$ is a time varying function of the observed covariates; and $m(\cdot)$ is a group-specific function of observed covariates and time invariant unobservables that could capture time invariant differences across groups in untreated potential outcomes. We do not put any restrictions on the relationship between X_i and ς_i. And the distribution of $\varsigma_i|X_i$ can differ between the treated and untreated groups. We assume that for all t, $\nu_{it}|X_i, \varsigma_i \sim \mathrm{F}_\nu(\cdot)$; that is, the time varying unobservables are independent of the covariates and time invariant unobservables and their distribution does not change over time. This assumption allows for serial correlation of ν_{it}.

PROPOSITION 1. *In the model of Equation 1.9, the Conditional Copula Stability Assumption is satisfied and the Conditional Distributional Difference in Differences Assumption is satisfied.*

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

This model generalizes many common econometric models. It allows for non-stationarity in outcomes. For example, in the empirical application on job training, aggregate time effects such as macroeconomic shocks are allowed in the model. Several other common models are special cases of this model. For example, the result covers the two-way fixed effects models with individual specific fixed effects and aggregate time fixed effects.

$$Y_{0it} = c_i + \theta_t + X_i\beta + \nu_{it}$$

where c_i is a time invariant fixed effect, θ_t is an aggregate time effect for the treated group, and ν_{it} is white noise. This result also covers a special case of the random trend model (Heckman and Hotz, 1989).[13]

$$Y_{0it} = c_i + g_i t + X_i\beta + \nu_{it}$$

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

where we restrict the random coefficient on the trend g_i to be a constant g across all individuals. Other models are also covered by Proposition 1.

A few additional comments are also in order. The model in Equation 1.9 is quite general and provides a case where both the Conditional Distributional Difference in Differences Assumption and the Copula Stability Assumption hold. We have shown that estimation of the model is much simpler when an unconditional copula stability assumption holds in place of the conditional on covariates assumption we have made thus far. Equation 1.9 violates the unconditional copula stability assumption. However, a sufficient condition for the unconditional copula stability assumpiton is that $h_t(X_i) = h(X_i)\gamma_t t$ where γ_t is a scalar parameter so that $h_t(\cdot)$ is linear in t. In this case, the combination of an unconditional copula stability assumption and the Conditional Distributional Difference in Differences Assumption are valid.

Finally, Equation 1.9 allows us to formalize the restrictions on the model relative to Mean Difference in Differences that the Conditional Distributional Difference in Differences Assumption and the Conditional Copula Stability Assumption require. First, one could consider the very general model

$$Y_{0it} = g_t(X_i, \varsigma_i, \nu_{it})$$

In this situation, the model can change in every period and more structure is required to apply even Mean Difference in Differences. The next model does exactly that.

$$Y_{0it} = g(X_i, \varsigma_i, \nu_{it}) + h_t(X_i, \nu_{it})$$

The key restriction here is that time does not interact with the time invariant unobservables (which are allowed to differ by group). This means that average change in outcomes for the untreated group is equal to the average change in outcomes for the treated group. However, one can show that neither the Conditional Distributional Difference in Differences Assumption nor the Conditional Copula Stability Assumption holds in this case as additional restrictions are needed. Consider the following model

$$Y_{0it} = h_t(X_i, \nu_{it}) + m(X_i, \varsigma_i)$$

where in this model the Mean Difference in Differences Assumption and the Conditional Distributional Difference in Differences Assumption hold, but the Conditional Copula Stability Assumption does not hold. The key extra requirement is to limit how time invariant unobservables (whose distribution can differ across the treated and untreated group) interact with time varying unobservables. In the case where only the Mean Difference in Differences Assumption held, when taking the difference of untreated potential outcomes and allowing for the interaction of ς_i and ν_{it}, the difference averages out to zero, but the distribution itself may not be the same for the treated and untreated groups. In the current model, when considering the term involving time invariant unobservables, the difference is exactly equal to zero. And then, the Conditional Distributional Difference in Differences Assumption holds because ν_{it} has the same distribution in the treated group as in the untreated group. Next, consider

$$Y_{0it} = g(X_i, \varsigma_i, \nu_{it}) + h_t(X_i)$$

where in this model the Mean Difference in Differences Assumption and the Conditional Copula Stability Assumption hold, but the Conditional Distributional Difference in Differences Assumption does not hold. For the Conditional Copula Stability Assumption to hold, the key extra requirement is to limit the interaction time and time-varying variables. Combining the last two models leads to the result in Equation 1.9.

Finally, consider the case where there are time varying covariates and partition the covariates into $X_{it} = (X_{iC}, X_{iVt})$ where X_{iC} are the covariates that are time constant and X_{iVt}

are the covariates that are time varying. Then, one can show that the model given by

$$Y_{0it} = g(X_{iC}, \nu_{it}) + h_t(X_{it}) + m(X_{iC}, \varsigma_i)$$

satisfies both the Conditional Distributional Difference in Differences Assumption and the Conditional Copula Stability Assumption. Although this model severely restricts the way that time varying observed covariates can interact with unobservables, it is still important because it implies that the linear, two-way fixed effects model with time varying regressors and (possibly) time varying coefficients

$$Y_{0it} = c_i + g_i t + X_{it}\beta_t + \nu_{it}$$

satisfies both assumptions.

1.5 Estimation Details

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

We estimate

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

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

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

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

The estimator for $\hat{\text{F}}^{-1}_{Y_{0t}|D_t=1}(\tau)$ is more complicated. The distribution $\hat{\text{F}}_{Y_{0t}|D_t=1}(y_{0t})$ is identified by Distributional Difference in Differences Assumption or as in Theorem 3 depending on the situation. We use this result to provide an estimator of the quantiles of that distribution in the following way:

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

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

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

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

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

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

Then, an estimator of $\mathrm{F}^{-1}_{\Delta Y_{0t}|D_t=1}(\nu)$ can be obtained in the following way. Let $\Delta Y_{t,i}(n)$ denote the ordered values of the change in outcomes from smallest to largest, and let

$\Delta Y_{t,i}(j)$ denote the jth value of $\Delta Y_{t,i}$ in the ordered sequence. Then, $\hat{\mathrm{F}}^{-1}_{\Delta Y_{0t}|D_t=1}(\nu) = \Delta Y_{t,i}(J^*)$ where $J^* = \inf\{J : \frac{1}{n}\sum_{j=1}^{J} \frac{(1-D_{jt})}{p} \frac{\hat{p}(X_j)}{(1-\hat{p}(X_j))} \geq \nu\}$.

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

1.5.1 Inference

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

1.5.1.1 No Covariates Case

In the case with no covariates, the following result holds

THEOREM 4. *Consistency under Distributional Difference in Differences Assumption*

Under Assumption 2, Assumption 3, and Assumption 4

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

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

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

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

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

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

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

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

and

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

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

and

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

THEOREM 5. *Asymptotic Normality*

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

$$\sqrt{n}(\hat{QTT}(\tau) - QTT(\tau)) \xrightarrow{d} N(0,V)$$

where

$$\begin{aligned} V = &\frac{1}{\left\{f_{Y_{0t}|D_t=1}(F^{-1}_{Y_{0t}|D_t=1}(\tau))\right\}^2} V_0 \\ &+ \frac{1}{\left\{f_{Y_{1t}|D_t=1}(F^{-1}_{Y_{1t}|D_t=1}(\tau))\right\}^2} V_1 \\ &- \frac{2}{f_{Y_{0t}|D_t=1}(F^{-1}_{Y_{0t}|D_t=1}(\tau)) \cdot f_{Y_{1t}|D_t=1}(F^{-1}_{Y_{1t}|D_t=1}(\tau))} V_{01} \end{aligned}$$

and

$$\begin{aligned} V_0 = &\frac{1-p}{p^2} \mathrm{E}\left[\mu_{12}(\Delta Y_t; F^{-1}_{Y_{0t}|D_t=1}(\tau))^2 | D_t = 0\right] \\ &+ \frac{1}{p} \mathrm{E}\Big[\Big(\mu_{22}(\Delta Y_{t-1}; F^{-1}_{Y_{0t}|D_t=1}(\tau)) + \mu_{32}(Y_{t-1}; F^{-1}_{Y_{0t}|D_t=1}(\tau)) \\ &\qquad + \mu_{42}(Y_{t-2}; F^{-1}_{Y_{0t}|D_t=1}(\tau)) + \mu_5(\Delta Y_{t-1}, Y_{t-2}; F^{-1}_{Y_{0t}|D_t=1}(\tau))\Big)^2 | D_t = 1\Big] \end{aligned}$$

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

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

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

1.5.1.2 Distributional Difference in Differences Assumption holds conditional on covariates

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

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

ASSUMPTION 7. *(Distribution of X)*

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

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

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

(i) $p(x)$ *is continuously differentiable of order* $s \geq 7r$ *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 9. *(Series Logit Estimator)*

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

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

THEOREM 6. *Consistency under Conditional Distributional Difference in Differences Assumption and when the propensity score is estimated nonparametrically*

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

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

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

$$\Psi_{N12}(z;y) = \mathrm{E}_{Y_{0t-2}|D_t=1}\left\{\frac{p(x)}{(1-p(x))}1\{\Delta \leq y - F^{-1}_{Y_{0t-1}|D_t=1}(F_{Y_{0t-2}|D_t=2}(Y_{0t-2}))\}\right. \tag{1.10}$$

$$\left. - \mathrm{E}_{X,\Delta Y_t|D_t=0}\left[\frac{p(X)}{(1-p(X))}1\{\Delta Y_t \leq y - F^{-1}_{Y_{0t-1}|D_t=1}(F_{Y_{0t-2}|D_t=2}(Y_{0t-2}))\}\right]\right\} \tag{1.11}$$

$$\Psi_{N22}(z;y) = \mathrm{E}_{Y_{0t-2}|D_t=1}\left[\frac{\mathrm{E}[1\{\Delta Y_t \leq y - F^{-1}_{Y_{0t-1}|D_t=1}(F_{Y_{0t-2}|D_t=2}(Y_{0t-2}))|X=x, D_t=0\}]}{(1-p(x))}\right] \tag{1.12}$$

$$\times (d - p(x)) \tag{1.13}$$

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

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

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

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

THEOREM 7. *Asymptotic Normality under Conditional Distributional Difference in Dif-*

ferences Assumption and when the propensity score is estimated nonparametrically

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

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

where

$$\begin{aligned} V_N = &\frac{1}{\left\{ f_{Y_{0t}|D_t=1}(F^{-1}_{Y_{0t}|D_t=1}(\tau)) \right\}^2} V_{0N} \\ &+ \frac{1}{\left\{ f_{Y_{1t}|D_t=1}(F^{-1}_{Y_{1t}|D_t=1}(\tau)) \right\}^2} V_{1N} \\ &- \frac{2}{f_{Y_{0t}|D_t=1}(F^{-1}_{Y_{0t}|D_t=1}(\tau)) \cdot f_{Y_{1t}|D_t=1}(F^{-1}_{Y_{1t}|D_t=1}(\tau))} V_{01N} \end{aligned}$$

and

$$\begin{aligned} V_{0N} = \frac{1}{p^2} \mathrm{E}\Big[\Big(&\Psi_{N22}(Z_i; F^{-1}_{Y_{0t}|D_t=1}(\tau)) + (1 - D_t)\Psi_{N12}(Z_i; F^{-1}_{Y_{0t}|D_t=1}(\tau) \\ &+ D_t \mu_{22}(Y_{t-2}; F^{-1}_{Y_{0t}|D_t=1}(\tau)) + D_t \mu_{32}(Y_{t-2}; F^{-1}_{Y_{0t}|D_t=1}(\tau)) \\ &+ D_t \mu_{42}(Y_{t-2}; F^{-1}_{Y_{0t}|D_t=1}(\tau)) + D_t \mu_5(\Delta Y_{t-1}, Y_{t-2}; F^{-1}_{Y_{0t}|D_t=1}(\tau)) \Big)^2 \Big] \end{aligned}$$

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

$$\begin{aligned} V_{01N} = \frac{1}{p^2} \mathrm{E}\Big[&D_t \psi(Y_{1t}; F^{-1}_{Y_{1t}|D_t=1}(\tau)) \\ &\times \Big(\Psi_{N22}(Z_i; F^{-1}_{Y_{0t}|D_t=1}(\tau)) + \mu_{22}(Y_{t-2}; F^{-1}_{Y_{0t}|D_t=1}(\tau)) + \mu_{32}(Y_{t-2}; F^{-1}_{Y_{0t}|D_t=1}(\tau)) \\ &+ \mu_{42}(Y_{t-2}; F^{-1}_{Y_{0t}|D_t=1}(\tau)) + \mu_5(\Delta Y_{t-1}, Y_{t-2}; F^{-1}_{Y_{0t}|D_t=1}(\tau)) \Big) \Big] \end{aligned}$$

REMARK 1. *When the true propensity score is known up to a finite set of parameters so that* $p(x) = G(x^\top \gamma_0)$ *and* $G(\cdot)$ *is a known function that is typically the cdf of the normal distribution or the logistic function, then consistency and asymptotic normality continue to hold. The proof is identical to the nonparametric case with the following exceptions. First, the propensity score* $p(x)$ *should be replaced everywhere by* $G(x^\top \gamma_0)$*. Second, the following assumption replaces Assumption 8 and Assumption 9.*

ASSUMPTION 10. *(Parametric Propensity Score)*

(i) γ_0 *is an interior point of a compact set* $\Gamma \subset \mathbb{R}^d$

(ii) $E[XX^\top]$ *is non-singular*

(iii) Let $\upsilon = \{x^\top \gamma : x \in \mathcal{X}, \gamma \in \Gamma\}$*. Then, for* $\upsilon \in \Upsilon$*,* $G(\upsilon)$ *is bounded away from 0 and 1, strictly increasing, and continuously differentiable with derivative* $g(\upsilon)$ *that is bounded away from zero and infinity.*

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

$$\begin{aligned}
\Psi_{P22}(z;y) = \mathrm{E}_{Y_{0t-2}|D_t=1}\Bigg\{\mathrm{E}_{X,\Delta Y_t|D_t=0}\Bigg[\frac{1}{1-G(X^\top\gamma_0)}\left(1+\frac{G(X^\top\gamma_0)}{1-G(X^\top\gamma_0)}\right) \\
\times\ 1\{\Delta Y_t \leq y - F^{-1}_{Y_{0t-1}|D_t=1}(F_{Y_{0t-2}|D_t=1}(Y_{0t-2}))\} g(X^\top\gamma_0) X^\top\Bigg]\Bigg\} \\
\times \mathrm{E}_{X,D_t}\left[\left(\frac{D_t - G(X^\top\gamma_0)}{G(X^\top\gamma_0)(1-G(X^\top\gamma_0))}\right)^2 g(X^\top\gamma_0)^2 XX^\top\right]^{-1} \\
\times \frac{d - G(x^\top\gamma_0)}{G(x^\top\gamma_0)(1-G(x^\top\gamma_0))} g(x^\top\zeta_0) x
\end{aligned} \tag{1.14}$$

REMARK 2. *The key step in the the proof for the case without covariates is to show the counterfactual distribution of untreated potential outcomes can be written in the following way*

$$\sqrt{n}(\hat{\mathrm{P}}(Y_{0t} \leq y|D_t = 1) - \mathrm{P}(Y_{0t} \leq y|D_t = 1)) = \sqrt{n}(\hat{\mu}_1 + \hat{\mu}_2 + \hat{\mu}_3 + \hat{\mu}_4 + \hat{\mu}_5) + o_p(1) \tag{1.15}$$

where

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Then, using standard results on V-statistics, Equation 1.15 can be written as

$$
\begin{aligned}
&\sqrt{n}(\hat{\mathrm{P}}(Y_{0t} \le y|D_t = 1) - \mathrm{P}(Y_{0t} \le y|D_t = 1)) \\
&\quad = \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n}\left\{\frac{(1-D_t)}{(1-p)}\mu_{12}(\Delta Y_t; F^{-1}_{Y_{0t}|D_t=1}(\tau))\right.\right. \\
&\qquad\qquad + \frac{D_t}{p}\left[\mu_{22}(\Delta Y_{t-1}; F^{-1}_{Y_{0t}|D_t=1}(\tau)) + \mu_{32}(Y_{t-1}; F^{-1}_{Y_{0t}|D_t=1}(\tau))\right. \\
&\qquad\qquad\qquad \left.\left.\left. + \mu_{42}(Y_{t-2}; F^{-1}_{Y_{0t}|D_t=1}(\tau)) + \mu_5(\Delta Y_{t-1}, Y_{t-2}; F^{-1}_{Y_{0t}|D_t=1}(\tau))\right]\right\}\right) + o_p(1)
\end{aligned}
$$

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

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

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

In this section, we use a well known dataset from LaLonde (1986) that consists of (i) data from randomly assigning job training program applicants to a job training program and (ii) a second dataset consisting of observational data consisting of some individuals who are treated and some who are not treated. This dataset has been widely used in the program evaluation literature. Having access to both a randomized control and an observational control group is a powerful tool for evaluating the performance of observational methods in estimating the effect of treatment. The original contribution of LaLonde (1986) is that many typically used methods (least squares regression, Difference in Differences, and the Heckman selection model) did not perform very well in estimating the average effect of participation in the job training program. An important subsequent literature argued that observational methods can effectively estimate the effect of a job training program, but the results are sensitive to the implementation (Heckman and Hotz, 1989; Heckman, Ichimura, and Todd, 1997; Heckman, Ichimura, Smith, and Todd, 1998; Dehejia and Wahba, 1999; Smith and Todd, 2005). Finally, Firpo (2007) has used this dataset to study the quantile treatment effects of participating in the job training program under the selection on observables assumption.

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

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

1.6.1 Data

The job training data is from the National Supported Work (NSW) Demonstration. The program consisted of providing extensive training to individuals who were unemployed (or working very few hours) immediately prior to participating in the program. Detailed descriptions of the program are available in Hollister, Kemper, and Maynard (1984), LaLonde (1986), and Smith and Todd (2005). Our analysis focuses on the all-male subset used in Dehejia and Wahba (1999). This subset has been the most frequently studied. In particular, Firpo (2007) uses this subset. Importantly for applying the method presented in this chapter, this subset contains data on participant earnings in 1974, 1975, and 1978.[14]

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

Summary statistics for earnings by treatment status (treated, randomized controls, observational controls) are presented in Table 1.1. Average earnings are very similar between the treated group and the randomized control group in the two years prior to treatment. After treatment, average earnings are about $1700 higher for the treated group than the control group indicating that treatment has, on average, a positive effect on earnings. Average earnings for the observational control group are well above the earnings of the treated group in all periods (including the after treatment period).

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

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

of the large differences in earnings outcomes.

1.6.2 Results

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

Main Results

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

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

The PanelQTT method and the CI method admit estimation based on a first step estimate of the propensity score. For both of these methods, we estimate parametric versions of the propensity score using the three specifications mentioned above. Additionally, we also include an additional set of results based on nonparametric estimate of the propensity score using a series logit method. In practice, the PanelQTT method and the CI method use slightly different series logit estimates. For the PanelQTT method, we select the number of approximating terms using a cross-validation method. We use only covariates available from the UNEM covariate set as it would not be appropriate to condition on lags of the dependent variable. We do condition on lags of unemployment. For the CI method, we use the series logit specification used in Firpo (2007). The key difference between the two is that the CI method can condition on lags of the dependent variable real earnings in addition to all the other available covariates.

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

The first section of Table 2 reports estimates of the QTT using the PanelQTT method. The first row provides results where the propensity score is estimated nonparametrically using series logit. The estimated QTT is positive and statistically significant at each of the 0.7, 0.8, and 0.9-quantiles though the estimates tend to be larger than the experimental results. These estimates are statistically different from the experimental results at the 0.8 and 0.9-quantiles. These results also indicate that the QTT is increasing at larger quantiles which is in line with the experimental results. The second row provides results using the COV conditioning set. In our view, this specification is likely to be what an empirical researcher would estimate given the available data and if he were to use the PanelQTT method. Out of all 16 method-covariate set estimates presented in Table 2, the QTTs come closest to matching the experimental results using the PanelQTT method and the COV conditioning set. The point estimate for each of the 0.7, 0.8, and 0.9-quantiles are somewhat smaller than the ATT indicating that the gain from the job training program was either similar across quantiles or slightly at lower income parts of the distribution than at higher income parts of the distribution. The experimental dataset gives precisely the opposite conclusion though: gains at the higher income part of the distribution were somewhat larger than average gains. The difference in conclusions results mainly from a large difference in the estimated ATT[15] and the experimental ATT. When using the UNEM conditioning set, the estimates of the QTT are very similar to the nonparametric specification. Finally, the NO COV conditioning set tends to perform the most poorly. The QTT is estimated to be close to zero at each quantile and is statistically different from the experimental results for the 0.7 and 0.9-quantile.

The second section presents results using cross sectional data. The results in the first row come from estimating the propensity score nonparametrically using series logit where the conditioning set can include lags of the dependent variable real earnings. If we had imposed linearity (and momentarily ignoring the nonparametric estimation of the propensity score), the difference between the CI and the PanelQTT model is that the CI model would include lags of the dependent variable but no fixed effect while the PanelQTT model would include a fixed effect but no lags of the dependent variable. Just as in the case of the linear model, the choice of which model to use depends on the application and the de-

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

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

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

1.7 Conclusion

This chapter 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 show that under conditional versions of both of the main assumptions, the QTT is still identified. Moreover, under the somewhat stronger assumption that the Distributional Difference in Differences Assumption holds conditional on covariates and the Copula Stability Assumption holds unconditionally, we provide very simple estimators of the QTT using propensity score re-weighting. In an application where we compare the results using several available methods to estimate the QTT on observational data to results obtained from an experiment, we find that our method performs at least as well as other available methods.

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

1.8 Proofs

1.8.1 Identification

1.8.1.1 Identification without covariates

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

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

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

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

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

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

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

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

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

$$\begin{aligned} = \int_{\mathcal{Y}_{t-1|D_t=1}} \int_{\Delta\mathcal{Y}_{t|D_t=1}} & 1\{\Delta y_{0t} \leq y - y_{0t-1}\} \\ \times \quad & c_t(F_{\Delta Y_{0t}|D_t=1}(\Delta y_{0t}), F_{Y_{0t-1}|D_t=1}(y_{0t-1})) \\ \times \quad & f_{\Delta Y_{0t}|D_t=1}(\Delta y_{0t}) f_{Y_{0t-1}|D_t=1}(y_{0t-1}) \mathrm{d}\Delta y_{0t} \mathrm{d}y_{0t-1} \end{aligned} \tag{1.16}$$

$$\begin{aligned} = \int_{\mathcal{Y}_{t-1|D_t=1}} \int_{\Delta\mathcal{Y}_{t|D_t=1}} & 1\{\Delta y_{0t} \leq y - y_{0t-1}\} \\ \times \quad & c_{t-1}(F_{\Delta Y_{0t}|D_t=1}(\Delta y_{0t}), F_{Y_{0t-1}|D_t=1}(y_{0t-1})) \\ \times \quad & f_{\Delta Y_{0t}|D_t=1}(\Delta y_{0t}) f_{Y_{0t-1}|D_t=1}(y_{0t-1}) \mathrm{d}\Delta y_{0t} \mathrm{d}y_{0t-1} \end{aligned} \tag{1.17}$$

$$\begin{aligned} = \int_{\mathcal{Y}_{t-1|D_t=1}} \int_{\Delta\mathcal{Y}_{t|D_t=1}} & 1\{\Delta y_{0t} \leq y - y_{0t-1}\} \\ \times \quad & \varphi_{t-1} \left\{ F^{-1}_{\Delta Y_{0t-1}|D_t=1}(F_{\Delta Y_{0t}|D_t=1}(\Delta y_{0t})), F^{-1}_{Y_{0t-2}|D_t=1}(F_{Y_{0t-1}|D_t=1}(y_{0t-1})) \right\} \\ \times \quad & \frac{f_{\Delta Y_{0t}|D_t=1}(\Delta y_{0t})}{f_{\Delta Y_{0t-1}|D_t=1}(F^{-1}_{\Delta Y_{0t-1}|D_t=1}(F_{\Delta Y_{0t}|D_t=1}(\Delta y_{0t})))} \\ \times \quad & \frac{f_{Y_{0t-1}|D_t=1}(y_{0t-1})}{f_{Y_{0t-2}|D_t=1}(F^{-1}_{Y_{0t-2}|D_t=1}(F_{Y_{0t-1}|D_t=1}(y_{0t-1})))} \mathrm{d}\Delta y_{0t} \mathrm{d}y_{0t-1} \end{aligned} \tag{1.18}$$

Equation 1.16 rewrites the joint distribution in terms of the copula pdf using Lemma 1; Equation 1.17 uses the copula stability assumption; Equation 1.18 rewrites the copula pdf as the joint distribution (now in period $t-1$) using Lemma 2.

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

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

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

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

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

Plugging in (1)-(4) in Equation 1.18 and noticing that the substitutions for $\mathrm{d}\Delta y_{0t}$ and $\mathrm{d}y_{0t-1}$ cancel out the fractional terms in the third and fourth lines of Equation 1.18 implies

$$\text{Equation 1.18} = \int_{\mathcal{Y}_{t-2|D_t=1}} \int_{\Delta\mathcal{Y}_{t-1|D_t=1}} 1\{F^{-1}_{\Delta Y_{0t}|D_t=1}(F_{\Delta Y_{0t-1}|D_t=1}(u)) \leq y - F^{-1}_{Y_{0t-1}|D_t=1}(F_{Y_{0t-2}|D_t=1}(v))\} \tag{1.19}$$

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

$$= \mathrm{E}\left[1\{F^{-1}_{\Delta Y_{0t}|D_t=1}(F_{\Delta Y_{0t-1}|D_t=1}(\Delta Y_{0t-1})) \leq y - F^{-1}_{Y_{0t-1}|D_t=1}(F_{Y_{0t-2}|D_t=1}(Y_{0t-2}))\}|D_t=1\right] \tag{1.20}$$

$$= \mathrm{E}\left[1\{F^{-1}_{\Delta Y_{0t}|D_t=0}(F_{\Delta Y_{0t-1}|D_t=1}(\Delta Y_{0t-1})) \leq y - F^{-1}_{Y_{0t-1}|D_t=1}(F_{Y_{0t-2}|D_t=1}(Y_{0t-2}))\}|D_t=1\right] \tag{1.21}$$

where Equation 1.19 follows from the discussion above, Equation 1.20 follows by the definition of expectation, and Equation 1.21 follows from the Distributional Difference in Differences Assumption. □

1.8.1.2 Identification with covariates

In this section, we prove Theorem 3.

Proof. All of the results from the proof of Theorem 1 are still valid. Therefore, all that needs to be shown is that Equation 1.7 holds. Notice,

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

$$= \mathrm{E}\left[\frac{p(X)}{p}\mathrm{P}(\Delta Y_{0t} \leq \Delta y|X, D_t = 0)\right] \tag{1.22}$$

$$= \mathrm{E}\left[\frac{p(X)}{p}\mathrm{E}[(1-D_t)1\{\Delta Y_t \leq \Delta y)\}|X, D_t = 0]\right] \tag{1.23}$$

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

$$= \mathrm{E}\left[\frac{1-D_t}{1-p(X)}\frac{p(X)}{p}1\{\Delta Y_t \leq \Delta y\}\right] \tag{1.24}$$

where Equation 1.22 holds by Conditional Distributional Difference in Differences Assumption. Equation 1.23 holds by replacing $P(\cdot)$ with $\mathrm{E}(1\{\cdot\})$ 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 1.24 simply applies the Law of Iterated Expectations to conclude the proof. $\square$

1.8.2 Proof of Proposition 1

Proof. We are interested in showing that the Copula Stability Assumption holds in the case of the model in Equation 1.9 in Proposition 1. First, recall the definition of the copula for the change in untreated potential outcomes for the treated group and the initial level of untreated potential outcomes for the treated group.

$$\mathrm{C}_{\Delta Y_{0t}, Y_{0t-1}|X, D_t=1}(v, w|x) = \mathrm{P}\left(\mathrm{F}_{\Delta Y_{0t}|X, D_t=1}(\Delta Y_{0t}|x) \leq v, \mathrm{F}_{Y_{0t-1}|X|D_t=1}(Y_{0t-1}|x) \leq w|X = x, D_t = 1\right) \tag{1.25}$$

The model that we consider is the following

$$Y_{0it} = g(X_i, \nu_{it}) + h_t(X_i) + m(X_i, \varsigma_i)$$

which we assume holds for all time periods. We assume that $\nu_{it}|X_i, \varsigma_i \sim \mathrm{F}_\nu(\cdot)$; we place no restrictions on the relationship between X_i and ς_i, and we allow for the distribution of $\varsigma_i|X_i$ to differ across treated and untreated groups. This implies

$$\begin{aligned}
\mathrm{F}_{Y_{0t-1}|X,D_t=1}(y|x) &= \mathrm{P}(Y_{0t-1} \leq y|X = x, D_t = 1) \\
&= \mathrm{P}(g(x, \nu_{it-1}) + h_t(x) + m(x, \varsigma_i) \leq y|X = x, D_t = 1) \\
&= \mathrm{E}_{\varsigma,\nu_{t-1}|X,D_t=1}[1\{g(x, \nu_{t-1}) + h_t(x) + m(x, \varsigma) \leq y\}|X = x, D_t = 1]
\end{aligned}$$

This also implies

$$\mathrm{F}_{Y_{0t-1}|X,D_t=1}(\tilde{Y}_{0t-1}|x) = \mathrm{E}_{\varsigma,\nu_{t-1}|X,D_t=1}[1\{g(x, \nu_{t-1}) + m(x, \varsigma) \leq g(x, \tilde{\nu}_{t-1}) + m(x, \tilde{\varsigma})\}|X = x, D_t = 1] \tag{1.26}$$

and this distribution does not depend on time because the distribution of ν_{t-1} does not change over time, x does not change over time and the functions $g(\cdot)$ and $m(\cdot)$ do not change over time. Similarly,

$$\begin{aligned}
\mathrm{F}_{\Delta Y_{0t}|X,D_t=1}(\Delta|x) &= \mathrm{P}(\Delta Y_{0t} \leq \Delta|X = x, D_t = 1) \\
&= \mathrm{P}(g(x, \nu_{it}) - g(x, \nu_{it-1}) + h_t(x) - h_{t-1}(x) \leq \Delta|X = x, D_t = 1) \quad (1.27) \\
&= \mathrm{E}_{\varsigma,\nu_t,\nu_{t-1}|X,D_t=1}[1\{g(x, \nu_t) - g(x, \nu_{t-1}) + h_t(x) - h_{t-1}(x)\} \leq \Delta\}|X = x, D_t = 1)]
\end{aligned}$$

This implies

$$\mathrm{F}_{\Delta Y_{0t}|X}(\Delta\tilde{Y}_{0t}|x) = \mathrm{E}_{\varsigma,\nu_t,\nu_{t-1}|X,D_t=1}[1\{g(x, \nu_t) - g(x, \nu_{t-1}) \leq g(x, \tilde{\nu}_t) - g(x, \tilde{\nu}_{t-1})\}|X = x, D_t = 1)] \tag{1.28}$$

which does not depend on time because the conditional distribution of ν_t does not change over time. Since neither Equation 1.26 nor Equation 1.28 change over time, the Conditional Copula Stability Assumption holds.

Finally, one can show that

$$\mathrm{P}(\Delta Y_{0t} \leq \Delta | X = x, D_t = 0) = \mathrm{P}(g(x, \nu_{it}) - g(x, \nu_{it-1}) + h_t(x) - h_{t-1}(x) \leq \Delta | X = x, D_t = 0)$$

which is equal to Equation 1.27 because the distribution of (ν_{it}, ν_{it-1}) is independent of whether or not an individual is treated or untreated. This implies that the Conditional Distributional Difference in Differences Assumption holds.

$\square$

1.8.3 Consistency

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

LEMMA 3. *Pointwise convergence of empirical distribution function*

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

LEMMA 4. *Pointwise convergence of empirical quantiles*

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

LEMMA 5. *Uniform convergence of empirical distribution function*

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

LEMMA 6. *Uniform convergence of empirical quantiles*

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

Proof. See Athey and Imbens (2006, Lemma A.3) □

Lemma 7 and Lemma 8 are helpful to work with empirical distributions and empirical quantiles. They are used both in the proof of consistency and in the proof of asymptotic normality.

LEMMA 7.

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

Proof. Because Y is continuously distributed,

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

which implies the result. □

LEMMA 8.

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

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 7 implies the result. □

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

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

Case 1: No covariates

As a first step, we show that $\sup_y |\hat{\mathrm{F}}_{Y_{0t}|D_t=1}(y) - \mathrm{F}_{Y_{0t}|D_t=1}(y)| \xrightarrow{p} 0$. To show this, notice that

$$\sup_y |\hat{\mathrm{F}}_{Y_{0t}|D_t=1}(y) - \mathrm{F}_{Y_{0t}|D_t=1}(y)| \tag{1.29}$$

$$= \sup_y \Bigg| \frac{1}{n_T} \sum_{i \in \mathcal{T}} \left[1\{\hat{F}^{-1}_{\Delta Y_{0t}|D_t=0}(\hat{F}_{\Delta Y_{0t-1}|D_t=1}(\Delta Y_{0t-1,i})) \le y - \hat{F}^{-1}_{Y_{0t-1}|D_t=1}(\hat{F}_{Y_{0t-2}|D_t=1}(Y_{0t-2,i}))\} \right] \tag{1.30}$$

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

$$\le \sup_y \Bigg| \frac{1}{n_T} \sum_{i \in \mathcal{T}} \left[1\{\hat{F}^{-1}_{\Delta Y_{0t}|D_t=0}(\hat{F}_{\Delta Y_{0t-1}|D_t=1}(\Delta Y_{0t-1,i})) \le y - \hat{F}^{-1}_{Y_{0t-1}|D_t=1}(\hat{F}_{Y_{0t-2}|D_t=1}(Y_{0t-2,i}))\} \right] \tag{1.31}$$

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

$$+ \sup_y \Bigg| \frac{1}{n_T} \sum_{i \in \mathcal{T}} \left[1\{F^{-1}_{\Delta Y_{0t}|D_t=0}(\hat{F}_{\Delta Y_{0t-1}|D_t=1}(\Delta Y_{0t-1,i})) \le y - \hat{F}^{-1}_{Y_{0t-1}|D_t=1}(\hat{F}_{Y_{0t-2}|D_t=1}(Y_{0t-2,i}))\} \right] \tag{1.32}$$

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

$$+ \sup_y \Bigg| \frac{1}{n_T} \sum_{i \in \mathcal{T}} \left[1\{F^{-1}_{\Delta Y_{0t}|D_t=0}(F_{\Delta Y_{0t-1}|D_t=1}(\Delta Y_{0t-1,i})) \le y - \hat{F}^{-1}_{Y_{0t-1}|D_t=1}(\hat{F}_{Y_{0t-2}|D_t=1}(Y_{0t-2,i}))\} \right] \tag{1.33}$$

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

$$+\sup_y \left| \frac{1}{n_T} \sum_{i\in\mathcal{T}} \left[1\{F^{-1}_{\Delta Y_{0t}|D_t=0}(F_{\Delta Y_{0t-1}|D_t=1}(\Delta Y_{0t-1,i})) \leq y - F^{-1}_{Y_{0t-1}|D_t=1}(\hat{F}_{Y_{0t-2}|D_t=1}(Y_{0t-2,i}))\} \right] \right. \tag{1.34}$$

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

$$+\sup_y \left| \frac{1}{n_T} \sum_{i\in\mathcal{T}} \left[1\{F^{-1}_{\Delta Y_{0t}|D_t=0}(F_{\Delta Y_{0t-1}|D_t=1}(\Delta Y_{0t-1,i})) \leq y - F^{-1}_{Y_{0t-1}|D_t=1}(F_{Y_{0t-2}|D_t=1}(Y_{0t-2,i}))\} \right] \right. \tag{1.35}$$

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

Next, we show that each of the numbered equations above converges to 0.

Equation 1.31

$$\sup_y \left| \frac{1}{n_T} \sum_{i\in\mathcal{T}} \left[1\{\hat{F}^{-1}_{\Delta Y_{0t}|D_t=0}(\hat{F}_{\Delta Y_{0t-1}|D_t=1}(\Delta Y_{0t-1,i})) \leq y - \hat{F}^{-1}_{Y_{0t-1}|D_t=1}(\hat{F}_{Y_{0t-2}|D_t=1}(Y_{0t-2,i}))\} \right] \right. \tag{1.36}$$

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

$$\leq \sup_z \left| \frac{1}{n_T} \sum_{i\in\mathcal{T}} \left[1\{\hat{F}^{-1}_{\Delta Y_{0t}|D_t=0}(\hat{F}_{\Delta Y_{0t-1}|D_t=1}(\Delta Y_{0t-1,i})) \leq z\} \right] \right. \tag{1.37}$$

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

$$\leq \sup_z |\hat{F}_{\Delta Y_{0t}|D_t=0}(z) - F_{\Delta Y_{0t}|D_t=0}(z)| + o_p(1) \tag{1.38}$$

where Equation 1.38 holds by applying Lemma 7, Lemma 8, and Lemma 15 (below), and the result holds by uniform convergence of empirical distributions as in Lemma 5.

Equation 1.32

$$\sup_y \left| \frac{1}{n_T} \sum_{i\in\mathcal{T}} \left[1\{F^{-1}_{\Delta Y_{0t}|D_t=0}(\hat{F}_{\Delta Y_{0t-1}|D_t=1}(\Delta Y_{0t-1,i})) \leq y - \hat{F}^{-1}_{Y_{0t-1}|D_t=1}(\hat{F}_{Y_{0t-2}|D_t=1}(Y_{0t-2,i}))\} \right] \right. \tag{1.39}$$

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

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

$$\left.-\frac{1}{n_T}\sum_{i\in\mathcal{T}}\left[1\{\Delta Y_{0t-1,i}\leq F^{-1}_{\Delta Y_{0t-1}|D_t=1}(q)\}\right]\right|+o_p(1) \tag{1.40}$$

$$=\sup_{q\in[0,1]}\left|F_{\Delta Y_{0t-1}|D_t=1}(\hat{F}^{-1}_{\Delta_{Y_{0t-1}|D_t=1}}(q))-F_{\Delta Y_{0t-1}|D_t=1}(F^{-1}_{\Delta_{Y_{0t-1}|D_t=1}}(q))\right|+o_p(1) \tag{1.41}$$

$$=\sup_{q\in[0,1]}\left|f_{\Delta Y_{0t-1}|D_t=1}(F^{-1}_{\Delta Y_{0t-1}}(q))(\hat{F}^{-1}_{\Delta_{Y_{0t-1}|D_t=1}}(q)-\hat{F}^{-1}_{\Delta_{Y_{0t-1}|D_t=1}}(q))\right|+o_p(1) \tag{1.42}$$

$$\leq \sup_{q\in[0,1]}\left|f_{\Delta Y_{0t-1}|D_t=1}(F^{-1}_{\Delta Y_{0t-1}}(q))\right|\sup_{q\in[0,1]}\left|\hat{F}^{-1}_{\Delta_{Y_{0t-1}|D_t=1}}(q)-\hat{F}^{-1}_{\Delta_{Y_{0t-1}|D_t=1}}(q)\right|+o_p(1) \tag{1.43}$$

where Equation 1.42 holds by a Taylor Expansion Equation 1.43 applies the Cauchy-Schwarz inequality. The first term in Equation 1.43 is bounded from above by assumption while the second term converges to 0 by Lemma 6.

Equation 1.33

$$\sup_{y}\left|\frac{1}{n_T}\sum_{i\in\mathcal{T}}\left[1\{F^{-1}_{\Delta Y_{0t}|D_t=0}(F_{\Delta Y_{0t-1}|D_t=1}(\Delta Y_{0t-1,i}))\leq y-\hat{F}^{-1}_{Y_{0t-1}|D_t=1}(\hat{F}_{Y_{0t-2}|D_t=1}(Y_{0t-2,i}))\}\right]\right. \tag{1.44}$$

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

$$\leq \sup_{y,q\in[0,1]}\left|\hat{F}_{\Delta Y_{0t-1}|D_t=1}(F^{-1}_{\Delta Y_{0t-1}|D_t=1}(F_{\Delta Y_{0t}|D_t=0}(y-\hat{F}^{-1}_{Y_{0t-1}|D_t=1}(q))))\right. \tag{1.45}$$

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

$$=\sup_{y,q\in[0,1]}\left|F_{\Delta Y_{0t}|D_t=0}(y-\hat{F}^{-1}_{Y_{0t-1}|D_t=1}(q))-F_{\Delta Y_{0t}|D_t=0}(y-F^{-1}_{Y_{0t-1}|D_t=1}(q))\right|+o_p(1) \tag{1.46}$$

$$= \sup_{y,q\in[0,1]} \left| f_{\Delta Y_{0t}|D_t=0}(y - F^{-1}_{Y_{0t-1}|D_t=1}(q))(\hat{F}^{-1}_{Y_{0t-1}|D_t=1}(q) - F^{-1}_{Y_{0t-1}|D_t=1}(q)) \right| + o_p(1) \tag{1.47}$$

$$\leq \sup_{\Delta} \left| f_{\Delta Y_{0t}|D_t=0}(\Delta) \right| \sup_{q\in[0,1]} \left| \hat{F}^{-1}_{Y_{0t-1}|D_t=1}(q) - F^{-1}_{Y_{0t-1}|D_t=1}(q) \right| + o_p(1) \tag{1.48}$$

where Equation 1.46 follows from Lemma 15 (below); Equation 1.47 is a Taylor expansion of Equation 1.46; and Equation 1.48 follows from an application of the Cauchy-Schwarz inequality. The first term in Equation 1.48 is bounded because $f_{\Delta Y_{0t}|D_t=1}(\cdot)$ is bounded; the second term converges to 0 by Lemma 6.

Equation 1.34

$$\sup_{y} \left| \frac{1}{n} \sum 1\{\mathrm{F}^{-1}_{\Delta Y_{0t}|D_t=0}(\mathrm{F}_{\Delta Y_{0t-1}|D_t=1}(\Delta Y_{t-1,i})) \leq y - \mathrm{F}^{-1}_{Y_{0t-1}|D_t=1}(\hat{\mathrm{F}}_{Y_{0t-2}|D_t=1}(Y_{t-2},i))\} \right.$$
$$\left. - \frac{1}{n} \sum 1\{\mathrm{F}^{-1}_{\Delta Y_{0t}|D_t=0}(\mathrm{F}_{\Delta Y_{0t-1}|D_t=1}(\Delta Y_{t-1,i})) \leq y - \mathrm{F}^{-1}_{Y_{0t-1}|D_t=1}(\mathrm{F}_{Y_{0t-2}|D_t=1}(Y_{t-2},i))\} \right|$$

$$\leq \sup_{y,z} |\hat{\mathrm{F}}_{\Delta Y_{0t-1}|D_t=1}(\mathrm{F}^{-1}_{\Delta Y_{0t-1}|D_t=1}(\mathrm{F}_{\Delta Y_{0t}|D_t=0}(y - \mathrm{F}^{-1}_{Y_{0t-1}|D_t=1}(\hat{\mathrm{F}}_{Y_{0t-2}|D_t=1}(z))))) \tag{1.49}$$
$$- \hat{\mathrm{F}}_{\Delta Y_{0t-1}|D_t=1}(\mathrm{F}^{-1}_{\Delta Y_{0t-1}|D_t=1}(\mathrm{F}_{\Delta Y_{0t}|D_t=0}(y - \mathrm{F}^{-1}_{Y_{0t-1}|D_t=1}(\mathrm{F}_{Y_{0t-2}|D_t=1}(z)))))|$$

$$= \sup_{y,z} |\mathrm{F}_{\Delta Y_{0t}|D_t=0}(y - \mathrm{F}^{-1}_{Y_{0t-1}|D_t=1}(\hat{\mathrm{F}}_{Y_{0t-2}|D_t=1}(z))) \tag{1.50}$$
$$- \mathrm{F}_{\Delta Y_{0t}|D_t=0}(y - \mathrm{F}^{-1}_{Y_{0t-1}|D_t=1}(\mathrm{F}_{Y_{0t-2}|D_t=1}(z)))| + o_p(1)$$

$$= \sup_{y,z} \left| - \frac{f_{\Delta Y_{0t}|D_t=0}(y - \mathrm{F}^{-1}_{Y_{0t-1}|D_t=1}(\mathrm{F}_{Y_{0t-2}|D_t=1}(z)))}{f_{Y_{0t-1}|D_t=1}(\mathrm{F}^{-1}_{Y_{0t-1}|D_t=1}(\mathrm{F}_{Y_{0t-2}|D_t=1}(z)))} (\hat{\mathrm{F}}_{Y_{0t-2}|D_t=1}(z) - \mathrm{F}_{Y_{0t-2}|D_t=1}(z)) \right| + o_p(1) \tag{1.51}$$

$$\leq \sup_{\Delta,z} \left| \frac{f_{\Delta Y_{0t}|D_t=0}(\Delta)}{f_{Y_{0t-1}|D_t=1}(\mathrm{F}^{-1}_{Y_{0t-1}|D_t=1}(\mathrm{F}_{Y_{0t-2}|D_t=1}(z)))} \right| \sup_{z} |\hat{\mathrm{F}}_{Y_{0t-2}|D_t=1}(z) - \mathrm{F}_{Y_{0t-2}|D_t=1}(z)| \tag{1.52}$$

where Equation 1.50 holds by an application of Lemma 15 (below). Equation 1.51 is a Taylor expansion of Equation 1.50. Equation 1.52 applies the Cauchy-Schwarz inequality. The first term is bounded because $f_{\Delta Y_{0t}|D_t=0}(\cdot)$ is bounded from above and $f_{Y_{0t-1}|D_t=1}(\cdot)$ is bounded away from 0; and the second term converges to 0 by Lemma 5.

Equation 1.35

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

converges to 0 by the uniform law of large numbers.

Next, $\hat{F}_{Y_{0t}|D_t=1}(y)$ converges to $F_{Y_{0t}|D_t=1}(y)$ uniformly in y implies

$$\hat{F}_{Y_{0t}|D_t=1}(y) \xrightarrow{a.s.} F_{Y_{0t}|D_t=1}(y)$$

for all y. Let $V \sim \Phi(\cdot) \equiv N(0,1)$. This implies that for $q \in (0,1)$,

$$\Phi(\hat{F}^{-1}_{Y_{0t}|D_t=1}(q)) = \mathrm{P}(\hat{F}_{Y_{0t}|D_t=1}(V) \le q) \tag{1.53}$$

Moreover, because $\hat{F}_{Y_{0t}|D_t=1}(y) \xrightarrow{a.s.} F_{Y_{0t}|D_t=1}(y)$,

$$\mathrm{P}(\hat{F}_{Y_{0t}|D_t=1}(V) \le q) \xrightarrow{a.s.} \mathrm{P}(F_{Y_{0t}|D_t=1}(V) \le q) \tag{1.54}$$

Then, applying the continuous transformation $\Phi^{-1}(\cdot)$ to the left hand side of Equation 1.53 and to the right hand side of Equation 1.54 implies $\hat{F}^{-1}_{Y_{0t}|D_t=1}(\tau) \xrightarrow{a.s.} F^{-1}_{Y_{0t}|D_t=1}(\tau)$. The result then follows by the convergence of $\hat{F}^{-1}_{Y_{1t}|D_t=1}(\tau)$ and Slutsky's Lemma.

Case 2: Covariates

The preceding results will continue to go through provided we show two additional things (i) $\sup_\Delta |\hat{F}_{\Delta Y_{0t}|D_t=1}(\Delta) - F_{\Delta Y_{0t}|D_t=1}(\Delta)| \xrightarrow{p} 0$ and (ii) a result similar to Lemma 8 that allows us to move the empirical quantiles of this distribution to the other side of the inequality inside of an indicator function.

For (i), notice that

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

$$\leq \sup_{\Delta} \left| \frac{1}{n} \sum_{i=1}^{n} \frac{1-D_{it}}{p} \frac{\hat{p}(X_i)}{1-\hat{p}(X_i)} 1\{\Delta Y_{it} \leq \Delta\} - \frac{1}{n} \sum_{i=1}^{n} \frac{1-D_{it}}{p} \frac{p(X_i)}{1-p(X_i)} 1\{\Delta Y_{it} \leq \Delta\} \right| \tag{1.55}$$

$$+ \sup_{\Delta} \left| \frac{1}{n} \sum_{i=1}^{n} \frac{1-D_{it}}{p} \frac{p(X_i)}{1-p(X_i)} 1\{\Delta Y_{it} \leq \Delta\} - \mathrm{E}\left[\frac{1-D_t}{p} \frac{p(X)}{1-p(X)} 1\{\Delta Y_t \leq \Delta\} \right] \right| \tag{1.56}$$

Notice that Equation 1.55 is equal to

$$\sup_{\Delta} \left| \frac{1}{n} \sum_{i=1}^{n} \frac{1-D_{it}}{p} \left(\frac{\hat{p}(X_i) - p(X_i)}{(1-\hat{p}(X_i))(1-p(X_i))} \right) 1\{\Delta Y_{it} \leq \Delta\} \right|$$

$$\leq C \sup_{x} |\hat{p}(x) - p(x)| \xrightarrow{p} 0$$

which follows because of the uniform convergence of the estimated propensity score, p is bounded away from 0 by Assumption 8 , $p(\cdot)$ is bounded away from 1 by Assumption 8, and $\hat{p}(\cdot)$ is bounded away from 1 with probability 1 by the uniform convergence of the of the estimated propensity score. The uniform convergence of the propensity score estimated by series logit under identical conditions to those in the current chapter is established in Hirano, Imbens, and Ridder (2003, Lemma 1). Uniform convergence of the propensity score when it is estimated parametrically is guaranteed by the conditions of Assumption 10.

Equation 1.56 converges to 0 by the uniform law of large numbers.

For (ii), we first show that

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

This follows because

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

where C is an arbitrary constant and the result holds because the difference is equal to 0 if $q \in Range(\hat{F}_{\Delta Y_{0t}|D_t=1})$ 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 8. $\square$

1.8.4 Asymptotic Normality

In this section we derive the asymptotic distribution of $\hat{\text{QTT}}$. We make use of several lemmas in this section and state these first.

LEMMA 9. *(i)* $\sqrt{n}(\hat{F}_X(x) - F_X(x)) \xrightarrow{d} N(0, p(1-p))$ *where* $p = F_X(x)$*, and (ii)* $\sqrt{n}(\hat{F}_X(x) - F_X(x)) = \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} 1\{X_i \leq x\} - \mathrm{E}[1\{X_i \leq x\}]\right)$.

LEMMA 10. *For* $0 \leq \tau \leq 1$*, (i)* $\sqrt{n}(\hat{F}_X^{-1}(\tau) - F_X^{-1}(\tau)) \xrightarrow{d} N\left(0, \frac{\tau(1-\tau)}{f_X^2(F_X^{-1}(\tau))}\right)$*, and (ii)* $\sqrt{n}(\hat{F}_X^{-1}(\tau) - F_X^{-1}(\tau)) = \frac{1}{f_X(F_X^{-1}(\tau))}\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} 1\{X_i \leq F_X^{-1}(\tau)\} - \mathrm{E}[1\{X_i \leq F_X^{-1}(\tau)\}]\right) + o_p(1)$

Proof. See Van der Vaart (2000, pp. 307-308) $\square$

LEMMA 11. *Uniform Convergence of empricial distribution For* $0 \leq \delta < 1/2$*,* $\sup_x n^{\delta}|\hat{F}_X(x) - F_X(x)| \xrightarrow{p} 0$

LEMMA 12. *Uniform Convergence of empricial quantiles For* $0 \leq \delta < 1/2$*,* $\sup_{q \in (0,1)} n^{\delta}|\hat{F}_X^{-1}(q) - F_X^{-1}(q)| \xrightarrow{p} 0$

LEMMA 13. $\sqrt{n}(\hat{F}_X(\hat{F}_X^{-1}(\tau)) - \tau) \xrightarrow{p} 0$

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

LEMMA 14. $\sup_x \sqrt{n}|\hat{F}_X(x) - F_X(x) - f_X(x)(\hat{F}_X^{-1}(F_X(x)) - x)| \xrightarrow{p} 0$

Proof. The result holds because

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

where the second equality uses Lemma 4. □

LEMMA 15. *For* $\delta > 1/2$, $y \in \mathcal{Y}$, $(y+x) \in \mathcal{Y}$, $\sup_{x \leq n^{-\delta}} \sqrt{n}|\hat{F}_Y(y+x) - \hat{F}_Y(y) - (F_Y(y+x) - F_Y(y))| \xrightarrow{p} 0$.

Proof. This is a special case of Lemma A.5 in Athey and Imbens (2006) □

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

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

Proof. First, note that by Taylor expansions

$$F_Z(\hat{F}_Y^{-1}(\hat{F}_X(x))) - F_Z(F_Y^{-1}(\hat{F}_X(x))) = f_Z(F_Y^{-1}(\hat{F}_X(x)))(\hat{F}_Y^{-1}(\hat{F}_X(x)) - F_Y^{-1}(\hat{F}_X(x))) + o_p(1) \tag{1.57}$$

and

$$\begin{aligned}
F_Z(\hat{F}_Y^{-1}(F_X(x))) - F_Z(F_Y^{-1}(F_X(x))) &= f_Z(F_Y^{-1}(F_X(x)))(\hat{F}_Y^{-1}(F_X(x)) - F_Y^{-1}(F_X(x))) + o_p(1) \\
&= \left(f_Z(F_Y^{-1}(F_X(x))) - f_Z(F_Y^{-1}(\hat{F}_X(x)))\right)(\hat{F}_Y^{-1}(F_X(x)) - F_Y^{-1}(F_X(x))) \\
&\quad + f_Z(F_Y^{-1}(\hat{F}_X(x)))(\hat{F}_Y^{-1}(F_X(x)) - F_Y^{-1}(F_X(x)))) + o_p(1)
\end{aligned} \tag{1.58}$$

This implies that

$$
\begin{aligned}
&\sup_x \sqrt{n} \left| F_Z(\hat{F}_Y^{-1}(\hat{F}_X(x))) - F_Z(F_Y^{-1}(\hat{F}_X(x))) - \left\{ F_Z(\hat{F}_Y^{-1}(F_X(x))) - F_Z(F_Y^{-1}(F_X(x))) \right\} \right| \\
&\quad \le \sup_x \sqrt{n} \left| f_Z(F_Y^{-1}(\hat{F}_X(x))) \left(\hat{F}_Y^{-1}(\hat{F}_X(x)) - F_Y^{-1}(\hat{F}_X(x))) - \left\{ \hat{F}_Y^{-1}(F_X(x)) - F_Y^{-1}(F_X(x))) \right\} \right) \right| \\
&\qquad + \sup_x \sqrt{n} \left| \left(f_Z(F_Y^{-1}(F_X(x))) - f_Z(F_Y^{-1}(\hat{F}_X(x))) \right) (\hat{F}_Y^{-1}(F_X(x)) - F_Y^{-1}(F_X(x))) \right| + o_p(1)
\end{aligned}
\tag{1.59}
$$

$$\xrightarrow{p} 0$$

The first term in Equation 1.59 converges to 0 because $f_Z(F_Y^{-1}(\hat{F}_X(x)))$ is bounded by assumption and Lemma 15 implies $\sqrt{n} \left(\hat{F}_Y^{-1}(\hat{F}_X(x)) - F_Y^{-1}(\hat{F}_X(x))) - \left\{ \hat{F}_Y^{-1}(F_X(x)) - F_Y^{-1}(F_X(x))) \right\} \right)$ converges to 0. For the second term, $(\hat{F}_Y^{-1}(F_X(x)) - F_Y^{-1}(F_X(x)))$ is clearly $O_p(1/\sqrt{n})$. The term $\left(f_Z(F_Y^{-1}(F_X(x))) - f_Z(F_Y^{-1}(\hat{F}_X(x))) \right)$ is also $O_p(1/\sqrt{n})$ which can be seen by taking a Taylor approximation and using the assumptions that $f_Y(y)$ and $\frac{\partial f_Z}{\partial z}(z)$ are bounded. This implies the result.

$$
\begin{aligned}
&\sqrt{n} \left(\hat{F}_Y^{-1}(\hat{F}_X(x)) - F_Y^{-1}(\hat{F}_X(x))) - \left\{ \hat{F}_Y^{-1}(F_X(x)) - F_Y^{-1}(F_X(x))) \right\} \right) \\
&\quad = \frac{1}{f_Y(F_Y^{-1}(\hat{F}_X(x)))} \left\{ \hat{F}_Y(F_Y^{-1}(\hat{F}_X(x))) - F_Y(F_Y^{-1}(\hat{F}_X(x))) \right\} \\
&\qquad - \frac{1}{f_Y(F_Y^{-1}(F_X(x)))} \left\{ \hat{F}_Y(F_Y^{-1}(F_X(x))) - F_Y(F_Y^{-1}(F_X(x))) \right\}
\end{aligned}
$$

□

LEMMA 17. *One Sample V-Statistic*

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

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

LEMMA 18. *Two Sample V-Statistic*

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

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

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

LEMMA 19. *Asymptotic Representation of* $\sqrt{n}\left(\hat{P}(Y_{0t} \leq y|D_t = 1) - P(Y_{0t} \leq y|D_t = 1)\right)$

Let $\hat{\mu}_1$*,* $\hat{\mu}_2$*,* $\hat{\mu}_3$*,* $\hat{\mu}_4$*, and* $\hat{\mu}_5$ *be defined as in the main text and restated here:*[16]

$$\begin{aligned}\hat{\mu}_1 &= \frac{1}{n_T n_C}\sum_{i\in\mathcal{T}}\sum_{j\in\mathcal{C}} 1\{\Delta Y_{0t,j} \leq (y - F^{-1}_{Y_{0t-1}|D_t=1}(F_{Y_{0t-2}|D_t=1}(Y_{0t-2,i})))\} \\ &\qquad - F_{\Delta Y_{0t}|D_t=0}(y - F^{-1}_{Y_{0t-1}|D_t=1}(F_{Y_{0t-2}|D_t=1}(Y_{0t-2,i}))) \\ &\equiv \frac{1}{n_T n_C}\sum_{i\in\mathcal{T}}\sum_{j\in\mathcal{C}} \phi_1(Y_{0t-2,i}, \Delta Y_{0t,j})\end{aligned}$$

$$\begin{aligned}\hat{\mu}_2 &= \frac{1}{n_T^2}\sum_{i\in\mathcal{T}}\sum_{j\in\mathcal{T}} 1\{\Delta Y_{0t-1,j} \leq F^{-1}_{\Delta Y_{0t-1|D_t=1}}(F_{\Delta Y_{0t}|D_t=0}(y - F^{-1}_{Y_{0t-1}|D_t=1}(F_{Y_{0t-2|D_t=1}}(Y_{0t-2,i}))))\} \\ &\qquad - F_{\Delta Y_{0t-1}|D_t=1}(F^{-1}_{\Delta Y_{0t-1|D_t=1}}(F_{\Delta Y_{0t}|D_t=0}(y - F^{-1}_{Y_{0t-1}|D_t=1}(F_{Y_{0t-2}|D_t=1}(Y_{0t-2,i}))))] \\ &\equiv \frac{1}{n_T^2}\sum_{i\in\mathcal{T}}\sum_{j\in\mathcal{T}} \phi_2(Y_{t-2,i}, \Delta Y_{t-1,j})\end{aligned}$$

$$\begin{aligned}\hat{\mu}_3 &= \frac{1}{n_T^2}\sum_{i\in\mathcal{T}}\sum_{j\in\mathcal{T}} \frac{f_{\Delta Y_{0t}|D_t=0}(y - F^{-1}_{Y_{0t-1}|D_t=1}(F_{Y_{0t-2}|D_t=1}(Y_{0t-2,i})))}{f_{Y_{0t-1}|D_t=1}(F_{Y_{0t-2}|D_t=1}(Y_{0t-2,i}))} \\ &\qquad \times \Big(1\{Y_{0t-1,j} \leq F^{-1}_{Y_{0t-1}|D_t=1}(F_{Y_{0t-2}|D_t=1}(Y_{0t-2,i}))\} \\ &\qquad\qquad - F_{Y_{0t-1}|D_t=1}(F^{-1}_{Y_{0t-1}|D_t=1}(F_{Y_{0t-2}|D_t=1}(Y_{0t-2,i})))\Big)\end{aligned}$$

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

$$\equiv \frac{1}{n_T^2} \sum_{i \in \mathcal{T}} \sum_{j \in \mathcal{T}} \phi_3(Y_{t-1,i}, Y_{t-2,j})$$

$$\hat{\mu}_4 = \frac{1}{n_T^2} \sum_{i \in \mathcal{T}} \sum_{j \in \mathcal{T}} \frac{f_{\Delta Y_{0t}|D_t=0}(y - F^{-1}_{Y_{0t-1}|D_t=1}(F_{Y_{0t-2}|D_t=1}(Y_{0t-2,i})))}{f_{Y_{0t-1}|D_t=1}(F^{-1}_{Y_{0t-1}|D_t=1}(F_{Y_{0t-2}|D_t=1}(Y_{0t-2,i})))}$$
$$\times \left(1\{Y_{0t-2}, j \leq Y_{0t-2,i}\} - F_{Y_{0t-2}|D_t=1}(Y_{0t-2,i})]\right)$$
$$\equiv \frac{1}{n_T^2} \sum_{i \in \mathcal{T}} \sum_{j \in \mathcal{T}} \phi_4(Y_{t-2,i}, Y_{t-2,j})$$

$$\hat{\mu}_5 = \frac{1}{n_T} \sum_{i \in \mathcal{T}} \left[1\{F^{-1}_{\Delta Y_{0t}|D_t=0}(F_{\Delta Y_{0t-1}|D_t=1}(\Delta Y_{0t-1,i})) \leq y - F^{-1}_{Y_{0t-1}|D_t=1}(F_{Y_{0t-2}|D_t=1}(Y_{0t-2,i}))\}\right]$$
$$- \mathrm{E}\left[1\{F^{-1}_{\Delta Y_{0t}|D_t=0}(F_{\Delta Y_{0t-1}|D_t=1}(\Delta Y_{0t-1,i})) \leq y - F^{-1}_{Y_{0t-1}|D_t=1}(F_{Y_{0t-2}|D_t=1}(Y_{0t-2,i}))\}|D_t = 1\right]$$
$$\equiv \frac{1}{n_T} \sum_{i \in \mathcal{T}} \phi_5(\Delta Y_{t-1,i}, Y_{0t-2,i})$$

Then,

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

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

$$\sqrt{n}\left(\hat{P}(Y_{0t} \leq y|D_t = 1) - P(Y_{0t} \leq y|D_t = 1) - \hat{\mu}_1 - \hat{\mu}_2 - \hat{\mu}_3 - \hat{\mu}_4 - \hat{\mu}_5\right)$$
$$= \sqrt{n}\left(\frac{1}{n_T} \sum_{i \in \mathcal{T}} \left[1\{\hat{F}^{-1}_{\Delta Y_{0t}|D_t=0}(\hat{F}_{\Delta Y_{0t-1}|D_t=1}(\Delta Y_{0t-1,i})) \leq y - \hat{F}^{-1}_{Y_{0t-1}|D_t=1}(\hat{F}_{Y_{0t-2}|D_t=1}(Y_{0t-2,i}))\}\right]\right.$$
$$- \mathrm{E}\left[1\{F^{-1}_{\Delta Y_{0t}|D_t=0}(F_{\Delta Y_{0t-1}|D_t=1}(\Delta Y_{0t-1,i})) \leq y - F^{-1}_{Y_{0t-1}|D_t=1}(F_{Y_{0t-2}|D_t=1}(Y_{0t-2,i}))\}|D_t = 1\right]$$
$$\left. - \hat{\mu}_1 - \hat{\mu}_2 - \hat{\mu}_3 - \hat{\mu}_4 - \hat{\mu}_5\right) \tag{1.60}$$
$$= \sqrt{n}\left(\frac{1}{n_T} \sum_{i \in \mathcal{T}} \left[1\{\hat{F}^{-1}_{\Delta Y_{0t}|D_t=0}(\hat{F}_{\Delta Y_{0t-1}|D_t=1}(\Delta Y_{0t-1,i})) \leq y - \hat{F}^{-1}_{Y_{0t-1}|D_t=1}(\hat{F}_{Y_{0t-2}|D_t=1}(Y_{0t-2,i}))\}\right]\right.$$

$$- \frac{1}{n_T} \sum_{i \in \mathcal{T}} \left[1\{F^{-1}_{\Delta Y_{0t}|D_t=0}(\hat{F}_{\Delta Y_{0t-1}|D_t=1}(\Delta Y_{0t-1,i})) \leq y - \hat{F}^{-1}_{Y_{0t-1}|D_t=1}(\hat{F}_{Y_{0t-2}|D_t=1}(Y_{0t-2,i}))\} \right] - \hat{\mu}_1 \Bigg) \tag{1.61}$$

$$+ \sqrt{n} \Bigg(\frac{1}{n_T} \sum_{i \in \mathcal{T}} \left[1\{F^{-1}_{\Delta Y_{0t}|D_t=0}(\hat{F}_{\Delta Y_{0t-1}|D_t=1}(\Delta Y_{0t-1,i})) \leq y - \hat{F}^{-1}_{Y_{0t-1}|D_t=1}(\hat{F}_{Y_{0t-2}|D_t=1}(Y_{0t-2,i}))\} \right]$$
$$- \frac{1}{n_T} \sum_{i \in \mathcal{T}} \left[1\{F^{-1}_{\Delta Y_{0t}|D_t=0}(F_{\Delta Y_{0t-1}|D_t=1}(\Delta Y_{0t-1,i})) \leq y - \hat{F}^{-1}_{Y_{0t-1}|D_t=1}(\hat{F}_{Y_{0t-2}|D_t=1}(Y_{0t-2,i}))\} \right] - \hat{\mu}_2 \Bigg) \tag{1.62}$$

$$+ \sqrt{n} \Bigg(\frac{1}{n_T} \sum_{i \in \mathcal{T}} \left[1\{F^{-1}_{\Delta Y_{0t}|D_t=0}(F_{\Delta Y_{0t-1}|D_t=1}(\Delta Y_{0t-1,i})) \leq y - \hat{F}^{-1}_{Y_{0t-1}|D_t=1}(\hat{F}_{Y_{0t-2}|D_t=1}(Y_{0t-2,i}))\} \right]$$
$$- \frac{1}{n_T} \sum_{i \in \mathcal{T}} \left[1\{F^{-1}_{\Delta Y_{0t}|D_t=0}(F_{\Delta Y_{0t-1}|D_t=1}(\Delta Y_{0t-1,i})) \leq y - F^{-1}_{Y_{0t-1}|D_t=1}(\hat{F}_{Y_{0t-2}|D_t=1}(Y_{0t-2,i}))\} \right] - \hat{\mu}_3 \Bigg) \tag{1.63}$$

$$+ \sqrt{n} \Bigg(\frac{1}{n_T} \sum_{i \in \mathcal{T}} \left[1\{F^{-1}_{\Delta Y_{0t}|D_t=0}(F_{\Delta Y_{0t-1}|D_t=1}(\Delta Y_{0t-1,i})) \leq y - F^{-1}_{Y_{0t-1}|D_t=1}(\hat{F}_{Y_{0t-2}|D_t=1}(Y_{0t-2,i}))\} \right]$$
$$- \frac{1}{n_T} \sum_{i \in \mathcal{T}} \left[1\{F^{-1}_{\Delta Y_{0t}|D_t=0}(F_{\Delta Y_{0t-1}|D_t=1}(\Delta Y_{0t-1,i})) \leq y - F^{-1}_{Y_{0t-1}|D_t=1}(F_{Y_{0t-2}|D_t=1}(Y_{0t-2,i}))\} \right] - \hat{\mu}_4 \Bigg) \tag{1.64}$$

$$+ \sqrt{n} \Bigg(\frac{1}{n_T} \sum_{i \in \mathcal{T}} \left[1\{F^{-1}_{\Delta Y_{0t}|D_t=0}(F_{\Delta Y_{0t-1}|D_t=1}(\Delta Y_{0t-1,i})) \leq y - F^{-1}_{Y_{0t-1}|D_t=1}(F_{Y_{0t-2}|D_t=1}(Y_{0t-2,i}))\} \right]$$
$$- \mathrm{E}\left[1\{F^{-1}_{\Delta Y_{0t}|D_t=0}(F_{\Delta Y_{0t-1}|D_t=1}(\Delta Y_{0t-1,i})) \leq y - F^{-1}_{Y_{0t-1}|D_t=1}(F_{Y_{0t-2}|D_t=1}(Y_{0t-2,i}))\}|D_t = 1 \right] - \hat{\mu}_5 \Bigg) \tag{1.65}$$

Next, we show that Equation 1.61, Equation 1.62, Equation 1.63, and Equation 1.64 each converge to 0. We analyze each equation in turn.

Equation 1.61:

Recall

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

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

Next, notice that

$$= \sqrt{n}\left| \frac{1}{n_T}\sum_{i\in\mathcal{T}}\left[1\{\hat{F}^{-1}_{\Delta Y_{0t}|D_t=0}(\hat{F}_{\Delta Y_{0t-1}|D_t=1}(\Delta Y_{0t-1,i})) \leq y - \hat{F}^{-1}_{Y_{0t-1}|D_t=1}(\hat{F}_{Y_{0t-2}|D_t=1}(Y_{0t-2,i}))\}\right] \right.$$
$$\left. - \frac{1}{n_T}\sum_{i\in\mathcal{T}}\left[1\{F^{-1}_{\Delta Y_{0t}|D_t=0}(\hat{F}_{\Delta Y_{0t-1}|D_t=1}(\Delta Y_{0t-1,i})) \leq y - \hat{F}^{-1}_{Y_{0t-1}|D_t=1}(\hat{F}_{Y_{0t-2}|D_t=1}(Y_{0t-2,i}))\}\right] - \hat{\mu}_1 \right| \tag{1.66}$$

$$= \sqrt{n}\left| \frac{1}{n_T}\sum_{i\in\mathcal{T}}\left[1\{\hat{F}^{-1}_{\Delta Y_{0t}|D_t=0}(\hat{F}_{\Delta Y_{0t-1}|D_t=1}(\Delta Y_{0t-1,i})) \leq g(Y_{0t-2,i})\}\right] \right.$$
$$\left. - \frac{1}{n_T}\sum_{i\in\mathcal{T}}\left[1\{F^{-1}_{\Delta Y_{0t}|D_t=0}(\hat{F}_{\Delta Y_{0t-1}|D_t=1}(\Delta Y_{0t-1,i})) \leq g(Y_{0t-2,i})\}\right] - \hat{\mu}_1 \right| \tag{1.67}$$

$$\leq \sqrt{n}\left| \frac{1}{n_T}\sum_{i\in\mathcal{T}}\left[1\{\hat{F}^{-1}_{\Delta Y_{0t}|D_t=0}(\hat{F}_{\Delta Y_{0t-1}|D_t=1}(\Delta Y_{0t-1,i})) \leq g(Y_{0t-2,i})\}\right] \right.$$
$$\left. - \frac{1}{n_T}\sum_{i\in\mathcal{T}}\left[1\{\Delta Y_{0t-1,i} \leq \hat{F}^{-1}_{\Delta Y_{0t-1}|D_t=1}(\hat{F}_{\Delta Y_{0t}|D_t=0}(g(Y_{0t-2,i})))\}\right] \right| \tag{1.68}$$

$$+ \sqrt{n}\left| \frac{1}{n_T}\sum_{i\in\mathcal{T}}\left[1\{\Delta Y_{0t-1,i} \leq \hat{F}^{-1}_{\Delta Y_{0t-1}|D_t=1}(F_{\Delta Y_{0t}|D_t=0}(g(Y_{0t-2,i})))\}\right] \right.$$
$$\left. - \frac{1}{n_T}\sum_{i\in\mathcal{T}}\left[1\{F^{-1}_{\Delta Y_{0t}|D_t=0}(\hat{F}_{\Delta Y_{0t-1}|D_t=1}(\Delta Y_{0t-1,i})) \leq g(Y_{0t-2,i})\}\right] \right| \tag{1.69}$$

$$+ \sqrt{n}\left| \frac{1}{n_T}\sum_{i\in\mathcal{T}}\left[1\{\Delta Y_{0t-1,i} \leq \hat{F}^{-1}_{\Delta Y_{0t-1}|D_t=1}(\hat{F}_{\Delta Y_{0t}|D_t=0}(g(Y_{0t-2,i})))\}\right] \right.$$
$$\left. - \frac{1}{n_T}\sum_{i\in\mathcal{T}}\left[1\{\Delta Y_{0t-1,i} \leq \hat{F}^{-1}_{\Delta Y_{0t-1}|D_t=1}(F_{\Delta Y_{0t}|D_t=0}(g(Y_{0t-2,i})))\}\right] - \hat{\mu}_1 \right| \tag{1.70}$$

$$\leq \sup_z \sqrt{n}\left| \frac{1}{n_T}\sum_{i\in\mathcal{T}}\left[1\{\hat{F}^{-1}_{\Delta Y_{0t}|D_t=0}(\hat{F}_{\Delta Y_{0t-1}|D_t=1}(\Delta Y_{0t-1,i})) \leq g(z)\}\right] \right.$$
$$\left. - \frac{1}{n_T}\sum_{i\in\mathcal{T}}\left[1\{\Delta Y_{0t-1,i} \leq \hat{F}^{-1}_{\Delta Y_{0t-1}|D_t=1}(\hat{F}_{\Delta Y_{0t}|D_t=0}(g(z)))\}\right] \right| \tag{1.71}$$

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

$$
\begin{aligned}
&\quad - \frac{1}{n_T}\sum_{i\in\mathcal{T}}\Big[1\{F^{-1}_{\Delta Y_{0t}|D_t=0}(\hat{F}_{\Delta Y_{0t-1}|D_t=1}(\Delta Y_{0t-1,i})) \le g(z)\}\Big]\Bigg| \quad (1.72)\\
&+\sqrt{n}\Bigg|\frac{1}{n_T}\sum_{i\in\mathcal{T}}\Big[1\{\Delta Y_{0t-1,i} \le \hat{F}^{-1}_{\Delta Y_{0t-1}|D_t=1}(\hat{F}_{\Delta Y_{0t}|D_t=0}(g(Y_{0t-2,i})))\}\Big]\\
&\quad - \frac{1}{n_T}\sum_{i\in\mathcal{T}}\Big[1\{\Delta Y_{0t-1,i} \le \hat{F}^{-1}_{\Delta Y_{0t-1}|D_t=1}(F_{\Delta Y_{0t}|D_t=0}(g(Y_{0t-2,i})))\}\Big] - \hat{\mu}_1\Bigg| \quad (1.73)\\
&\le \sqrt{n}\Bigg|\frac{1}{n_T}\sum_{i\in\mathcal{T}}\Big[1\{\Delta Y_{0t-1,i} \le \hat{F}^{-1}_{\Delta Y_{0t-1}|D_t=1}(\hat{F}_{\Delta Y_{0t}|D_t=0}(g(Y_{0t-2,i})))\}\Big]\\
&\quad - \frac{1}{n_T}\sum_{i\in\mathcal{T}}\Big[1\{\Delta Y_{0t-1,i} \le \hat{F}^{-1}_{\Delta Y_{0t-1}|D_t=1}(F_{\Delta Y_{0t}|D_t=0}(g(Y_{0t-2,i})))\}\Big]\\
&\quad - \frac{1}{n_T n_C}\sum_{i\in\mathcal{T}}\sum_{j\in\mathcal{C}} 1\{\Delta Y_{0t,j} \le g(Y_{0t-2,i})\} - F_{\Delta Y_{0t}|D_t=0}(g(Y_{0t-2,i}))\Bigg| \quad (1.74)\\
&+\sqrt{n}\Bigg|\frac{1}{n_T n_C}\sum_{i\in\mathcal{T}}\sum_{j\in\mathcal{C}}\Big[1\{\Delta Y_{0t,j} \le y - \hat{F}^{-1}_{Y_{0t-1}|D_t=1}(\hat{F}_{Y_{0t-2}|D_t=1}(Y_{0t-1,i}))\}\\
&\qquad -F_{\Delta Y_{0t}|D_t=0}(y - \hat{F}^{-1}_{Y_{0t-1}|D_t=1}(\hat{F}_{Y_{0t-2}|D_t=1}(Y_{0t-1,i})))\Big]\\
&\quad - \frac{1}{n_T n_C}\sum_{i\in\mathcal{T}}\sum_{j\in\mathcal{C}}\Big[1\{\Delta Y_{0t,j} \le y - F^{-1}_{Y_{0t-1}|D_t=1}(F_{Y_{0t-2}|D_t=1}(Y_{0t-1,i}))\}\\
&\qquad -F_{\Delta Y_{0t}|D_t=0}(y - F^{-1}_{Y_{0t-1}|D_t=1}(F_{Y_{0t-2}|D_t=1}(Y_{0t-1,i})))\Big]\Bigg| + o_p(1) \quad (1.75)\\
&\le \sup_z \sqrt{n}\Big|\hat{F}_{\Delta Y_{0t-1}|D_t=1}(\hat{F}^{-1}_{\Delta Y_{0t-1}|D_t=1}(\hat{F}_{\Delta Y_{0t}|D_t=0}(g(z))))\\
&\quad - \hat{F}_{\Delta Y_{0t-1}|D_t=1}(\hat{F}^{-1}_{\Delta Y_{0t-1}|D_t=1}(F_{\Delta Y_{0t}|D_t=0}(g(z))))\\
&\quad - \big\{\hat{F}_{\Delta Y_{0t}|D_t=0}(g(z)) - F_{\Delta Y_{0t}|D_t=0}(g(z))\big\}\Big| \quad (1.76)\\
&+\sup_z \sqrt{n}\Big|\hat{F}_{\Delta Y_{0t}|D_t=0}(y - \hat{F}^{-1}_{Y_{0t-1}|D_t=1}(\hat{F}_{Y_{0t-2}|D_t=1}(z)))\\
&\qquad - F_{\Delta Y_{0t}|D_t=0}(y - \hat{F}^{-1}_{Y_{0t-1}|D_t=1}(\hat{F}_{Y_{0t-2}|D_t=1}(z)))\\
&\quad - \hat{F}_{\Delta Y_{0t}|D_t=0}(y - F^{-1}_{Y_{0t-1}|D_t=1}(F_{Y_{0t-2}|D_t=1}(z)))\\
&\qquad -F_{\Delta Y_{0t}|D_t=0}(y - F^{-1}_{Y_{0t-1}|D_t=1}(F_{Y_{0t-2}|D_t=1}(z)))\Big| + o_p(1) \quad (1.77)\\
&\xrightarrow{p} 0
\end{aligned}
$$

Equation 1.71 and Equation 1.72 converge to 0 by Lemma 7 and Lemma 8, respectively. Equation 1.76 converges to 0 by an application of Lemma 13 followed by some simple cancellations. Equation 1.77 converges to 0 by Lemma 15.

Equation 1.62: First, recall that

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

Then,

$$\begin{aligned}&\sqrt{n}\Bigg(\frac{1}{n_T}\sum_{i\in\mathcal{T}}\Big[1\{F^{-1}_{\Delta Y_{0t}|D_t=0}(\hat{F}_{\Delta Y_{0t-1}|D_t=1}(\Delta Y_{0t-1,i})) \leq y - \hat{F}^{-1}_{Y_{0t-1}|D_t=1}(\hat{F}_{Y_{0t-2}|D_t=1}(Y_{0t-2,i}))\}\Big] \\ &\qquad - \frac{1}{n_T}\sum_{i\in\mathcal{T}}\Big[1\{F^{-1}_{\Delta Y_{0t}|D_t=0}(F_{\Delta Y_{0t-1}|D_t=1}(\Delta Y_{0t-1,i})) \leq y - \hat{F}^{-1}_{Y_{0t-1}|D_t=1}(\hat{F}_{Y_{0t-2}|D_t=1}(Y_{0t-2,i}))\}\Big] - \hat{\mu}_2\Bigg) \\ &\leq \sqrt{n}\Bigg|\frac{1}{n_T}\sum_{i\in\mathcal{T}}\Big[1\{F^{-1}_{\Delta Y_{0t}|D_t=0}(\hat{F}_{\Delta Y_{0t-1}|D_t=1}(\Delta Y_{0t-1,i})) \leq y - \hat{F}^{-1}_{Y_{0t-1}|D_t=1}(\hat{F}_{Y_{0t-2}|D_t=1}(Y_{0t-2,i}))\}\Big] \\ &\qquad - \frac{1}{n_T}\sum_{i\in\mathcal{T}}\Big[1\{\Delta Y_{0t-1,i} \leq \hat{F}^{-1}_{\Delta Y_{0t-1}|D_t=1}(F_{\Delta Y_{0t}|D_t=0}(y - \hat{F}^{-1}_{Y_{0t-1}|D_t=1}(\hat{F}_{Y_{0t-2}|D_t=1}(Y_{0t-2,i}))))\}\Big]\Bigg| \\ &\quad - \Bigg|\frac{1}{n_T}\sum_{i\in\mathcal{T}}\Big[1\{\Delta Y_{0t-1,i} \leq \hat{F}^{-1}_{\Delta Y_{0t-1}|D_t=1}(F_{\Delta Y_{0t}|D_t=0}(y - \hat{F}^{-1}_{Y_{0t-1}|D_t=1}(\hat{F}_{Y_{0t-2}|D_t=1}(Y_{0t-2,i}))))\}\Big] \\ &\qquad - \frac{1}{n_T}\sum_{i\in\mathcal{T}}\Big[1\{\Delta Y_{0t-1,i} \leq F^{-1}_{\Delta Y_{0t-1}|D_t=1}(F_{\Delta Y_{0t}|D_t=0}(y - \hat{F}^{-1}_{Y_{0t-1}|D_t=1}(\hat{F}_{Y_{0t-2}|D_t=1}(Y_{0t-2,i}))))\}\Big] - \hat{\mu}_2\Bigg|\end{aligned} \tag{1.78}$$

$$\begin{aligned}&\leq \sup_z \sqrt{n}\Bigg|\frac{1}{n_T}\sum_{i\in\mathcal{T}}\Big[\Delta Y_{0t-1,i} \leq \hat{F}^{-1}_{\Delta Y_{0t-1}|D_t=1}(F_{\Delta Y_{0t}|D_t=0}(g(z)))\Big] \\ &\qquad - \frac{1}{n_T}\sum_{i\in\mathcal{T}}\Big[\Delta Y_{0t-1,i} \leq F^{-1}_{\Delta Y_{0t-1|D_t=1}}(F_{\Delta Y_{0t}|D_t=0}(g(z)))\Big] \\ &\qquad - \frac{1}{n_T^2}\sum_{i\in\mathcal{T}}\sum_{j\in\mathcal{T}} 1\{\Delta Y_{0t-1,j} \leq F^{-1}_{\Delta Y_{0t-1|D_t=1}}(F_{\Delta Y_{0t}|D_t=0}(y - F^{-1}_{Y_{0t-1}|D_t=1}(F_{Y_{0t-2}|D_t=1}(z)))\}\end{aligned}$$

$$-F_{\Delta Y_{0t-1}|D_t=1}(F^{-1}_{\Delta Y_{0t-1|D_t=1}}(F_{\Delta Y_{0t}|D_t=0}(y - F^{-1}_{Y_{0t-1}|D_t=1}(F_{Y_{0t-2}|D_t=1}(z))))] \Big| + o_p(1) \tag{1.79}$$

$$\begin{aligned}
\leq \sup_z \sqrt{n} \Big| & \hat{F}_{\Delta Y_{0t-1}|D_t=1}(\hat{F}^{-1}_{\Delta Y_{0t-1}|D_t=1}(F_{\Delta Y_{0t}|D_t=0}(g(z)))) \\
& - \hat{F}_{\Delta Y_{0t-1}|D_t=1}(F^{-1}_{\Delta Y_{0t-1|D_t=1}}(F_{\Delta Y_{0t}|D_t=0}(g(z)))) \\
& - \Big(F_{\Delta Y_{0t-1}|D_t=1}(\hat{F}^{-1}_{\Delta Y_{0t-1}|D_t=1}(F_{\Delta Y_{0t}|D_t=0}(g(z)))) \\
& - F_{\Delta Y_{0t-1}|D_t=1}(F^{-1}_{\Delta Y_{0t-1|D_t=1}}(F_{\Delta Y_{0t}|D_t=0}(g(z)))) \Big) \Big| \qquad (1.80) \\
& + \sup_z \sqrt{n} \Big| F_{\Delta Y_{0t-1}|D_t=1}(\hat{F}^{-1}_{\Delta Y_{0t-1}|D_t=1}(F_{\Delta Y_{0t}|D_t=0}(g(z)))) \\
& - F_{\Delta Y_{0t-1}|D_t=1}(F^{-1}_{\Delta Y_{0t-1|D_t=1}}(F_{\Delta Y_{0t}|D_t=0}(g(z)))) \Big) \\
& - \Big(F_{\Delta Y_{0t-1}|D_t=1}(\hat{F}^{-1}_{\Delta Y_{0t-1}|D_t=1}(F_{\Delta Y_{0t}|D_t=0}(y - F^{-1}_{Y_{0t-1}|D_t=1}(F_{Y_{0t-2}|D_t=1}(z)))))) \\
& - F_{\Delta Y_{0t-1}|D_t=1}(F^{-1}_{\Delta Y_{0t-1|D_t=1}}(F_{\Delta Y_{0t}|D_t=0}(y - F^{-1}_{Y_{0t-1}|D_t=1}(F_{Y_{0t-2}|D_t=1}(z)))))) \Big) \Big| \qquad (1.81) \\
& + \sup_z \sqrt{n} \Big| F_{\Delta Y_{0t-1}|D_t=1}(\hat{F}^{-1}_{\Delta Y_{0t-1}|D_t=1}(F_{\Delta Y_{0t}|D_t=0}(y - F^{-1}_{Y_{0t-1}|D_t=1}(F_{Y_{0t-2}|D_t=1}(z)))))) \\
& - F_{\Delta Y_{0t-1}|D_t=1}(F^{-1}_{\Delta Y_{0t-1|D_t=1}}(F_{\Delta Y_{0t}|D_t=0}(y - F^{-1}_{Y_{0t-1}|D_t=1}(F_{Y_{0t-2}|D_t=1}(z)))))) \Big) \\
& - F_{\Delta Y_{0t-1}|D_t=1}(\hat{F}^{-1}_{\Delta Y_{0t-1}|D_t=1}(F_{\Delta Y_{0t}|D_t=0}(y - F^{-1}_{Y_{0t-1}|D_t=1}(F_{Y_{0t-2}|D_t=1}(z))) \\
& - F_{\Delta Y_{0t-1}|D_t=1}(F^{-1}_{\Delta Y_{0t-1|D_t=1}}(F_{\Delta Y_{0t}|D_t=0}(y - F^{-1}_{Y_{0t-1}|D_t=1}(F_{Y_{0t-2}|D_t=1}(z))))] \Big| + o_p(1) \qquad (1.82) \\
\xrightarrow{p} 0 &
\end{aligned}$$

where Equation 1.80 converges to 0 by Lemma 15, Equation 1.81 converges to 0 by several Taylor expansions (the result is similar to the proof of Lemma 16), and Equation 1.82 converges to 0 by first noticing the following Taylor expansion

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

$$= f_{\Delta Y_{0t-1}|D_t=1}(F^{-1}_{\Delta Y_{0t-1|D_t=1}}(F_{\Delta Y_{0t}|D_t=0}(y - F^{-1}_{Y_{0t-1}|D_t=1}(F_{Y_{0t-2}|D_t=1}(z)))))$$
$$\times \sqrt{n}\Big(\hat{F}^{-1}_{\Delta Y_{0t-1}|D_t=1}(F_{\Delta Y_{0t}|D_t=0}(y - F^{-1}_{Y_{0t-1}|D_t=1}(F_{Y_{0t-2}|D_t=1}(z))))$$
$$-F^{-1}_{\Delta Y_{0t-1}|D_t=1}(F_{\Delta Y_{0t}|D_t=0}(y - F^{-1}_{Y_{0t-1}|D_t=1}(F_{Y_{0t-2}|D_t=1}(z))))\Big) + o_p(1) \tag{1.83}$$

and then noting that

$$\sqrt{n}\Big(\hat{F}^{-1}_{\Delta Y_{0t-1}|D_t=1}(F_{\Delta Y_{0t}|D_t=0}(y - F^{-1}_{Y_{0t-1}|D_t=1}(F_{Y_{0t-2}|D_t=1}(z))))$$
$$-F^{-1}_{\Delta Y_{0t-1}|D_t=1}(F_{\Delta Y_{0t}|D_t=0}(y - F^{-1}_{Y_{0t-1}|D_t=1}(F_{Y_{0t-2}|D_t=1}(z))))\Big)$$
$$= \frac{1}{f_{\Delta Y_{0t-1}|D_t=1}(F^{-1}_{\Delta Y_{0t-1|D_t=1}}(F_{\Delta Y_{0t}|D_t=0}(y - F^{-1}_{Y_{0t-1}|D_t=1}(F_{Y_{0t-2}|D_t=1}(z)))))}$$
$$\times \sqrt{n}\Big(\hat{F}_{\Delta Y_{0t-1}|D_t=1}(F^{-1}_{\Delta Y_{0t-1}|D_t=1}(F_{\Delta Y_{0t}|D_t=0}(y - F^{-1}_{Y_{0t-1}|D_t=1}(F_{Y_{0t-2}|D_t=1}(z)))))$$
$$-F_{\Delta Y_{0t}|D_t=0}(y - F^{-1}_{Y_{0t-1}|D_t=1}(F_{Y_{0t-2}|D_t=1}(z)))\Big) \tag{1.84}$$

which holds by Lemma 10. Combining Equation 1.83 and Equation 1.84 completes the result.

Equation 1.63: Recall that

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

Then,

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

$$
\begin{aligned}
&\leq \sqrt{n}\left|\frac{1}{n_T}\sum_{i\in\mathcal{T}}\Big[1\{\Delta Y_{0t-1,i}\leq F^{-1}_{\Delta Y_{0t-1}|D_t=1}(F_{\Delta Y_{0t}|D_t=0}(y-\hat{F}^{-1}_{Y_{0t-1}|D_t=1}(\hat{F}_{Y_{0t-2}|D_t=1}(Y_{0t-2,i})))))\}\Big]\right.\\
&\qquad\left.-\frac{1}{n_T}\sum_{i\in\mathcal{T}}\Big[1\{\Delta Y_{0t-1,i}\leq F^{-1}_{\Delta Y_{0t-1}|D_t=1}(F_{\Delta Y_{0t}|D_t=0}(y-F^{-1}_{Y_{0t-1}|D_t=1}(\hat{F}_{Y_{0t-2}|D_t=1}(Y_{0t-2,i})))))\}\Big]-\hat{\mu}_3\right|\\
&\leq \sup_z\sqrt{n}\Big|\hat{F}_{\Delta Y_{0t-1}|D_t=1}(F^{-1}_{\Delta Y_{0t-1}|D_t=1}(F_{\Delta Y_{0t}|D_t=0}(y-\hat{F}^{-1}_{Y_{0t-1}|D_t=1}(\hat{F}_{Y_{0t-2}|D_t=1}(z)))))\\
&\qquad\quad-\hat{F}_{\Delta Y_{0t-1}|D_t=1}(F^{-1}_{\Delta Y_{0t-1}|D_t=1}(F_{\Delta Y_{0t}|D_t=0}(y-F^{-1}_{Y_{0t-1}|D_t=1}(\hat{F}_{Y_{0t-2}|D_t=1}(z)))))\\
&\qquad\quad-\frac{f_{\Delta Y_{0t}|D_t=0}(y-F^{-1}_{Y_{0t-1}|D_t=1}(F_{Y_{0t-2}|D_t=1}(z)))}{f_{Y_{0t-1}|D_t=1}(F_{Y_{0t-2}|D_t=1}(z))}\\
&\qquad\qquad\times\Big(\hat{F}_{Y_{0t-1|D_t=1}}(F^{-1}_{Y_{0t-1}|D_t=1}(F_{Y_{0t-2}|D_t=1}(z)))\\
&\qquad\qquad\quad-F_{Y_{0t-1}|D_t=1}(F^{-1}_{Y_{0t-1}|D_t=1}(F_{Y_{0t-2}|D_t=1}(z)))\Big)\Big| \qquad (1.85)\\
&\leq \sup_z\sqrt{n}\Big|\hat{F}_{\Delta Y_{0t-1}|D_t=1}(F^{-1}_{\Delta Y_{0t-1}|D_t=1}(F_{\Delta Y_{0t}|D_t=0}(y-\hat{F}^{-1}_{Y_{0t-1}|D_t=1}(\hat{F}_{Y_{0t-2}|D_t=1}(z)))))\\
&\qquad\quad-\hat{F}_{\Delta Y_{0t-1}|D_t=1}(F^{-1}_{\Delta Y_{0t-1}|D_t=1}(F_{\Delta Y_{0t}|D_t=0}(y-F^{-1}_{Y_{0t-1}|D_t=1}(\hat{F}_{Y_{0t-2}|D_t=1}(z)))))\\
&\qquad-\Big(F_{\Delta Y_{0t-1}|D_t=1}(F^{-1}_{\Delta Y_{0t-1}|D_t=1}(F_{\Delta Y_{0t}|D_t=0}(y-\hat{F}^{-1}_{Y_{0t-1}|D_t=1}(\hat{F}_{Y_{0t-2}|D_t=1}(z)))))\\
&\qquad\quad-F_{\Delta Y_{0t-1}|D_t=1}(F^{-1}_{\Delta Y_{0t-1}|D_t=1}(F_{\Delta Y_{0t}|D_t=0}(y-F^{-1}_{Y_{0t-1}|D_t=1}(\hat{F}_{Y_{0t-2}|D_t=1}(z)))))\Big)\Big|\\
&\qquad (1.86)\\
&+\sup_z\sqrt{n}\Big|F_{\Delta Y_{0t}|D_t=0}(y-\hat{F}^{-1}_{Y_{0t-1}|D_t=1}(\hat{F}_{Y_{0t-2}|D_t=1}(z)))\\
&\qquad\quad-F_{\Delta Y_{0t}|D_t=0}(y-F^{-1}_{Y_{0t-1}|D_t=1}(\hat{F}_{Y_{0t-2}|D_t=1}(z)))\\
&\qquad-\Big(F_{\Delta Y_{0t}|D_t=0}(y-\hat{F}^{-1}_{Y_{0t-1}|D_t=1}(F_{Y_{0t-2}|D_t=1}(z)))\\
&\qquad\quad-F_{\Delta Y_{0t}|D_t=0}(y-F^{-1}_{Y_{0t-1}|D_t=1}(F_{Y_{0t-2}|D_t=1}(z)))\Big)\Big| \qquad (1.87)\\
&+\sup_z\sqrt{n}\Big|F_{\Delta Y_{0t}|D_t=0}(y-\hat{F}^{-1}_{Y_{0t-1}|D_t=1}(F_{Y_{0t-2}|D_t=1}(z)))\\
&\qquad\quad-F_{\Delta Y_{0t}|D_t=0}(y-F^{-1}_{Y_{0t-1}|D_t=1}(F_{Y_{0t-2}|D_t=1}(z)))\\
&\qquad-\frac{f_{\Delta Y_{0t}|D_t=0}(y-F^{-1}_{Y_{0t-1}|D_t=1}(F_{Y_{0t-2}|D_t=1}(z)))}{f_{Y_{0t-1}|D_t=1}(F_{Y_{0t-2}|D_t=1}(z))}\\
&\qquad\qquad\times\Big(\hat{F}_{Y_{0t-1|D_t=1}}(F^{-1}_{Y_{0t-1}|D_t=1}(F_{Y_{0t-2}|D_t=1}(z)))\\
&\qquad\qquad\quad-F_{Y_{0t-1}|D_t=1}(F^{-1}_{Y_{0t-1}|D_t=1}(F_{Y_{0t-2}|D_t=1}(z)))\Big)\Big| \qquad (1.88)
\end{aligned}
$$

$$\xrightarrow{p} 0$$

Equation 1.88 converges to 0 by Lemma 15. Equation 1.80 converges to 0 by Lemma 16. For Equation 1.87, by a Taylor Expansion,

$$\begin{aligned}\sqrt{n}\left(F_{\Delta Y_{0t}|D_t=0}(y-\hat{F}^{-1}_{Y_{0t-1}|D_t=1}(F_{Y_{0t-2}|D_t=1}(z))) - F_{\Delta Y_{0t}|D_t=0}(y-F^{-1}_{Y_{0t-1}|D_t=1}(F_{Y_{0t-2}|D_t=1}(z)))\right) \\ = f_{\Delta Y_{0t}|D_t=0}(y-F^{-1}_{Y_{0t-1}|D_t=1}(F_{Y_{0t-2}|D_t=1}(z)))\sqrt{n}\left(\hat{F}^{-1}_{Y_{0t-1}|D_t=1}(F_{Y_{0t-2}|D_t=1}(z))\right. \\ \left. -F^{-1}_{Y_{0t-1}|D_t=1}(F_{Y_{0t-2}|D_t=1}(z))\right) + o_p(1)\end{aligned} \tag{1.89}$$

The result is then obtained by using Lemma 10 on the term

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

in Equation 1.89.

Equation 1.64 Recall that:

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

Then,

$$\begin{aligned}\sqrt{n}\left(\frac{1}{n_T}\sum_{i\in\mathcal{T}}\left[1\{F^{-1}_{\Delta Y_{0t}|D_t=0}(F_{\Delta Y_{0t-1}|D_t=1}(\Delta Y_{0t-1,i}))\leq y-F^{-1}_{Y_{0t-1}|D_t=1}(\hat{F}_{Y_{0t-2}|D_t=1}(Y_{0t-2,i}))\}\right]\right. \\ \left. -\frac{1}{n_T}\sum_{i\in\mathcal{T}}\left[1\{F^{-1}_{\Delta Y_{0t}|D_t=0}(F_{\Delta Y_{0t-1}|D_t=1}(\Delta Y_{0t-1,i}))\leq y-F^{-1}_{Y_{0t-1}|D_t=1}(F_{Y_{0t-2}|D_t=1}(Y_{0t-2,i}))\}\right]-\hat{\mu}_4\right)\end{aligned} \tag{1.90}$$

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

$$
\begin{aligned}
& - \frac{f_{\Delta Y_{0t}|D_t=0}(y - F^{-1}_{Y_{0t-1}|D_t=1}(F_{Y_{0t-2}|D_t=1}(z)))}{f_{Y_{0t-1}|D_t=1}(F^{-1}_{Y_{0t-1}|D_t=1}(F_{Y_{0t-2}|D_t=1}(z)))} \\
& \quad \times \left(\hat{F}_{Y_{0t-2}|D_t=1}(z) - F_{Y_{0t-2}|D_t=1}(z)]\right)\Big| \qquad (1.91) \\
\leq & \sup_z \sqrt{n} \Big| \hat{F}_{\Delta Y_{0t-1}|D_t=1}(F^{-1}_{\Delta Y_{0t-1}|D_t=1}(F_{\Delta Y_{0t}|D_t=1}(y - F^{-1}_{Y_{0t-1}|D_t=1}(\hat{F}_{Y_{0t-2}|D_t=1}(z))))) \\
& - \hat{F}_{\Delta Y_{0t-1}|D_t=1}(F^{-1}_{\Delta Y_{0t-1}|D_t=1}(F_{\Delta Y_{0t}|D_t=1}(y - F^{-1}_{Y_{0t-1}|D_t=1}(F_{Y_{0t-2}|D_t=1}(z))))) \\
& - \Big(F_{\Delta Y_{0t-1}|D_t=1}(F^{-1}_{\Delta Y_{0t-1}|D_t=1}(F_{\Delta Y_{0t}|D_t=1}(y - F^{-1}_{Y_{0t-1}|D_t=1}(\hat{F}_{Y_{0t-2}|D_t=1}(z))))) \\
& \quad - F_{\Delta Y_{0t-1}|D_t=1}(F^{-1}_{\Delta Y_{0t-1}|D_t=1}(F_{\Delta Y_{0t}|D_t=1}(y - F^{-1}_{Y_{0t-1}|D_t=1}(F_{Y_{0t-2}|D_t=1}(z))))) \Big) \Big| \\
& \qquad (1.92) \\
& + \sup_z \sqrt{n} \Big| F_{\Delta Y_{0t}|D_t=1}(y - F^{-1}_{Y_{0t-1}|D_t=1}(\hat{F}_{Y_{0t-2}|D_t=1}(z))) \\
& \quad - F_{\Delta Y_{0t}|D_t=1}(y - F^{-1}_{Y_{0t-1}|D_t=1}(F_{Y_{0t-2}|D_t=1}(z))) \\
& - \frac{f_{\Delta Y_{0t}|D_t=0}(y - F^{-1}_{Y_{0t-1}|D_t=1}(F_{Y_{0t-2}|D_t=1}(z)))}{f_{Y_{0t-1}|D_t=1}(F^{-1}_{Y_{0t-1}|D_t=1}(F_{Y_{0t-2}|D_t=1}(z)))} \\
& \quad \times \left(\hat{F}_{Y_{0t-2}|D_t=1}(z) - F_{Y_{0t-2}|D_t=1}(z)]\right)\Big| \qquad (1.93) \\
\xrightarrow{p} & \; 0
\end{aligned}
$$

Equation 1.92 converges to 0 by Lemma 15. For Equation 1.93, notice that by a Taylor expansion,

$$
\begin{aligned}
& \sqrt{n} \left(F_{\Delta Y_{0t}|D_t=1}(y - F^{-1}_{Y_{0t-1}|D_t=1}(\hat{F}_{Y_{0t-2}|D_t=1}(z))) - F_{\Delta Y_{0t}|D_t=1}(y - F^{-1}_{Y_{0t-1}|D_t=1}(F_{Y_{0t-2}|D_t=1}(z))) \right) \\
& \quad = f_{\Delta Y_{0t}|D_t=1}(y - F^{-1}_{Y_{0t-1}|D_t=1}(\hat{F}_{Y_{0t-2}|D_t=1}(z))) \\
& \qquad \times \sqrt{n} \left(F^{-1}_{Y_{0t-1}|D_t=1}(\hat{F}_{Y_{0t-2}|D_t=1}(z)) - F^{-1}_{Y_{0t-1}|D_t=1}(F_{Y_{0t-2}|D_t=1}(z)) \right) + o_p(1) \qquad (1.94)
\end{aligned}
$$

Then, by a second Taylor expansion,

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

$$= \frac{1}{f_{Y_{0t-1}|D_t=1}(F^{-1}_{Y_{0t-1}|D_t=1}(F_{Y_{0t-2}|D_t=1}(z)))} \sqrt{n}\left(\hat{F}_{Y_{0t-2}|D_t=1}(z) - F_{Y_{0t-2}|D_t=1}(z)\right) + o_p(1) \tag{1.95}$$

and combining Equation 1.94 and Equation 1.95 implies the result.

Equation 1.72 Since

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

Equation 1.72 is equal to 0.

□

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

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

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

$$\hat{\mu}_1 = \frac{1}{n_C} \sum_{i \in \mathcal{C}} \mu_{12}(\Delta Y_{t,i}) + o_p(1)$$
$$\hat{\mu}_2 = \frac{1}{n_T} \sum_{i \in \mathcal{T}} \mu_{22}(Y_{t-2,i}) + o_p(1)$$
$$\hat{\mu}_3 = \frac{1}{n_T} \sum_{i \in \mathcal{T}} \mu_{31}(Y_{t-1,i}) + o_p(1)$$

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

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

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

$$\hat{\mu}_1 = \frac{1}{n_T n_C}\sum_{i\in\mathcal{T}}\sum_{j\in\mathcal{C}} \mu_1(Y_{0t-2,i}, \Delta Y_{0t,j}) \tag{1.96}$$

$$= \frac{1}{n_T}\sum_{i\in\mathcal{T}} \mu_{11}(Y_{0t-2,i}) + \frac{1}{n_C}\sum_{i\in\mathcal{C}} \mu_{12}(\Delta Y_{0t,i}) - \mathrm{E}[\mu_1(Y_{0t-2}, \Delta Y_{0t})] + o_p(1) \tag{1.97}$$

$$= \frac{1}{n_C}\sum_{i\in\mathcal{C}} \mu_{12}(\Delta Y_{0t,i}) + o_p(1) \tag{1.98}$$

Equation 1.97 uses Lemma 17. It is easy to show that $\mu_{11}(x) = 0$ and $\mathrm{E}[\mu_1(Y_{0t-2}, \Delta Y_{0t})] = 0$. □

Proof of Theorem 5. Let $\mu_j(\cdot; y)$ be the $\mu_j(\cdot)$ used in the previous lemmas with the dependence on the value of y in $\mathrm{P}(Y_{0t} < y | D_t = 1)$ explicit, and likewise for $\mu_{jk}(\cdot; y)$.

As a first step, notice that

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

$$= \frac{1}{f_{Y_{1t}|D_t=1}(F^{-1}_{Y_{1t}|D_t=1}(\tau))}\sqrt{n}\left(\frac{1}{n_T}\sum_{i\in\mathcal{T}} 1\{Y_{1t} \leq F^{-1}_{Y_{1t}|D_t=1}(\tau)\} - \tau\right) \tag{1.99}$$

$$\equiv \frac{1}{f_{Y_{1t}|D_t=1}(F^{-1}_{Y_{1t}|D_t=1}(\tau))}\sqrt{n}\left(\frac{1}{n_T}\sum_{i\in\mathcal{T}} \psi(Y_{1t,i}; F^{-1}_{Y_{1t}|D_t=1}(\tau))\right) \tag{1.100}$$

Second, based on Lemma 19 and Lemma 20[18]

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

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

$$
\begin{aligned}
&= \frac{1}{f_{Y_{0t}|D_t=1}(F^{-1}_{Y_{0t}|D_t=1}(\tau))}\sqrt{n}\left(\hat{F}_{Y_{0t}|D_t=1}(F^{-1}_{Y_{0t}|D_t=1}(\tau)) - F_{Y_{0t}|D_t=1}(F^{-1}_{Y_{0t}|D_t=1}(\tau))\right) + o_p(1) \\
&= \frac{1}{f_{Y_{0t}|D_t=1}(F^{-1}_{Y_{0t}|D_t=1}(\tau))}\sqrt{n}\left(\frac{1}{n_C}\sum_{i\in\mathcal{C}}\mu_{12}(\Delta Y_{t,i}; F^{-1}_{Y_{0t}|D_t=1}(\tau))\right. \\
&\qquad + \frac{1}{n_T}\sum_{i\in\mathcal{T}}\mu_{22}(\Delta Y_{t-1,i}; F^{-1}_{Y_{0t}|D_t=1}(\tau)) + \frac{1}{n_T}\sum_{i\in\mathcal{T}}\mu_{32}(Y_{t-1,i}; F^{-1}_{Y_{0t}|D_t=1}(\tau)) \\
&\qquad \left. + \frac{1}{n_T}\sum_{i\in\mathcal{T}}\mu_{42}(Y_{t-2,i}; F^{-1}_{Y_{0t}|D_t=1}(\tau)) + \frac{1}{n_T}\sum_{i\in\mathcal{T}}\mu_5(\Delta Y_{t-1,i}, Y_{0t-2,i}; F^{-1}_{Y_{0t}|D_t=1}(\tau))\right) + o_p(1)
\end{aligned}
\tag{1.101}
$$

where, as defined in the text,

$$
\begin{aligned}
\mu_{12}(z;y) &= \mathrm{E}\left[1\{z \leq (y - \mathrm{F}^{-1}_{Y_{0t-1}|D_t=1}(\mathrm{F}_{Y_{0t-2}|D_t=1}(Y_{0t-2})))\}\right. \\
&\qquad \left. - \mathrm{F}_{\Delta Y_{0t}|D_t=0}(y - \mathrm{F}^{-1}_{Y_{0t-1}|D_t=1}(\mathrm{F}_{Y_{0t-2}|D_t=1}(Y_{0t-2})))|D_t = 1\right] \\
\mu_{22}(z;y) &= \mathrm{E}\left[1\{z \leq \mathrm{F}^{-1}_{\Delta Y_{0t-1}|D_t=1}(\mathrm{F}_{\Delta Y_{0t}|D_t=0}(y - \mathrm{F}^{-1}_{Y_{0t-1}|D_t=1}(\mathrm{F}_{Y_{0t-2}|D_t=1}(Y_{0t-2}))))\}\right. \\
&\qquad \left. - \mathrm{F}_{\Delta Y_{0t}|D_t=0}(y - \mathrm{F}^{-1}_{Y_{0t-1}|D_t=1}(\mathrm{F}_{Y_{0t-2}|D_t=1}(Y_{0t-2})))|D_t = 1\right] \\
\mu_{32}(z;y) &= \mathrm{E}\left[\frac{f_{\Delta Y_{0t}|D_t=0}(y - \mathrm{F}^{-1}_{Y_{0t-1}|D_t=1}(\mathrm{F}_{Y_{0t-2}|D_t=1}(Y_{0t-2})))}{f_{Y_{0t-1}|D_t=1}(\mathrm{F}_{Y_{0t-2}|D_t=1}(Y_{0t-2}))}\right. \\
&\qquad \times \left(1\{z \leq \mathrm{F}^{-1}_{Y_{0t-1}|D_t=1}(\mathrm{F}_{Y_{0t-2}|D_t=1}(Y_{0t-2}))\}\right. \\
&\qquad \left.\left. - \mathrm{F}_{Y_{0t-1}|D_t=1}(\mathrm{F}^{-1}_{Y_{0t-1}|D_t=1}(\mathrm{F}_{Y_{0t-2}|D_t=1}(Y_{0t-2})))\right)|D_t = 1\right] \\
\mu_{42}(z;y) &= \mathrm{E}\left[\frac{f_{\Delta Y_{0t}|D_t=0}(y - \mathrm{F}^{-1}_{Y_{0t-1}|D_t=1}(\mathrm{F}_{Y_{0t-2}|D_t=1}(Y_{0t-2})))}{f_{Y_{0t-1}|D_t=1}(\mathrm{F}_{Y_{0t-2}|D_t=1}(Y_{0t-2}))}\right. \\
&\qquad \left. \times \left(1\{z \leq Y_{0t-2}\} - \mathrm{F}_{Y_{0t-2}|D_t=1}(Y_{0t-2})\right)|D_t = 1\right] \\
\mu_5(z_1, z_2; y) &= 1\{\mathrm{F}^{-1}_{\Delta Y_{0t}|D_t=0}(\mathrm{F}_{\Delta Y_{0t-1}|D_t=1}(z_1)) \leq y - \mathrm{F}^{-1}_{Y_{0t-1}|D_t=1}(\mathrm{F}_{Y_{0t-2}|D_t=1}(z_2))\} \\
&\qquad - \mathrm{E}\left[1\{\mathrm{F}^{-1}_{\Delta Y_{0t}|D_t=0}(\mathrm{F}_{\Delta Y_{0t-1}|D_t=1}(z_1)) \leq y - \mathrm{F}^{-1}_{Y_{0t-1}|D_t=1}(\mathrm{F}_{Y_{0t-2}|D_t=1}(z_2))\}|D_t = 1\right]
\end{aligned}
$$

and

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

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

Since

$$\begin{aligned}
&\sqrt{n}\left(\hat{\mathrm{QTT}}(\tau) - \mathrm{QTT}(\tau)\right) \\
&\quad = \sqrt{n}\left(\hat{F}^{-1}_{Y_{1t}|D_t=1}(\tau) - F^{-1}_{Y_{1t}|D_t=1}(\tau)\right) - \sqrt{n}\left(\hat{F}^{-1}_{Y_{0t}|D_t=1}(\tau) - F^{-1}_{Y_{0t}|D_t=1}(\tau)\right) \\
&\quad = \sqrt{n}\Bigg\{ \frac{1}{f_{Y_{1t}|D_t=1}(F^{-1}_{Y_{1t}|D_t=1}(\tau))} \frac{1}{p} \frac{1}{n} \sum_{i=1}^{n} D_t \psi(Y_{1t,i}; F^{-1}_{Y_{1t}|D_t=1}(\tau)) \\
&\qquad + \frac{1}{f_{Y_{0t}|D_t=1}(F^{-1}_{Y_{0t}|D_t=1}(\tau))} \frac{1}{1-p} \frac{1}{n} \sum_{i=1}^{n} (1-D_t)\mu_{12}(\Delta Y_{t,i}; F^{-1}_{Y_{0t}|D_t=1}(\tau)) \\
&\qquad + \frac{1}{f_{Y_{0t}|D_t=1}(F^{-1}_{Y_{0t}|D_t=1}(\tau))} \frac{1}{p} \frac{1}{n} \sum_{i=1}^{n} D_t \Big(\mu_{22}(\Delta Y_{t-1,i}; F^{-1}_{Y_{0t}|D_t=1}(\tau)) \\
&\qquad\qquad + \mu_{32}(Y_{t-1,i}; F^{-1}_{Y_{0t}|D_t=1}(\tau)) + \mu_{42}(Y_{t-2,i}; F^{-1}_{Y_{0t}|D_t=1}(\tau)) \\
&\qquad\qquad + \mu_5(\Delta Y_{t-1,i}, Y_{0t-2,i}; F^{-1}_{Y_{0t}|D_t=1}(\tau)) \Big) \Bigg\}
\end{aligned}$$

the result then follows from an application of the Central Limit Theorem. □

Asymptotic Normality of propensity score reweighted estimator

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

We consider two cases that are likely to be most useful to empirical researchers: (i) when the propensity score is known up to a finite number of parameters, and (ii) when the propensity score is estimated nonparametrically using a series logit estimator as in

Hirano, Imbens, and Ridder (2003). We also have results (available upon request) that provides the conditions and proof of asymptotic normality when the propensity score is semiparametrically using the method of Klein and Spady (1993).

Proof of Theorem 7. At a high level, almost all of the proof of Theorem 5 carries over to Theorem 7). Only Equation 1.61 and $\hat{\mu}_1$ need to be changed. As a first step, we find an asymptotically linear representation of $\sqrt{n}(\hat{\mathrm{F}}_{\Delta Y_{0t}|D_t=1}(\Delta) - \mathrm{F}_{\Delta Y_{0t}|D_t=1}(\Delta))$. Then, we show how this result can be combined with previous results to show asymptotic normality of the estimate of the QTT. When the propensity score is estimated nonparametrically,

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

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

When the propensity score is estimated parametrically,

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

$$
\begin{aligned}
&\times \mathrm{E}\left[\left(\frac{D-G(X^\top\zeta_0)}{G(X^\top\zeta_0)(1-G(X^\top\zeta_0))}\right)^2 g(X^\top\zeta_0)^2 XX^\top\right] \\
&\times \frac{D_{it}-G(X_i^\top\zeta_0)}{G(X_i,\zeta_0)(1-G(X_i^\top\zeta_0))} g(X_i^\top\zeta_0)\Bigg\}\Bigg) \\
\equiv \sqrt{n}\left(\frac{1}{n_C}\frac{(1-p)}{p}\sum_{i\in\mathcal{C}}\Psi_{P1}(Z_i;\Delta y)+\frac{1}{n}\sum_{i=1}^{n}\Psi_{P2}(Z_i;\Delta y)\right)
\end{aligned}
$$

Let[19]

$$
\begin{aligned}
\hat{\mu}_{1N} &= \sqrt{n}\frac{1}{n_C n_T}\frac{(1-p)}{p}\sum_{i\in\mathcal{C}}\sum_{j\in\mathcal{T}}\Psi_{N1}(Z_i; y-F^{-1}_{Y_{0t-1}|D_t=1}(F_{Y_{0t-2}|D_t=1}(Y_{0t-2,j}))) \\
&+\sqrt{n}\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{D_{tj}}{p}\Psi_{N2}(Z_i; y-F^{-1}_{Y_{0t-1}|D_t=1}(F_{Y_{0t-2}|D_t=1}(Y_{0t-2,j})))
\end{aligned}
$$

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

Starting from Equation 1.61 except with $\hat{\mathrm{F}}^{-1}_{\Delta Y_{0t}|D_t=0}(\cdot)$ replaced by the propensity score reweighted $\hat{\mathrm{F}}^{-1}_{\Delta Y_{0t}|D_t=1}(\cdot)$,

$$
\begin{aligned}
&\sqrt{n}\Bigg(\frac{1}{n_T}\sum_{i\in\mathcal{T}}\left[1\{\hat{F}^{-1}_{\Delta Y_{0t}|D_t=1}(\hat{F}_{\Delta Y_{0t-1}|D_t=1}(\Delta Y_{0t-1,i}))\le y-\hat{F}^{-1}_{Y_{0t-1}|D_t=1}(\hat{F}_{Y_{0t-2}|D_t=1}(Y_{0t-2,i}))\}\right] \\
&\qquad -\frac{1}{n_T}\sum_{i\in\mathcal{T}}\left[1\{F^{-1}_{\Delta Y_{0t}|D_t=1}(\hat{F}_{\Delta Y_{0t-1}|D_t=1}(\Delta Y_{0t-1,i}))\le y-\hat{F}^{-1}_{Y_{0t-1}|D_t=1}(\hat{F}_{Y_{0t-2}|D_t=1}(Y_{0t-2,i}))\}\right]-\hat{\mu}_{1N}\Bigg) \\
&\le \sup_z \sqrt{n}\Bigg|\frac{1}{n_T}\sum_{i\in\mathcal{T}}\left[1\{\Delta Y_{0t-1,i}\le \hat{F}^{-1}_{\Delta Y_{0t-1}|D_t=1}(\hat{F}_{\Delta Y_{0t}|D_t=1}(y-\hat{F}^{-1}_{Y_{0t-1}|D_t=1}(\hat{F}_{Y_{0t-2}|D_t=1}(z))))\}\right] \\
&\qquad -\frac{1}{n_T}\sum_{i\in\mathcal{T}}\left[1\{\Delta Y_{0t-1,i}\le \hat{F}^{-1}_{\Delta Y_{0t-1}|D_t=1}(\mathrm{F}_{\Delta Y_{0t}|D_t=1}(y-\hat{F}^{-1}_{Y_{0t-1}|D_t=1}(\hat{F}_{Y_{0t-2}|D_t=1}(z))))\}\right] \\
&\qquad -\frac{1}{n_C}\frac{(1-p)}{p}\sum_{i\in\mathcal{C}}\Psi_{N1}(Z_i; y-F^{-1}_{Y_{0t-1}|D_t=1}(F_{Y_{0t-2}|D_t=1}(z))) \\
&\qquad\qquad +\frac{1}{n}\sum_{i=1}^{n}\Psi_{N2}(Z_i; y-F^{-1}_{Y_{0t-1}|D_t=1}(F_{Y_{0t-2}|D_t=1}(z)))\Bigg|+o_p(1) \\
&\le \sup_z \sqrt{n}\Big|\hat{F}_{\Delta Y_{0t}|D_t=1}(y-\hat{F}^{-1}_{Y_{0t-1}|D_t=1}(\hat{F}_{Y_{0t-2}|D_t=1}(z)))
\end{aligned}
$$

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

$$
\begin{aligned}
&- F_{\Delta Y_{0t}|D_t=1}(y - \hat{F}^{-1}_{Y_{0t-1}|D_t=1}(\hat{F}_{Y_{0t-2}|D_t=1}(z))) \\
&- \Big(\hat{F}_{\Delta Y_{0t}|D_t=1}(y - F^{-1}_{Y_{0t-1}|D_t=1}(F_{Y_{0t-2}|D_t=1}(z))) \\
&\qquad\qquad - F_{\Delta Y_{0t}|D_t=1}(y - F^{-1}_{Y_{0t-1}|D_t=1}(F_{Y_{0t-2}|D_t=1}(z)))\Big)\Big| + o_p(1)
\end{aligned}
$$

which converges to 0 based on Lemma 15. □

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

$$
\begin{aligned}
&\frac{1}{n_C n_T}\frac{(1-p)}{p}\sum_{i\in\mathcal{C}}\sum_{j\in\mathcal{T}}\Psi_{N1}(Z_i; y - F^{-1}_{Y_{0t-1}|D_t=1}(F_{Y_{0t-2}|D_t=1}(Y_{0t-2,j}))) \\
&\qquad = \frac{1}{n_C}\frac{(1-p)}{p}\sum_{i\in\mathcal{C}}\mathrm{E}_{Y_{0t-2}|D_t=1}\left[\Psi_{N2}(Z_i; y - F^{-1}_{Y_{0t-1}|D_t=1}(F_{Y_{0t-2}|D_t=1}(Y_{0t-2}))\right] + o_p(1) \\
&\qquad \equiv \hat{\mu}_{1Ca} + o_p(1) \\
&\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{D_{tj}}{p}\Psi_{P2}(Z_i; y - F^{-1}_{Y_{0t-1}|D_t=1}(F_{Y_{0t-2}|D_t=1}(Y_{0t-2,j}))) \\
&\qquad = \frac{1}{n}\sum_{i=1}^{n}\mathrm{E}_{Y_{0t-2}|D_t=1}\left[\Psi_{P2}(Z_i; y - F^{-1}_{Y_{0t-1}|D_t=1}(F_{Y_{0t-2}|D_t=1}(Y_{0t-2}))\right] + o_p(1) \\
&\qquad \equiv \hat{\mu}_{1Cb} + o_p(1)
\end{aligned}
$$

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

Table 1.1: Summary Statistics

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

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

Table 1.2: QTT Estimates for Job Training Program

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

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

Chapter 2

Job Displacement of Older Workers during the Great Recession: Tight Bounds on Distributional Treatment Effect Parameters using Panel Data

2.1 Introduction

From the official beginning of the Great Recession in December 2007 to October 2009, the U.S. economy shed 8.4 million jobs and the unemployment rate doubled from 5.0% to 10.0%. Reemployment has been slow; many workers have exited the labor force, remain unemployed, or have moved into part time employment (Farber, 2015). This chapter studies the effect of job displacement on older workers during the Great Recession. Using standard Difference in Differences techniques, I find that annual earnings of displaced workers were, on average, 40% lower in 2012 than they would have been had the worker not been displaced. However, the effect of job displacement may be quite heterogeneous across workers. Some older workers may quickly move into similar jobs, some may move into jobs with lower wages or into part time employment, and others may remain unemployed. Understanding this heterogeneity is of interest to researchers and policymakers. For example, the policy response may be quite different if the effect of job displacement is very similar for all individuals compared to the case with very heterogeneous effects.

To understand the heterogeneous effects of job displacement, I develop new tight bounds on distributional treatment effect parameters that exploit having access to panel data. These bounds are much tighter than existing bounds and provide a credible alternative to point identifying assumptions that are not likely to hold in the current application. I find that workers in the 95th percentile of earnings losses due to displacement lose between 90% and 99% of earnings relative to counterfactual earnings had they not been displaced. I also find that at least 13% of workers have higher earnings after displacement than they would have had if they had not been displaced. These findings indicate that there is substantial heterogeneity in the effect of job displacement, but they would not be available using standard approaches to program evaluation.

Learning about heterogeneity in the effect of job displacement requires knowledge of the joint distribution of displaced potential earnings and non-displaced potential earnings for the group of workers that are displaced. However, this joint distribution of potential outcomes is not identified under common identifying assumptions such as selection on observables or even when individuals are randomly assigned to treatment. In each of these cases, although the marginal distributions of displaced and non-displaced potential earnings are identified, the copula – which "couples" the marginal distributions into the joint distribution and captures the dependence between the marginal distributions – is not identified.

To give an example, suppose a researcher is interested in the fraction of workers who have higher earnings following displacement than they would have had if they not been displaced. Further, suppose hypothetically that workers are randomly assigned to being displaced or not being displaced. In this case, the average effect of job displacement is identified – it is given by the difference between average earnings of those who are randomly assigned to be displaced and those who are randomly assigned to not be displaced. But the fraction of workers that benefit from displacement is not identified because, for workers randomly assigned to be displaced (non-displaced), where they *would* be in the distribution of non-displaced (displaced) earnings is not known.

There are many important parameters that depend on the joint distribution of potential outcomes. These include the fraction of individuals that benefit from being treated, the correlation between treated and untreated potential outcomes, the variance of the treatment effect, the quantiles and distribution of the treatment effect itself, and the distribution of the treatment effect conditional on an individual's untreated potential outcome (Heckman, Smith, and Clements, 1997; Firpo and Ridder, 2008). These parameters are not just of theoretical importance. Policymakers may decide to implement a policy based on whether or not a large enough fraction of the population benefits from the policy rather than based on the average benefit of the policy. As another example, policymakers are likely to prefer programs with widespread though smaller benefits to ones where very few people benefit but have extremely large benefits. Finally, for some treatments that are not the direct result of policy decisions, such as job displacement, the response from policymakers may differ in cases where many individuals all experience a small effect of treatment compared to one where a few individuals are very affected.

Existing methods take two polar approaches to identifying the joint distribution of potential outcomes. One idea is to construct bounds on the joint distribution without imposing any assumptions on the unknown dependence (Heckman, Smith, and Clements, 1997; Fan and Park, 2009; Fan and Park, 2012). In the case of job displacement, these bounds are not very informative. The main implications of these bounds are that (i) at least 19% of workers have lower earnings due to displacement than they would have had if they had not been displaced and (ii) the median of the treatment effect is between 79% lower earnings and 118% higher earnings.

Another approach is to assume that the dependence is known. The leading choice is perfect positive dependence.[1] This assumption says that individuals at a given rank in the distribution of earnings following displacement would have the same rank in the distribution of non-displaced potential earnings. This is a very strong assumption as it imposes severe restrictions on how heterogeneous the effect of treatment can be; for example, it prohibits any workers at the top of the distribution of non-displaced earnings from retiring or taking a part time job following displacement. But the assumption is much stronger than that – it prohibits displacement from even swapping the rank of any workers relative to their rank had they not been displaced.

[1]This assumption was first implicitly made in the earliest work on estimating the distributional effects of treatment (Doksum, 1974; Lehmann, 1974) that compared the difference between treated quantiles and untreated quantiles and interpreted this difference as the treatment effect at that quantile. There is also recent work on testing the assumption of perfect positive dependence (Bitler, Gelbach, and Hoynes, 2006; Dong and Shen, 2015; Frandsen and Lefgren, 2015)

In light of (i) the implausibility of existing point-identifying assumptions and (ii) the wide bounds resulting from imposing no assumptions on the missing dependence, I develop new, tighter bounds on parameters that depend on the joint distribution of potential outcomes. Unlike existing work which considers the case of cross-sectional data, I exploit having access to panel data on older workers' annual earnings. Panel data presents a unique opportunity to observe, at least for some individuals, both their displaced and non-displaced potential earnings though these are observed at different points in time. With panel data and under plausible identifying assumptions, the bounds on the joint distribution are much tighter – in theory, the joint distribution could even be point identified. To implement my method requires at least three periods of panel data.

Even though panel data appears to be useful for identifying the joint distribution of potential outcomes, there are still some challenges. In the context of Difference in Differences models, previous work has used panel data to recover missing dependence in the current period from observed dependence in previous periods (Callaway and Li, 2015). But that approach is not possible in the current context because the dependence between displaced and non-displaced potential earnings is never observed – even in previous periods. Instead panel data is informative about the dependence between non-displaced potential earnings over time. One idea would be to assume perfect positive dependence between non-displaced potential earnings over time (Heckman and Smith, 1998). This assumption results in point identification. With three periods of panel data, a researcher could pretest this assumption by checking whether or not perfect positive dependence occurs in the periods before displacement. This assumption is rejected in the current application; intuitively, it requires no changes in ranks of annual earnings over time which is a very strong assumption.

Instead of assuming perfect positive dependence of non-displaced potential earnings over time, I assume that the dependence (or copula) of non-displaced potential earnings over time is the same over time. I call this assumption the Copula Stability Assumption.

Recent work on income mobility decomposes the income at two different points in time into the marginal distributions – which capture inequality – and the copula which captures income mobility (Chetty, Hendren, Kline, and Saez, 2014). Thus, in the context of job displacement, the Copula Stability Assumption requires that, in the absence of job displacement, earnings mobility would be constant over time for the group of displaced workers. Importantly, the Copula Stability Assumption does not restrict the distribution of earnings over time. For example, the distribution of earnings can shift to the right over time or the distribution of earnings can become increasingly unequal over time.

I provide two pieces of evidence in favor of the Copula Stability Assumption. First, I show that the Copula Stability Assumption is likely to be satisfied in a very general model of the type typically estimated in panel data settings. Second, there is empirical evidence in favor of the Copula Stability Assumption. In the United States, despite large increases in inequality, there has been remarkably little change in yearly earnings mobility since the middle of the 20th century (Kopczuk, Saez, and Song, 2010, and also see Figure 2.1).

The final requirement for using my method is that the counterfactual distribution of non-displaced potential earnings for the group of displaced workers must be identified. Because job displacement is not randomly assigned, this requires some type of identifying assumption. I use the Distributional Difference in Differences method (Callaway and Li, 2015) to identify this distribution though the results are not sensitive to using other methods such as selection on observables (Firpo, 2007) as long as a lag of earnings is included as a conditioning variable.

The bounds work in the following way. Let Y_{1t} be displaced potential earnings after displacement, Y_{0t} be non-displaced potential earnings after displacement, and Y_{0t-1} be observed non-displaced earnings before displacement. Existing bounds come from statistical bounds on bivariate distributions when the marginal distributions are known (Hoeffding, 1940; Fréchet, 1951). Under the setup in the current chapter, the joint distributions of (Y_{1t}, Y_{0t-1}) and (Y_{0t}, Y_{0t-1}) are also available. I utilize the following result: for three random variables, when two of the three bivariate joint distributions are known, then bounds on the third bivariate joint distribution are at least as tight as the bounds when only the marginal distributions are known (Joe, 1997).

Consider an extreme example. Suppose Y_{1t} and Y_{0t-1} are perfectly positively dependent and Y_{0t} and Y_{0t-1} are perfectly positively dependent, then Y_{1t} and Y_{0t} must also be perfectly positively dependent. In this case, the extra information from panel data results in point identification. In fact, point identification will occur when either (A) perfect positive dependence is observed between Y_{1t} and Y_{0t-1} or (B) perfect positive dependence is observed between Y_{0t-1} and Y_{0t-2}. The first case is very similar to the leading idea for point identification – perfect positive dependence across treated and untreated potential outcomes – though it also involves an additional time dimension. The second case is exactly the leading assumption for point identification with panel data – perfect positive dependence of untreated potential outcomes over time. Moreover, the bounds are tighter as either (Y_{1t}, Y_{0t-1}) or (Y_{0t}, Y_{0t-1}) becomes more positively dependent. This implies that even when these assumptions are violated, if these assumptions are "close" to holding, my method is robust to these deviations and will deliver tight bounds in precisely this case. Job displacement for older workers falls exactly into this category. Neither type of perfect positive dependence is observed; nonetheless, there is strong positive dependence which results in substantially tighter bounds.

Under the current setup, I am also able to study a parameter I call the Average Treatment Effect on the Treated Conditional on Previous Outcome (ATT-CPO). Although this parameter could be identified under some existing assumptions (for example, an experiment where panel data is also available), it is not available under the Difference in Differences approach used in the current chapter or in the parametric panel data models used in much of the job displacement literature. I find that, on average, earnings losses for workers with higher earnings before the recession are larger in magnitude than earnings losses for workers with lower earnings before the recession; but, as a fraction of earnings, average earnings losses are very similar across the distribution of pre-recession earnings.

In evaluating the effect of job displacement, I focus on estimating the Quantile of the Treatment Effect for the Treated (QoTET) and the ATT-CPO. As a first step, I estimate the counterfactual distribution of non-displaced potential earnings for the group of displaced workers under a Distributional Difference in Differences assumption (Callaway and Li, 2015). The conditions required for that method to estimate the counterfactual distribution hold in the setup of the current chapter. The key requirement for the Distributional Difference in Differences assumption is that the path of non-displaced earnings for the treated group must be the same as the path of earnings for the non-displaced group of workers who have the same observed characteristics.[2]

Next, estimating the QoTET involves estimating several conditional distribution functions which I estimate using local linear kernel estimators. These distributions are straightforward to estimate because they are conditional only on earnings before displacement. I provide point estimates for the QoTET. Estimation of the ATT-CPO is similar; it involves local linear kernel regressions. The estimate of the ATT-CPO converges at a non-parametric rate and is asymptotically normal. Inference for the ATT-CPO is straightforward because the first-step estimates converge at the parametric rate and can be ignored asymptotically.

There are two other approaches to bounding the joint distribution of potential outcomes that should be mentioned. Fan, Guerre, and Zhu (2015) bound parameters that depend on the joint distribution when covariates are available. This approach could theoretically be combined with the approach considered in the current chapter to obtain even tighter bounds at the cost of significantly more challenging estimation. Another assumption that can bound parameters that depend on the joint distribution of potential outcomes is the assumption of Monotone Treatment Response (MTR) (Manski, 1997). Kim (2014) combines this assumption with the statistical bounds approach. MTR would imply that earnings

[2]Being able to condition on observed characteristics is important in the current application because characteristics such as education that are related to whether or not a worker is displaced may also affect the path of earnings in the absence of displacement (Heckman and Smith, 1999; Abadie, 2005).

for displaced workers cannot be larger than earnings would have been had they not been displaced. This assumption is rejected by the bounds developed in the current chapter.

There is some empirical work studying the distributional effects of participating in a program. Djebbari and Smith (2008) use Fréchet-Hoeffding bounds to study the distributional effects of the PROGRESA program in Mexico. Carneiro, Hansen, and Heckman (2003) and Abbring and Heckman (2007), among others, use factor models to identify the joint distribution of treated and untreated potential outcomes.

The outline of the chapter is as follows. Section 2 provides a more specific discussion of the issues involved in identifying the joint distribution of potential outcomes and discusses several parameters of interest. Section 3 contains the main identification results in the chapter. Section 4 discusses estimation and inference. Section 5 applies these results to studying the distributional effects of job displacement for older workers during the Great Recession. Section 6 concludes.

2.2 Background

This section provides some context, motivation, and required details for studying distributional treatment effect parameters. After introducing some notation, it considers distributional treatment effect parameters that depend on the joint distribution of potential outcomes and why there are useful. Finally, it discusses in more detail why the joint distribution of potential outcomes is not identified under conventional identifying asssumptions as well as existing stronger assumptions that point identify the joint distribution.

2.2.1 Treatment Effects Setup

The notation used throughout the chapter is very similar to the notation used in the treatment effects literature in statistics and econometrics. All individuals in the population either participate or do not participate in a treatment. Let $D_t = 1$ for individuals that participate in the treatment and $D_t = 0$ for individuals who do not participate in the treatment (to minimize notation, a subscript i representing each individual is omitted). This chapter considers the case where panel data is available and therefore random variables have a time subscript t. Each individual has potential outcomes in the treated and untreated states at time t which are given by Y_{1t} and Y_{0t}, respectively. But, for each individual, only one of these potential outcomes is observed at each time period. For individuals that are treated in period t, Y_{1t} is observed, but Y_{0t} is not observed. For individuals that are untreated in period t, Y_{0t} is observed but Y_{1t} is unobserved. Let Y_t be the observed outcome in period t; one can then write

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

The main case considered in the chapter is the one where the researcher observes outcomes in three periods implying Y_t, Y_{t-1}, and Y_{t-2} are observed. The researcher may also observe a vector of covariates X which, for simplicity, I assume are time invariant. This assumption can be relaxed with only minor costs which I discuss in more detail below. Throughout most of the chapter, I focus on the case where (i) individuals are first treated in period t and (ii) exactly three periods of panel data are available. Both of these conditions can be relaxed, but they represent the most straightforward conditions for discussing identification in the current setup.

2.2.1.1 Commonly Estimated Parameters

The main problem for researchers interested in understanding the effect of participating in a treatment is that only one potential outcome is observed for any particular individual. This means that the treatment effect itself is never observed. Instead, researchers have focused on identifying functionals of treatment effects and the assumptions that identify these parameters. The most common examples are the Average Treatment Effect (ATE)

$$ATE = \mathrm{E}[Y_{1t} - Y_{0t}]$$

and the Average Treatment Effect on the Treated (ATT)

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

These parameters are identified when the researcher has access to an experiment where individuals are randomly assigned to treatment or under some identifying assumption such as selection on observables. But these average effects only provide a limited summary of the effect of being treated. The next section discusses parameters that are useful for understanding the distributional effects of particpating in a treatment.

Quantile Treatment Effect

The Quantile Treatment Effect (QTE), first studied in Doksum (1974) and Lehmann (1974), is a distributional treatment effect parameter that only requires the marginal distributions to be identified. It is the difference between the quantiles of treated potential outcomes and untreated potential outcomes

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

where, for a random variable X, the τ-quantile $x_\tau = \mathrm{F}^{-1}_X(\tau) \equiv \inf\{x : \mathrm{F}_X(x) \geq \tau\}$. For example, setting $\tau = 0.5$, QTE(0.5) gives the difference between the median of treated potential outcomes and the median of untreated potential outcomes. If policymakers do not care about the identity of individuals in each treatment state, then the QTE fully summarizes the distributional impacts of participating in a program (Sen, 1997; Carneiro, Hansen, and Heckman, 2001).

In most panel data cases, only some fraction of individuals are ever treated. In this case, panel data is typically useful for identifying the related parameter, the Quantile Treatment

Effect on the Treated (QTT), which is given by

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

Both the QTE and the QTT provide some information about the distributional effects of being treated. They are easier to identify than many other distributional treatment effect parameters because they depend only on the marginal distributions of potential outcomes. The next set of parameters depend on the joint distribution of potential outcomes.

2.2.2 Distributional Parameters of Interest

Identifying the joint distribution of potential outcomes is the main identification challenge in the current chapter. But the joint distribution itself is usually not the final object of interest especially because reporting the joint distribution involves a three dimensional plot that is difficult to interpret. This section relates parameters that are useful for understanding the distributional impacts of treatment that depend on the joint distribution of potential outcomes. In many cases, these parameters may be required to properly evaluate the effect of treatment. This section also discuss the differences between several of these parameters and the QTT. Finally, this section considers a parameter that is only available when the researcher has access to panel data which I call the Average Treatment Effect on the Treated Conditional on Previous Outcome (ATT-CPO). To the author's knowledge, it has not been considered before. This parameter proves to be particularly useful in the application to job displacement. If policymakers want to target assistance to individuals that tend to suffer the largest costs of job displacement, this parameter can inform this sort of targeting.

There are many parameters of interest that depend on the joint distribution of treated and untreated potential outcomes rather than just the marginal distributions. First, the fraction of treated individuals that would benefit from treatment depends on the joint distribution

$$\mathrm{P}(Y_{1t} > Y_{0t} | D_t = 1)$$

The distribution of the treatment effect for the treated group (DTET)

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

The fraction of the treated group that has benefits from treatment in a certain range

$$\mathrm{P}(\Delta' \leq Y_{1t} - Y_{0t} \leq \Delta'' | D_t = 1)$$

The distribution of the treatment effect conditional on being in some particular base state

$$\mathrm{P}(Y_{1t} - Y_{0t} \leq \Delta | Y_{0t} = y_0, D_t = 1)$$

The correlation of treated and untreated potential outcomes for the treated group

$$\mathrm{Cor}(Y_{1t}, Y_{0t} | D_t = 1)$$

And with panel data, the distribution of the treatment effect conditional on being in some base state in the previous period

$$\mathrm{P}(Y_{1t} - Y_{0t} \leq \Delta | Y_{0t-1} = y')$$

Also means or quantiles of any of the above distributions may also be of interest. For example inverting the DTET provides the QoTET which is useful for understanding heterogeneity of the treatment effect across individuals and is one of the main parameters considered in the analysis of the effect of job displacement on older workers.[3]

Average Treatment Effect on the Treated Conditional on Previous Outcome

One important limitation of the QTT and the QoTET is that they can fail to provide information on which types of individuals experience the largest effects of treatment. For example, with panel data one can address whether workers with high earnings or low earnings in the previous period experience larger effects of job displacement. This parameter, which I term the Average Treatment Effect on the Treated Conditional on Previous Outcomes (ATT-CPO)[4] is given by

$$\text{ATT-CPO}(y') = \mathrm{E}[Y_{1t} - Y_{0t} | Y_{0t-1} = y', D_t = 1]$$

In the current framework, this parameter is point identified because the joint distributions $\mathrm{F}_{Y_{1t}, Y_{0t-1}|D_t=1}(y_1, y')$ and $\mathrm{F}_{Y_{0t}, Y_{0t-1}|D_t=1}(y_0, y')$ are identified, and the ATT-CPO depends only on these joint distributions and not on the joint distribution of treated and untreated potential outcomes for the treated group in period t. The next example provides a case where a combination of the ATT-CPO and the QoTET can provide a better understanding of the distributional effect of treatement compared to the QTT.

[3]All the parameters mentioned above condition on being part of the treated group, but one may also be interested in these parameters for the entire population. Panel data is most useful for identifying parameters conditional on being part of the treated group. Using the techniques presented in the current chapter can still lead to bounds on parameters for the entire population by combining the bounds for the treated group presented in the current chapter with bounds for the untreated group coming from existing statistical bounds. These bounds will be tighter if a larger fraction of the population is treated. I do not pursue bounds on parameters for the entire population throughout the rest of the chapter.

[4]Heckman, Smith, and Clements (1997) suggest the related parameter $\mathrm{F}_{Y_{1t}-Y_{0t}|Y_{0t}=y'}(\Delta|y')$. This is the distribution of the treatment effect conditional on the base state Y_{0t} taking some particular value y. It is an interesting parameter, but it suffers from being difficult to display graphically because in most cases a researcher is interested in this parameter while varying y' in many values of its support. A plot of the result would be a three dimensional and difficult to interpret. An alternative measure is $\mathrm{E}[Y_{1t} - Y_{0t} | Y_{0t} = y']$. This is the average treatment effect conditional on the base state Y_{0t} taking some particular value y'. Varying y' results in an easy to interpret two dimensional plot. However, in the current setup, this parameter is not point identified because it depends on the joint distribution of treated and untreated potential outcomes which is only partially identified in the current setup.

Example 1

Suppose there are 10 workers in the population that are observed in two periods. In the first period, suppose each worker's earnings are given by their number; worker 1 has earnings of 1, worker 2 has earnings of 2, etc.). In the second period, in the absence of displacement, suppose each worker keeps the same earnings as in the first period. For worker 10, suppose his earnings decrease to 0 if he is displaced, but for the other workers, suppose they are able to find a new job with the same earnings. In this case, the QTE is constant everywhere and equal to -1. In much applied research, this effect would wrongly be interpreted as the effect of displacement being the same across workers with high and low earnings. On the other hand, $QoTE(0.1) = -10$ and $QoTE(\tau) = 0$ for $\tau > 0.1$. Immediately, this would imply that the effect of treatment is very heterogeneous – one worker has much lower earnings due to being displaced while most workers experience no effect of displacement of earnings. Also, ATT-CPO$(10) = -10$ but ATT-CPO$(y') = 0$ for other values of y'. This would imply the effect of displacement is strongest for workers with highest earnings.

This example is less extreme than it appears. If any older workers from the middle or top of the non-displaced potential earnings distribution move to the lower part of the distribution of earnings following displacement – which could happen due to difficulty finding new employment, moving to part time work, or retiring – then the QTT will be very difficult to interpret. However, the QoTET and the ATT-CPO can still be very helpful to understand the distributional effects of displacement.

2.2.3 The Identification Issue and Existing Solutions

This section explains in greater detail the fundamental reason why the joint distribution of potential outcomes is not point identified except under strong assumptions. First, I assume that both the marginal distribution of treated potential outcomes for the treated group

$F_{Y_{1t}|D_t=1}(y_1)$ and the marginal distribution of untreated potential outcomes for the treated group $F_{Y_{0t}|D_t=1}(y_0)$ are identified. The first can be obtained directly from the data; the second is obtained under some identifying assumption which is assumed to be available. Sklar (1959) demonstrates that joint distributions can be written as the copula function of marginal distributions in the following way

$$F_{Y_{1t},Y_{0t}|D_t=1}(y_1,y_0) = C_{Y_{1t},Y_{0t}|D_t=1}\left(F_{Y_{1t}|D_t=1}(y_1), F_{Y_{0t}|D_t=1}(y_0)\right) \tag{2.1}$$

where $C_{Y_{1t},Y_{0t}|D_t=1}(\cdot,\cdot) : [0,1]^2 \to [0,1]$. This representation highlights the key piece of missing information under standard assumptions – the copula function. Using results from the statistics literature, one can still construct the so-called Fréchet-Hoeffding bounds on the joint distribution. These bounds arise from considering two extreme cases: (i) when there is perfect positive dependence between the two marginal distributions and (ii) when there is perfect negative dependence between the two distributions. Heckman, Smith, and Clements (1997) follow this procedure and find that it leads to very wide bounds in general. Moreover, that paper points out that under strong forms of negative dependence, the bounds do not seem to make sense in an application on the treatment effect of participating in a job training program.

At the other extreme, one could posit a guess for the copula. In the cross-sectional case, the most common assumption is perfect positive dependence between treated potential outcomes and untreated potential outcomes for the treated group. This assumption can be written as

$$F_{Y_{1t}|D_t=1}(Y_{1t}) = F_{Y_{0t}|D_t=1}(Y_{0t})$$

which implies that

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

which means that for any individual in the treated group with observed outcome Y_{1t}, their counterfactual untreated potential outcome Y_{0t} is also known which implies that the joint distribution is point identified. Although this assumption might be more plausible than assuming independence or perfect negative dependence, it seems very unlikely to hold in practice because it severely restricts the ability of treatment to have different effects across different individuals. But the idea that different individuals can experience different effects of treatment is one of the central themes of the entire treatment effects literature. In the context of job displacement, perfect positive dependence seems unlikely to hold because it would prohibit individuals at the top of the pre-displacement earnings distribution from being unemployed or retiring following job displacement.

With panel data, perhaps a more plausible assumption is perfect positive dependence in untreated potential outcomes over time (this idea is mentioned in Heckman and Smith, 1998):

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

This assumption does not directly replace the unknown copula in Equation 2.1, but the next lemma establishes that this assumption also leads to point identification of the joint distribution of potential outcomes.

LEMMA 21. *Under perfect positive dependence between untreated potential outcomes for the treated group over time,*

$$\mathrm{F}_{Y_{1t},Y_{0t}|D_t=1}(y_1,y_0) = \mathrm{F}_{Y_{1t},Y_{0t-1}|D_t=1}(y_1,\mathrm{F}^{-1}_{Y_{0t-1}|D_t=1}(\mathrm{F}_{Y_{0t}|D_t=1}(y_0)))$$

When the researcher has access to more than two periods of panel data, one can apply a sort of pre-test to this assumption. That is, one can check whether perfect positive dependence in untreated potential outcomes holds between periods $t-1$ and $t-2$ and this can provide evidence as to whether or not perfect positive dependence is likely to hold between periods t and $t-1$. In the application in the current period, I find that this assumption does not hold, but it is not too far from holding; in other words, in the absence of being displaced, older workers do change ranks in the distribution of earnings over time, but, for the most part, the change in ranks is small.

2.3 Identification

This section provides the main identification results of the chapter. It provides bounds for the joint distribution of potential outcomes for the treated group, the distribution of the treatment effect for the treated group (DTET), and the quantile of the treatment effect for the treated group (QoTET). It also provides point identification results for the Average Treatment Effect on the Treated Conditional on Previous Outcome (ATT-CPO).

The following are the main assumptions used in the current chapter.

ASSUMPTION 11. *(Data)*

(a) There are three periods of panel data $\{Y_t, Y_{t-1}, Y_{t-2}, D_t, X\}$

(b) No one is treated before period t

Assumption 11(a) says that the researcher has access to three periods of panel data. The researcher possibly observes some covariates X which I assume, as is common in the treatment effects literature, are time invariant though this assumption can be relaxed with only minor costs. Assumption 11(b) is a standard assumption in the Difference in Differences literature. It can be relaxed at the cost of either (i) changing the identified parameters to be conditional on being part of the "newly treated group" - the group that first becomes treated at time period t, or (ii) some additional assumptions that say that the effect on the newly treated group is the same as the effect on the treated group overall. Under this assumption, the researcher observes untreated potential outcomes for members of both the treated and untreated group in periods $t-1$ and $t-2$. In period t, the researcher observes treated potential outcomes Y_{1t} for members of the treated group and untreated potential outcomes Y_{0t} for members of the untreated group.

I focus on the case with exactly three periods. Under the condition that no one is treated until the last period, having additional pre-treatment periods can lead to tighter bounds on the joint distribution of potential outcomes. In the more general case where individuals can first become treated at some period before the last period, one can still construct bounds on the joint distribution of potential outcomes using similar techniques.

ASSUMPTION 12. *(Marginal Distributions)*

(a) The marginal distributions $\mathrm{F}_{Y_{1t}|D_t=1}(y_1)$ *and* $\mathrm{F}_{Y_{0t}|D_t=1}(y_0)$ *are identified.*

(b) The conditional distributions $\mathrm{F}_{Y_{1t}|X,D_t=1}(y_1)$ *and* $\mathrm{F}_{Y_{0t}|X,D_t=1}(y_0)$ *are identified.*

The first part of Assumption 12(a) says that the distribution of treated potential outcomes for the treated group is identified. This follows directly from Assumption 11 because treated potential outcomes are observed for the treated group. The second part says that the distribution of untreated potential outcomes for the treated group is identified. This distribution is counterfactual and its identification requires some identifying assumption. But this is precisely the distribution that most work on identifying the QTT with observations over time identifies (examples include Athey and Imbens, 2006; Bonhomme and Sauder, 2011; Chernozhukov, Fernández-Val, Hahn, and Newey, 2013; Callaway and Li, 2015). This counterfactual distribution is also available under the selection on observables assumption (Firpo, 2007) and could be available under an instrumental variables assumption (Abadie, Angrist, and Imbens, 2002; Chernozhukov and Hansen, 2005; Carneiro and Lee, 2009; Frölich and Melly, 2013). Assumption 12(b) is a stronger assumption that says that the conditional versions of each marginal distribution are identified. These can be useful in the case where the Copula Stability Assumption (below) holds conditional on covariates. Also, when the conditional distributions are available, they can be useful for further tightening the bounds on the joint distribution (Fan, Guerre, and Zhu, 2015). If Assumption 12(b) holds, it implies that the marginal distributions in Assumption 12(a) will also be available; throughout most of the chapter, I use Assumption 12(a) for simplicity.

The next assumption is the main identifying assumption in the chapter.

Copula Stability Assumption. *For* $(u,v) \in [0,1]^2$

$$\mathrm{C}_{Y_{0t},Y_{0t-1}|D_t=1}(u,v) = \mathrm{C}_{Y_{0t-1},Y_{0t-2}|D_t=1}(u,v)$$

Conditional Copula Stability Assumption. *For* $(u,v) \in [0,1]^2$

$$\mathrm{C}_{Y_{0t},Y_{0t-1}|X,D_t=1}(u,v|x) = \mathrm{C}_{Y_{0t-1},Y_{0t-2}|X,D_t=1}(u,v|x)$$

The Copula Stability Assumption says that the dependence between untreated potential outcomes at periods t and $t-1$ is the same as the dependence between untreated potential outcomes at periods $t-1$ and $t-2$. This assumption is useful because the dependence between untreated potential outcomes at period t and period $t-1$ is not observed. Although, by assumption, the counterfactual distribution of untreated potential outcomes for the treated group, $\mathrm{F}_{Y_{0t}|D_t=1}(y_0)$, is identified and the distribution of untreated potential outcomes for the treated at period $t-1$, $\mathrm{F}_{Y_{0t-1}|D_t=1}(y')$, is identified because untreated potential outcomes are observed for the treated group at period $t-1$, their dependence is not identified because Y_{0t} and Y_{0t-1} are not simultaneously observed for the treated group. The Copula Stability Assumption recovers the missing dependence. This implies that the joint distribution of untreated potential outcomes at times t and $t-1$ for the treated group, $\mathrm{F}_{Y_{0t},Y_{0t-1}|D_t=1}(y_0,y')$, is identified. This joint distribution is not of primary interest in the current chapter. But knowledge of this joint distribution is important for deriving tight bounds on the distributions and parameters of interest.

To better understand the Copula Stability Assumption, it is helpful to consider some examples. As a first example, the Copula Stability Assumption says that if untreated potential outcomes at period $t-1$ are independent (or perfectly positively dependent) of untreated potential at period $t-2$, then untreated outcomes at period t will continue to be independent (or perfectly positively dependent) of untreated outcomes at period $t-1$. Or, for example, suppose the copula for $(Y_{0t-1},Y_{0t-2}|D_t=1)$ is Gaussian with parameter ρ, the

Copula Stability Assumption says that the copula for $(Y_{0t}, Y_{0t-1}|D_t = 1)$ is also Gaussian with parameter ρ though the marginal distributions of outcomes can change in unrestricted ways. For example, the distribution of earnings could be increasing over time or could become more unequal over time. Likewise, if the copula is Archimedean, the Copula Stability Assumption says that the generator function does not change over time. For Archimedean copulas with a scalar parameter having a one-to-one mapping to dependence parameters such as Kendall's Tau or Spearman's Rho (examples include common Archimedean copulas such as the Clayton, Frank, and Gumbel copulas), the Copula Stability Assumption says that the dependence parameter is the same over time.

The Conditional Copula Stability Assumption may be more plausible in many applications. It says that the dependence is the same over time conditional on some covariates X. For example, earnings over time may be more strongly positively dependent for older workers than for younger workers. It should be noted, however, that the unconditional Copula Stability Assumption does not preclude covariates affecting outcomes; but it does place some restrictions on how covariates can affect the outcome of interest and especially how the effect of covariates changes over time – this issue is discussed more in Section 2.3.1.1 below.

The next result is a simple application of Fréchet-Hoeffding bounds to a conditional distribution; it provides an important building block for constructing tighter bounds on the joint distribution of potential outcomes.

LEMMA 22.

$$\mathrm{F}^{L}_{Y_{1t},Y_{0t}|Y_{0t-1}=y',D_t=1}(y_1,y_0|y') \leq \mathrm{F}_{Y_{1t},Y_{0t}|Y_{0t-1}=y',D_t=1}(y_1,y_0|y') \leq \mathrm{F}^{U}_{Y_{1t},Y_{0t}|Y_{0t-1}=y',D_t=1}(y_1,y_0|y')$$

where

$$\mathrm{F}^{L}_{Y_{1t},Y_{0t}|Y_{0t-1}=y',D_t=1}(y_1,y_0|y') = \max\{\mathrm{F}_{Y_{1t}|Y_{0t-1},D_t=1}(y_1|y') + \mathrm{F}_{Y_{0t}|Y_{0t-1},D_t=1}(y_0|y') - 1, 0\}$$

$$F^U_{Y_{1t},Y_{0t}|Y_{0t-1}=y',D_t=1}(y_1,y_0|y') = \min\{F_{Y_{1t}|Y_{0t-1},D_t=1}(y_1|y'), F_{Y_{0t}|Y_{0t-1},D_t=1}(y_0|y')\}$$

THEOREM 8.

$$F^L_{Y_{1t},Y_{0t}|D_t=1}(y_1,y_0) \leq F_{Y_{1t},Y_{0t}|D_t=1}(y_1,y_0) \leq F^U_{Y_{1t},Y_{0t}|D_t=1}(y_1,y_0)$$

where

$$F^L_{Y_{1t},Y_{0t}|D_t=1}(y_1,y_0) = E[F^L_{Y_{1t},Y_{0t}|Y_{0t-1}=y',D_t=1}(y_1,y_0|Y_{0t-1})]$$
$$F^U_{Y_{1t},Y_{0t}|D_t=1}(y_1,y_0) = E[F^U_{Y_{1t},Y_{0t}|Y_{0t-1}=y',D_t=1}(y_1,y_0|Y_{0t-1})]$$

and these bounds are sharp.

The bounds in Theorem 8 warrant some more discussion. First, these bounds will be tighter than the bounds without using panel data unless Y_{0t-1} is independent of Y_{1t} and Y_{0t}. But in most applications in economics Y_{0t} and Y_{0t-1} are likely to be positively dependent. On the other hand, the joint distribution will be point identified if either (i) Y_{1t} and Y_{0t-1} are perfectly positively dependent or (ii) Y_{0t} and Y_{0t-1} are perfectly positively dependent. Item (i) is very similar to the assumption of perfect positive dependence across treated and untreated groups (though it also includes a time dimension); Item (ii) is exactly the condition of perfect positive dependence in untreated potential outcomes over time used as a point identifying assumption (Heckman and Smith, 1998). Together, these conditions imply that if either one of two natural limiting conditions hold in the data, then the joint distribution of potential outcomes will be point identified. Moreover, intuitively the bounds will be tighter in cases that are "closer" to either of these two limiting cases. This means that even in the case where the limiting conditions do not hold exactly, one is still able to (substantially) tighten the bounds that would arise in the case without panel data. I provide the intuition for this point next.

Example 2

Spearman's Rho is the correlation of the ranks of two random variables; i.e. $\rho_S = Corr(\mathrm{F}_1(X_1), \mathrm{F}_2(X_2))$. Bounds on Spearman's Rho can be derived when two out of three joint distributions and all marginal distributions (exactly our case) are known. Because the marginal distributions $\mathrm{F}_{Y_{1t}|D_t=1}(Y_{1t})$, $\mathrm{F}_{Y_{0t}|D_t=1}(Y_{0t})$, and $\mathrm{F}_{Y_{0t-1}|D_t=1}(Y_{0t-1})$ are uniformly distributed, their covariance matrix is given by

$$\mathrm{Cov}\left(\mathrm{F}_{Y_{1t}|D_t=1}(Y_{1t}), \mathrm{F}_{Y_{0t}|D_t=1}(Y_{0t}), \mathrm{F}_{Y_{0t-1}|D_t=1}(Y_{0t-1})\right) = \begin{pmatrix} 1 & \rho_{12} & \rho_{13} \\ \rho_{12} & 1 & \rho_{23} \\ \rho_{13} & \rho_{23} & 1 \end{pmatrix}$$

Consider the case where ρ_{13} and ρ_{23} are identified and ρ_{12} is not known. ρ_{12} is partially identified because the covariance matrix must be positive semi-definite.

This results in the condition that

$$\rho_{13}\rho_{23} - \sqrt{\rho_{13}^2\rho_{23}^2 + (1-\rho_{13}^2-\rho_{23}^2)} \leq \rho_{12} \leq \rho_{13}\rho_{23} + \sqrt{\rho_{13}^2\rho_{23}^2 + (1-\rho_{13}^2-\rho_{23}^2)}$$

The width of the bounds is given by

$$width = 2\sqrt{\rho_{13}^2\rho_{23}^2 + (1-\rho_{13}^2-\rho_{23}^2)}$$

It is easy to show that for fixed ρ_{23} with $|\rho_{23}| < 1$, the width of the bounds on ρ_{12} are decreasing as ρ_{13} increases for $\rho_{13} > 0$, and width of the bounds are decreasing as ρ_{13} decreases for $\rho_{13} < 0$. When either ρ_{13} or ρ_{23} is equal to one in absolute value, ρ_{12} is point identified. This corresponds exactly to the case of perfect positive dependence (or perfect negative dependence) mentioned above for point identification. The intuition of this result is that as the copula moves "closer" to perfect positive dependence or perfect negative dependence, the bounds on the joint distribution of interest shrink.

Remark

By a similar reasoning, knowledge of conditional distributions, conditional on X, will also serve to tighten the bounds. See especially Fan, Guerre, and Zhu (2015). The same logic applies in that case, though with covariates, the natural cases that lead to point identification with panel data do not have straightforward counterparts.

Just as knowledge of $\mathrm{F}_{Y_{1t},Y_{0t-1}|D_t=1}(y_1,y')$ and $\mathrm{F}_{Y_{0t},Y_{0t-1}|D_t=1}(y_0,y')$ leads to bounds on the joint distribution of interest $\mathrm{F}_{Y_{1t},Y_{0t}|D_t=1}(y_1,y_0)$, knowledge of these distributions can also be used to bound the DTET, the QoTET, and other parameters that depend on the joint distribution. These results are presented next.

Sharp bounds on the distribution of the treatment effect are known in the case where there is no additional information besides the marginal distributions (Fan and Park, 2010). These bounds are obtained using results from the statistics literature for the distribution of the difference of two random variables when the marginal distributions are fixed (Makarov, 1982; Rüschendorf, 1982; Frank, Nelsen, and Schweizer, 1987; Williamson and Downs, 1990). I use these same bounds for the conditional joint distribution.

LEMMA 23. *(Conditional Distribution of the Treatment Effect)*

$$\mathrm{F}^L_{Y_{1t}-Y_{0t}|Y_{0t-1},D_t=1}(\Delta|y') \leq \mathrm{F}_{Y_{1t}-Y_{0t}|Y_{0t-1},D_t=1}(\Delta|y') \leq \mathrm{F}^U_{Y_{1t}-Y_{0t}|Y_{0t-1},D_t=1}(\Delta|y')$$

where

$$\mathrm{F}^{L}_{Y_{1t}-Y_{0t}|Y_{0t-1},D_t=1}(\Delta|y') = \sup_{y} \max\{\mathrm{F}_{Y_{1t}|Y_{0t-1},D_t=1}(y|y') - \mathrm{F}_{Y_{0t}|Y_{0t-1},D_t=1}(y-\Delta|y'), 0\}$$

$$\mathrm{F}^{U}_{Y_{1t}-Y_{0t}|Y_{0t-1},D_t=1}(\Delta|y') = 1 + \inf_{y} \min\{\mathrm{F}_{Y_{1t}|Y_{0t-1},D_t=1}(y|y') - \mathrm{F}_{Y_{0t}|Y_{0t-1},D_t=1}(y-\Delta|y'), 0\}$$

THEOREM 9. *(Distribution of the Treatment Effect)*

$$\mathrm{F}^{L}_{Y_{1t}-Y_{0t}|D_t=1}(\Delta) = \mathrm{E}[\mathrm{F}^{L}_{Y_{1t}-Y_{0t}|Y_{0t-1},D_t=1}(\Delta|Y_{0t-1})]|D_t = 1]$$

$$\mathrm{F}^{U}_{Y_{1t}-Y_{0t}|D_t=1}(\Delta) = \mathrm{E}[\mathrm{F}^{U}_{Y_{1t}-Y_{0t}|Y_{0t-1},D_t=1}(\Delta|Y_{0t-1})]|D_t = 1]$$

where $\mathrm{F}^{L}_{Y_{1t}-Y_{0t}|Y_{0t-1},D_t=1}(\Delta|y')$ *and* $\mathrm{F}^{L}_{Y_{1t}-Y_{0t}|Y_{0t-1},D_t=1}(\Delta|y')$ *are given in Lemma 23. These bounds are sharp.*

Sharp bounds on the QoTET can be obtained from the bounds on the DTET. The upper bound on the QoTET comes from inverting the lower bound of the DTET, and the lower bound on the QoTET comes from inverting the upper bound on the DTET.

THEOREM 10. *(Quantile of the Treatment Effect)*

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

where

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

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

and $\mathrm{F}^{L}_{Y_{1t}-Y_{0t}|D_t=1}(\Delta)$ *and* $\mathrm{F}^{U}_{Y_{1t}-Y_{0t}|D_t=1}(\Delta)$ *are given in Theorem 9. These bounds are sharp.*

Point Identification of the ATT-CPO

Next, I show that the ATT-CPO is identified in the current setup. The reason why this parameter is point identified is that it depends on the joint distributions $\mathrm{F}_{Y_{1t},Y_{0t-1}|D_t=1}(y_1,y')$ and $\mathrm{F}_{Y_{0t},Y_{0t-1}|D_t=1}(y_0,y')$ which are both point identified, but it does not depend on the joint distribution of treated and untreated potential outcomes $\mathrm{F}_{Y_{1t},Y_{0t}|D_t=1}(y_1,y_0)$ which is only partially identified. Point identification of the ATT-CPO requires the following assumption

ASSUMPTION 13. *(Distribution of Untreated Potential Outcomes)*

Y_{0t-1} and Y_{0t-2} are continuously distributed.[5]

This assumption allows for the quantile functions in the expression below to be well-defined.

The next result provides an explicit expression for this result that can be estimated using the observed data and the identified marginal distributions.

THEOREM 11.

$$
\begin{aligned}
\textit{ATT-CPO}(y') &= \mathrm{E}[Y_{1t} - Y_{0t}|Y_{0t-1} = y'] \\
&= \mathrm{E}[Y_{1t}|Y_{0t-1} = y', D_t = 1] \\
&\quad - \mathrm{E}[\mathrm{F}^{-1}_{Y_{0t}|D_t=1}(\mathrm{F}_{Y_{0t-1}|D_t=1}(Y_{0t-1})) \mid Y_{0t-2} = \mathrm{F}^{-1}_{Y_{0t-2}|D_t=1}(\mathrm{F}_{Y_{0t-1}|D_t=1}(y'))]
\end{aligned}
$$

2.3.1 More Evidence on the Copula Stability Assumption

Because the key identifying assumption in the chapter, the Copula Stability Assumption, is new to the literature on evaluating the distributional impacts of program participation, this section considers whether or not it is likely to hold in applications. The first

[5]This condition could be weakened to

$$
\begin{aligned}
\mathrm{Range}(\mathrm{F}_{Y_{0t}|D_t=1}) &\subseteq \mathrm{Range}(\mathrm{F}_{Y_{0t-1}|D_t=1}) \\
\mathrm{Range}(\mathrm{F}_{Y_{0t-2}|D_t=1}) &\subseteq \mathrm{Range}(\mathrm{F}_{Y_{0t-1}|D_t=1})
\end{aligned}
$$

for the ATT-CPO to be identified which would allow for some mass points in the distributions of untreated potential outcomes.

contribution of this section is to consider several models of varying generality and discusses whether or not they are consistent with the Copula Stability Assumption. The key requirement for a model to satisfy the Copula Stability Assumption is that the way unobservables affect outcomes cannot change over time. This requirement allows for untreated outcomes to be a nonseparable function of observed covariates, time-varying unobservables, and time invariant unobservables that can be correlated with observed covariates plus a time varying function of observed covariates. Second, this section provides empirical evidence in favor of the Copula Stability Assumption in the case where the outcome is yearly earnings in the United States.

2.3.1.1 Models that are Consistent with the Copula Stability Assumption

An important question is whether or not the Copula Stability Assumption is likely to hold in the types of models that economists most frequently use. This section shows that the Copula Stability Assumption holds in most of the cases most frequently considered in panel data or Difference in Differences settings such as Abadie (2005). Examples include (i) unobserved heterogeity distributed differently across treated and control groups and (ii) different time trends in untreated potential outcomes for observations with different observable characteristics. The key restriction is that the effect of unobservables cannot change over time. When the effect of time invariant unobserved heterogeneity changes over time – for example, if the return to unobserved ability is increasing over time – then, the Copula Stability Assumption will not hold; however, in this case, panel data assumptions and Difference in Differences assumptions would also be violated which implies that this is not a unique restriction to the methods considered in the current chapter. On the other hand, if the effect of time varying unobservables changes over time, then a Difference in Differences approach to identifying the average effect of participating in the treatment would still be valid, but the Copula Stability Assumption would not hold.

Let c_i be a time invariant unobservable whose distribution can be different for the treated and untreated groups (though this is not important for the Copula Stability Assumption as we only consider untreated outcomes for the treated group) and v_{it} is a time varying unobservable satistfying $\mathrm{F}_{v_{it}|X,c_i}(v) = F_v(v)$ which allows for serial correlation. The data generating process for treated potential outcomes can be left completely unrestricted as the Copula Stability Assumption only concerns untreated potential outcomes.

Model 1:

$Y_{0t} = g(X_i, c_i, v_{it})$. This model is stationary in the sense that the same inputs produce the same outcomes in every time period though it is similar to models in recent work on identifying nonseparable models with panel data (for example Evdokimov, 2010; Chernozhukov, Fernández-Val, Hahn, and Newey, 2013). Both the unconditional Copula Stability Assumption and the Conditional Copula Stability Assumption hold in this model. Difference in Differences techniques would be useful for this model. It includes as a special case the model $Y_{0t} = X_i'\beta + c_i + v_{it}$.

Model 2:

$Y_{0t} = g(X_i, c_i, v_{it}) + h_t(X_i)$. This model generalizes the previous model in that it allows for a trend in outcomes that can differ based on observable characteristics. The Conditional Copula Stability Assumption holds in this model, but the unconditional Copula Stability Assumption does not. This model includes as a special case $Y_{0t} = X_i'\beta_t + c_i + \theta_t + v_{it}$ where θ_t is an aggregate time fixed effect. However, this model also allows for the possibility of a much more general trend as a function of the observed covariates. The common aggregate time effect is a sufficient condition for the unconditional Copula Stability Assumption to hold even though there are covariates present in the model. In this type of model, one can include time varying observable characteristics – the key requirement is that they be additively separable from unobservables.

Next, I consider two models where neither Copula Stability Assumption holds. In the first, it would be possible to use a Difference in Differences approach to estimate the ATT.

The second provides a case where neither a Difference in Differences Assumption nor the Copula Stability Assumption is valid.

Model 3:

$Y_{0t} = g(X_i, c_i, v_{it}) + h_t(X_i, v_{it})$. This model provides an example where the Copula Stability Assumption does not hold, but a semiparametric Difference in Differences approach would still be valid. The reason the Copula Stability Assumption does not hold is that the model effectively allows the effect of unobservables to change over time allowing an individual's place in the distribution of untreated potential outcomes to change in an unrestricted way that cannot be handled by the Copula Stability Assumption. For example, past evidence of very little mobility in outcomes over time does not provide evidence that there will be very little mobility in the next period in this model. In the context of parametric panel data models, there do not appear to be any well known cases that this model covers that are not covered by the Model 3.

Model 4:

$Y_{0t} = g(X_i, c_i, v_{it}) + h_t(X_i, c_i)$. The Copula Stability Assumption does not hold in this model for the same reason that it did not hold in Model 4 – the Copula Stability Assumption cannot allow for the effect of unobservables to change in an unrestricted way across time periods. Panel data techniques and semiparametric Difference in Differences would not work in this model either though because the path of untreated potential outcomes for the untreated group will, in general, not be the same as the path of untreated potential outcomes for the treated group. One example of this sort of model is the random growth model of Heckman and Hotz (1989): $Y_{0t} = X_i'\beta + c_i + g_i t + v_{it}$.

2.3.1.2 Empirical Evidence on the Copula Stability Assumption

This section provides some empirical evidence that the Copula Stability Assumption may be valid when the outcome of interest is yearly income – a leading case in labor economics. In this case, the Copula Stability Assumption says that income mobility, which has been interpreted as the copula of income over time in studies of mobility (Chetty, Hendren, Kline, and Saez, 2014) or very similarly as correlation between the ranks of income over time (Kopczuk, Saez, and Song, 2010),[6] is the same over time.[7]

A simple way to check if the copula is constant over time is to check if some dependence measure such as Spearman's Rho or Kendall's Tau is constant over time.[8] Using administrative data from 1937-2003, Kopczuk, Saez, and Song (2010) find that the rank correlation (Spearman's Rho) of yearly income is nearly constant in the U.S. Immediately following World War II, there was a slight decline in income mobility. Since then, there has been remarkable stability in income mobility (See Figure 2.1).

Moreover, Figure 2.1 also confirms the intuition that there is strong positive dependence of yearly income over time though the dependence is less than perfect positive dependence. This is precisely the case where the method developed in the current chapter is likely to (i) provide more credible results than employing a perfect positive dependence over time assumption while (ii) yielding much tighter bounds on the joint distribution of potential outcomes than would be available using other methods that rely on purely statistical results to bound distributional treatment effects that depend on the joint distribution of potential outcomes.

[6]The dependence measure Spearman's Rho is exactly the correlation of ranks. Dependence measures such as Spearman's Rho or Kendall's Tau are very closely related to copulas; for example, these dependence measures depend only on the copula of two random variables not the marginal distributions. Dependence measures also have the property of being ordered. For example, larger Spearman's Rho indicates more positive dependence; two copulas, on the other hand, cannot generally be ordered. See Nelsen (2007) and Joe (2015) for more discussion on the relationship between dependence measures and copulas.

[7]It is also very similar to other work in the income mobility literature that considers transitions from one quintile of earnings in one period to another quintile of earnings in another period (Hungerford, 1993; Gottschalk, 1997; Carroll, Joulfaian, and Rider, 2007)

[8]It is possible for a copula to change over time and have the same value of the dependence measure, but if the dependence measure changes over time, then the copula necessarily changes over time.

2.4 Estimation

This section shows how to estimate the QoTET and the ATT-CPO under the identification results presented above and supposing that an estimate of the counterfactual distribution of untreated potential outcomes for the treated group, $\hat{\mathrm{F}}_{Y_{0t}|D_t=1}(y_0)$, is available. The second part of this section discusses inference for the ATT-CPO.

Estimating the QoTET

Estimation of the QoTET is based on the results of Lemma 23, Theorem 9, and Theorem 10. Broadly speaking, it is possible to use plug-in estimators for every term except $\mathrm{F}_{Y_{0t}|Y_{0t-1},D_t=1}(y_0|y')$. This term is identified under the Copula Stability Assumption, but it is not immediate how to estimate it. I consider how to estimate this term in Step 2 below.

Step 1: Estimate $\mathrm{F}_{Y_{1t}|Y_{0t-1},D_t=1}(y_1|y')$*:*

To estimate the distribution of treated potential outcomes conditional on previous untreated potential outcomes for the treated group, $\mathrm{F}_{Y_{1t}|Y_{0t-1},D_t=1}(y_1|y')$, I use a local linear kernel estimator. This estimator solves

$$\min_{\alpha_1,\beta_1} \sum_{i\in\mathcal{T}} [1\{Y_{it} \leq y_1\} - \alpha_1 - (Y_{it-1} - y')\beta_1]^2 K_h(Y_{it-1} - y')$$

This is easy to estimate as it is simply weighted least squares. Let $\gamma_1(y_1|y') = [\alpha_1(y_1|y'), \beta_1(y_1|y')]'$, $\tilde{\mathcal{Y}}_1$ be an $n_T \times 1$ vector with ith component $1\{Y_{it} \leq y_1\}$, $\tilde{\mathcal{X}}_1$ be an $n_T \times 2$ matrix with ith row given by $(1, Y_{it-1} - y')$, and $\tilde{\mathcal{K}}_1(y')$ be an $n_T \times n_T$ diagonal matrix with ith diagonal element given by $K_h(Y_{it-1} - y')$. Then, a closed form expression for the estimate of $\gamma_1(y')$ is

$$\hat{\gamma}_1(y_1|y') = (\tilde{\mathcal{X}}_1'\tilde{\mathcal{K}}_1(y')\tilde{\mathcal{X}}_1)^{-1}\tilde{\mathcal{X}}_1'\tilde{\mathcal{K}}_1(y')\tilde{\mathcal{Y}}_1$$

The estimate of $\mathrm{F}_{Y_{1t}|Y_{0t-1},D_t=1}(y_1|y')$ is $\hat{\alpha}_1(y')$.

The advantage of using a local linear kernel estimator as opposed to a local constant estimator is that the bias of the local linear linear estimator is the same near the boundary of the support of Y_{0t-1} as it is in the interior of the support – this is not the case for the local constant estimator. However, using a local linear estimator introduces one additional issue: the estimate of $\mathrm{F}_{Y_{1t}|Y_{0t-1},D_t=1}(y_1|y')$ may be less than 0 or greater than 1 because the local weights can be negative (Hall, Wolff, and Yao, 1999). To alleviate this problem, I adopt the approach of Hansen (2004) and set negative weights to be equal to 0. This approach has an asymptotically negligible effect.

Step 2: Estimate $\mathrm{F}_{Y_{0t}|Y_{0t-1},D_t=1}(y_1|y')$*:*

The first requirement for estimating $\mathrm{F}_{Y_{0t}|Y_{0t-1},D_t=1}(y_0|y')$ is to write it in terms of objects that are observed and therefore estimable. The following lemma provides an estimable version of this conditional distribution.

LEMMA 24.

$$\mathrm{F}_{Y_{0t}|Y_{0t-1},D_t=1}(y_0|y') = \mathrm{F}_{Y_{0t-1}|Y_{0t-2},D_t=1}\left(\mathrm{F}^{-1}_{Y_{0t-1}|D_t=1}(\mathrm{F}_{Y_{0t}|D_t=1}(y_0))\middle|\mathrm{F}^{-1}_{Y_{0t-2}|D_t=1}(\mathrm{F}_{Y_{0t-1}|D_t=1}(y'))\right)$$

With Lemma 24 in hand, it is fairly straightforward to estimate $\mathrm{F}_{Y_{0t}|Y_{0t-1},D_t=1}(y_0|y')$. Once again, I use local linear kernel estimators. Compared to Step 1, the only additional issue here is that I need first step estimators of the distribution and quantile functions in the result of Lemma 24. To estimate distribution functions, I use empirical cdfs:

$$\hat{\mathrm{F}}_Z(z) = \frac{1}{n}\sum_{i=1}^{n} 1\{Z_i \leq z\}$$

To estimate quantile functions, I invert empirical cdfs:

$$\hat{\mathrm{F}}_Z^{-1}(\tau) = \inf\{z : \hat{\mathrm{F}}_Z(z) \geq \tau\}$$

With estimates of the distribution functions in hand, Lemma 24 can be estimated by the solution $\hat{\alpha}_0(y_0, y')$ to

$$\min_{\alpha_0,\beta_0} \sum_{i\in\mathcal{T}} \left(1\{Y_{it-1} \leq \hat{\mathrm{F}}^{-1}_{Y_{0t-1}|D_t=1}(\hat{\mathrm{F}}_{Y_{0t}|D_t=1}(y_0))\} - \alpha_0 - \left(Y_{it-2} - \hat{\mathrm{F}}^{-1}_{Y_{0t-2}|D_t=1}(\hat{\mathrm{F}}_{Y_{0t-1}}(y'))\right)\beta_0\right)$$
$$\times K_h\left(Y_{it-2} - \hat{\mathrm{F}}^{-1}_{Y_{0t-2}|D_t=1}(\hat{\mathrm{F}}_{Y_{0t-1}}(y'))\right)$$

Likewise, there is a closed form solution to this problem. Let $\gamma_0(y_0|y') = [\alpha_0(y_0|y'), \beta_0(y_0|y')]'$, $\hat{\mathcal{Z}}$ be an $n_T \times 1$ vector with ith component $1\{Y_{it-1} \leq \hat{\mathrm{F}}^{-1}_{Y_{0t-1}|D_t=1}(\hat{\mathrm{F}}_{Y_{0t}|D_t=1}(y_0))\}$, $\hat{\mathcal{X}}$ be an $n_T \times 2$ matrix with ith row given by $(1, Y_{it-2} - \hat{\mathrm{F}}^{-1}_{Y_{0t-2}|D_t=1}(\hat{\mathrm{F}}_{Y_{0t-1}|D_t=1}(y')))$, and $\hat{\mathcal{K}}(y')$ be an $n_T \times n_T$ diagonal matrix with ith diagonal element given by $K_h(Y_{it-2} - \hat{\mathrm{F}}^{-1}_{Y_{0t-2}|D_t=1}(\hat{\mathrm{F}}_{Y_{0t-1}|D_t=1}(y')))$. Then, a closed form expression for the estimate of $\gamma_0(y_0|y')$ is

$$\hat{\gamma}_0(y') = (\hat{\mathcal{X}}'\hat{\mathcal{K}}(y')\hat{\mathcal{X}})^{-1}\hat{\mathcal{X}}'\hat{\mathcal{K}}(y')\hat{\mathcal{Z}}$$

Step 3: Compute the Bounds on the Distribution of the Treatment Effect

To obtain the bounds on the distribution of the treatment effect, one can plug in the above estimates into the results of Lemma 23 and Theorem 9. Recall, the lower bound on the DTE is identified and given by

$$\mathrm{F}^L_{Y_{1t}-Y_{0t}|D_t=1}(\Delta) = \mathrm{E}[\mathrm{F}^L_{Y_{1t}-Y_{0t}|Y_{0t-1},D_t=1}(\Delta|Y_{0t-1})]|D_t=1] \tag{2.2}$$

Further, recall that

$$\mathrm{F}^L_{Y_{1t}-Y_{0t}|Y_{0t-1},D_t=1}(\Delta|y') = \sup_y \max\{\mathrm{F}_{Y_{1t}|Y_{0t-1},D_t=1}(y|y') - \mathrm{F}_{Y_{0t}|Y_{0t-1},D_t=1}(y-\Delta|y'), 0\}$$

which can be estimated by

$$\hat{F}^{L}_{Y_{1t}-Y_{0t}|Y_{0t-1},D_t=1}(\Delta|y') = \sup_{y} \max\{\hat{F}_{Y_{1t}|Y_{0t-1},D_t=1}(y|y') - \hat{F}_{Y_{0t}|Y_{0t-1},D_t=1}(y-\Delta|y'), 0\}$$

Then, an estimate of Equation 2.2 is given by

$$\hat{F}^{L}_{Y_{1t}-Y_{0t}|D_t=1}(\Delta) = \frac{1}{n_T}\sum_{i\in\mathcal{T}} \hat{F}^{L}_{Y_{1t}-Y_{0t}|Y_{0t-1},D_t=1}(\Delta|Y_{it-1})$$

Similarly, the upper bound on the DTE is given by

$$F^{U}_{Y_{1t}-Y_{0t}|D_t=1}(\Delta) = \mathrm{E}[F^{U}_{Y_{1t}-Y_{0t}|Y_{0t-1},D_t=1}(\Delta|Y_{0t-1})]|D_t = 1] \tag{2.3}$$

where

$$F^{U}_{Y_{1t}-Y_{0t}|Y_{0t-1},D_t=1}(\Delta|y') = 1 + \inf_{y} \min\{F_{Y_{1t}|Y_{0t-1},D_t=1}(y|y') - F_{Y_{0t}|Y_{0t-1},D_t=1}(y-\Delta|y'), 0\}$$

which can be estimated by

$$\hat{F}^{U}_{Y_{1t}-Y_{0t}|Y_{0t-1},D_t=1}(\Delta|y') = 1 + \inf_{y} \min\{\hat{F}_{Y_{1t}|Y_{0t-1},D_t=1}(y|y') - \hat{F}_{Y_{0t}|Y_{0t-1},D_t=1}(y-\Delta|y'), 0\}$$

and implies that an estimate of Equation 2.3 is given by

$$\hat{F}^{U}_{Y_{1t}-Y_{0t}|D_t=1}(\Delta) = \frac{1}{n_T}\sum_{i\in\mathcal{T}} \hat{F}^{U}_{Y_{1t}-Y_{0t}|Y_{0t-1},D_t=1}(\Delta|Y_{it-1})$$

Step 4: Estimate the Bounds on the QoTET

The upper bound on the QoTET is given by inverting the lower bound on the DTE, and the lower bound on the QoTET is given by inverting the upper bound on the DTE. Therefore,

$$\widehat{\text{QoTET}}^U(\tau) = \inf\{\Delta : \hat{\text{F}}^L_{Y_{1t}-Y_{0t}|D_t=1}(\Delta) \geq \tau\}$$

and

$$\widehat{\text{QoTET}}^L(\tau) = \inf\{\Delta : \hat{\text{F}}^U_{Y_{1t}-Y_{0t}|D_t=1}(\Delta) \geq \tau\}$$

Estimating the ATT-CPO

Recall that the ATT-CPO is given by

$$\begin{aligned} \text{ATT-CPO}(y') &= \text{E}[Y_{1t} - Y_{0t}|Y_{0t-1} = y', D_t = 1] \\ &= \text{E}[Y_{1t}|Y_{0t-1} = y', D_t = 1] \qquad (2.4) \\ &\quad - \text{E}[\text{F}^{-1}_{Y_{0t}|D_t=1}(\text{F}_{Y_{0t-1}|D_t=1}(Y_{0t-1})) \mid Y_{0t-2} = \text{F}^{-1}_{Y_{0t-2}|D_t=1}(\text{F}_{Y_{0t-1}|D_t=1}(y')), D_t = 1] \qquad (2.5) \end{aligned}$$

I estimate the ATT-CPO using local linear kernel estimators. An estimate of the term in Equation 2.4 comes from solving

$$\min_{a_1,b_1} \sum_{i \in \mathcal{T}} [Y_{it} - a_1 - (Y_{it-1} - y')b_1]^2 K_h(Y_{it-1} - y')$$

Just like for the QoTET, this is easy to estimate as it is simply weighted least squares. Let $\delta_1(y') = [a_1(y'), b_1(y')]'$, $\mathcal{Y}_1$ be an $n_T \times 1$ vector with ith component Y_{it}, $\mathcal{X}_1$ be an $n_T \times 2$ matrix with ith row given by $(1, Y_{it-1} - y')$, and $\mathcal{K}_1(y')$ be an $n_T \times n_T$ diagonal matrix with ith diagonal element given by $K_h(Y_{it-1} - y')$. Then, a closed form expression for the estimate of $\delta_1(y')$ is

$$\hat{\delta}_1(y') = (\mathcal{X}_1'\mathcal{K}_1(y')\mathcal{X}_1)^{-1}\mathcal{X}_1'\mathcal{K}_1(y')\mathcal{Y}_1$$

The estimate of ATT-CPO(y') is $\hat{a}_1(y')$.

Estimating the term in Equation 2.5 is more complicated because it depends on distribution functions and quantile functions. These need to be estimated in a first step. With estimates of the distribution functions in hand, Equation 2.4 can be estimated by the solution $\hat{a}_0(y')$ to

$$\min_{a_0,b_0}\sum_{i\in\mathcal{T}}\left(\hat{\mathrm{F}}^{-1}_{Y_{0t}|D_t=1}(\hat{\mathrm{F}}_{Y_{0t-1}|D_t=1}(Y_{it-1})) - a_0 - \left(Y_{it-2} - \hat{\mathrm{F}}^{-1}_{Y_{0t-2}|D_t=1}(\hat{\mathrm{F}}_{Y_{0t-1}|D_t=1}(y'))\right)b_0\right)$$
$$\times K_h\left(Y_{it-2} - \hat{\mathrm{F}}^{-1}_{Y_{0t-2}|D_t=1}(\hat{\mathrm{F}}_{Y_{0t-1}|D_t=1}(y'))\right)$$

Likewise, there is a closed form solution to this problem. Let $\delta_0(y') = [a_0(y'), b_0(y')]'$, $\hat{\mathcal{Z}}_0$ be an $n_T \times 1$ vector with ith component $\hat{\mathrm{F}}^{-1}_{Y_{0t}|D_t=1}(\hat{\mathrm{F}}_{Y_{0t-1}|D_t=1}(Y_{it-1}))$, $\hat{\mathcal{X}}_0$ be an $n_T \times 2$ matrix with ith row given by $(1, Y_{it-2} - \hat{\mathrm{F}}^{-1}_{Y_{0t-2}|D_t=1}(\hat{\mathrm{F}}_{Y_{0t-1}|D_t=1}(y')))$, and $\hat{\mathcal{K}}_0(y')$ be an $n_T \times n_T$ diagonal matrix with ith diagonal element given by $K_h(Y_{it-2} - \hat{\mathrm{F}}^{-1}_{Y_{0t-2}|D_t=1}(\hat{\mathrm{F}}_{Y_{0t-1}|D_t=1}(Y_{it-2})))$. Then, a closed form expression for the estimate of $\delta_0(y')$ is

$$\hat{\delta}_0(y') = (\hat{\mathcal{X}}_0'\hat{\mathcal{K}}_0(y')\hat{\mathcal{X}}_0)^{-1}\hat{\mathcal{X}}_0'\hat{\mathcal{K}}_0(y')\hat{\mathcal{Z}}_0$$

2.4.1 Inference

ATT-CPO

This section shows that the estimate of the ATT-CPO is consistent and asymptotically normal. Its rate of convergence is slower than $\sqrt{n}$ when the ATT-CPO is estimated nonparametrically. Following standard arguments on local linear regression (see, for example, Fan and Gijbels, 1996; Li and Racine, 2007), one can show that its estimate is consistent and asymptotically normal.

The only complication is that estimation depends on first step estimates of several distribution and quantile functions. The intuition for the following result is that the first step estimation of the distributions and quantile functions does not matter asymptotically because each of these can be estimated at the parametric $\sqrt{n}$ rate, but the final nonparametric rate converges at the slower rate $\sqrt{nh}$.

Let $g_1(y') = \mathrm{E}[Y_{1t}|Y_{0t-1} = y', D_t = 1]$, $\varepsilon_{it} = Y_{1it} - g_y(y')$, $\sigma^2_\varepsilon(y') = \mathrm{E}[\varepsilon^2_{it}|Y_{0t-1} = y', D_t = 1]$, and h_1 be a bandwidth parameter. Also, let $z = \mathrm{F}^{-1}_{Y_{0t-2}|D_t=1}(\mathrm{F}_{Y_{0t-1}|D_t=1}(y'))$, $g(z) = \mathrm{E}[\mathrm{F}^{-1}_{Y_{0t}|D_t=1}(\mathrm{F}_{Y_{0t-1}|D_t=1}(Y_{0t-1})) \bigm| Y_{0t-2} = z, D_t = 1]$,
$u_{it} = \mathrm{F}^{-1}_{Y_{0t}|D_t=1}(\mathrm{F}_{Y_{0t-1}|D_t=1}(Y_{0it-1})) - g(\mathrm{F}^{-1}_{Y_{0t-2}|D_t=1}(\mathrm{F}_{Y_{0t-1}|D_t=1}(Y_{0it-2})))$, $\sigma^2_u(z) = \mathrm{E}[u^2_{it} \bigm| Y_{0t-2} = z, D_t = 1]$, and h be a bandwidth parameter. Finally, let $k(\cdot)$ be a kernel function, $\kappa = \int k(v)^2\, \mathrm{d}v$, and $\kappa_2 = \int k(v)v^2\, \mathrm{d}v$. I make the following assumptions

ASSUMPTION 14.

(a) $g(z)$, $f_{Y_{0t-2}|D_t=1}(z)$, *and* $\sigma^2_u(z)$ *are twice continuously differentiable*

(b) $g_1(y')$, $f_{Y_{0t-1}|D_t=1}(y')$, *and* $\sigma^2_\varepsilon(y')$ *are twice continuously differentiable*

(c) $k(\cdot)$ *is a bounded second order kernel*

(d) As $n \to \infty$, $nh \to \infty$, $nh_1 \to \infty$, $nh^7 \to 0$, *and* $nh_1^7 \to 0$

(e) Y_{0t}, Y_{0t-1}, *and* Y_{0t-2} *have a common, compact support* $\mathcal{Y}$

(f) $f_{Y_{0t-1}|D_t=1}(\cdot)$ *and* $f_{Y_{0t-2}|D_t=1}(\cdot)$ *are bounded away from 0 on* $\mathcal{Y}$.

Under these assumptions, the following result holds,

THEOREM 12. *(Asymptotic Normality of ATT-CPO)*

$$\sqrt{nh}\{(\hat{a}_1(y') - \hat{a}_0(y')) - \mathit{ATT\text{-}CPO}(y') - \mathit{Bias}(y')\} \xrightarrow{d} N(0, V) \tag{2.6}$$

where

$$\text{Bias}(y') = \frac{\kappa_2}{2}\left(f_{Y_{0t-1}|D_t=1}(y')g_1''(z)h_1^2 - f_{Y_{0t-2}|D_t=1}(z)g''(z)h^2\right)$$

and

$$V = \frac{\sigma_\varepsilon^2(y')}{f_{Y_{0t-1}|D_t=1}(y')}\kappa + \frac{\sigma_u^2(\mathrm{F}^{-1}_{Y_{0t}|D_t=1}(\mathrm{F}_{Y_{0t-1}|D_t=1}(y')))}{f_{Y_{0t-2}|D_t=1}(\mathrm{F}^{-1}_{Y_{0t-2}|D_t=1}(\mathrm{F}_{Y_{0t-1}|D_t=1}(y')))}\kappa$$

2.5 Job Displacement of Older Workers during the Great Recession

This section studies the effect of job displacement during the Great Recession on yearly earnings of older workers. Using standard techniques, job displacement is estimated to decrease workers earnings by 40% relative to counterfactual earnings had they not been displaced. The size of this effect is 0-25% larger than existing estimates of the effect of job displacement for all workers during severe recessions. The size of the effect is also consistent with the ideas that (i) the effect of job displacement is larger for older workers than prime age workers and (ii) the effect of job displacement is larger during recessions.

Next, this section considers the distributional impacts of job displacement using the techniques developed in the chapter. There are two key findings that would not be available without these methods. First, using the panel data methods developed in the chapter provides substantially more identifying power for distributional treatment effects such as the QoTET than is available using existing bounds that do not exploit panel data. The reason the bounds are tighter is that relatively strong positive dependence is observed in non-displaced earnings over time for the displaced group of workers in the period before the Great Recession. These bounds are tight enough to rule out the assumption of perfect positive dependence between displaced and non-displaced potential earnings. This result implies that there is more heterogeneity (and potentially much more heterogeneity) in the effect of job displacement than would be implied by the estimate of the QoTET under the assumption of perfect positive dependence. The bounds also imply that some workers have higher earnings after being displaced than they would have had they not been displaced; this implies that the assumption of Monotone Treatment Response (Manski, 1997) is rejected. Second, workers with higher earnings before the recession experience larger decreases in the level of earnings following displacement, but as a fraction of earnings, the earnings loss is almost constant across workers with differing earnings prior to the recession.

These results can be compared to existing empirical work on job displacement. Broadly speaking, there are two key findings from the job displacement literature: (i) the effect of job displacement on earnings is large, and (ii) the effect of job displacement is persistent. The current chapter considers the effect of job displacement on earnings 2-4 years following displacement which is a somewhat shorter period than most existing work. The empirical literature on job displacement finds that workers suffer large earnings losses upon job displacement. To give some examples, Jacobson, LaLonde, and Sullivan (1993) study the effect of job displacement during a deep recession – the recession in the early 1980s. That paper finds that workers lose 40% of their earnings upon displacement and still have 25% lower earnings six years following displacement. Interestingly, it finds little difference in the path of earnings for older, prime-age, and younger workers. Couch and Placzek (2010) study job displacement in the smaller recession in the early 2000s. They find an initial 32% decrease in earnings following displacement, but earnings are only 13% lower six years after displacement. Using Social Security data that covers the entire U.S., Von Wachter, Song, and Manchester (2009) also study the effect of displacement during the early 1980s and find a 30% reduction in earnings upon displacement and earnings still 20% lower up to twenty years following displacement. Because they have data on the entire country, they can compare the effect of displacement including and excluding observations with 0 earnings in a particular year. Not surprisingly, the effect of job displacement is larger when 0 earnings are included, but the path of earnings is very similar – a large dip followed by some recovery but never complete recovery. That paper also finds older workers experience a 21% larger decrease in earnings than prime age workers following displacement though the results are more similar when only workers with non-negative earnings are included. Stevens (1997), using PSID data, finds that workers initially lose 25% of their earnings following job displacement and 9% lower earnings ten years later. Using the Displaced Worker Survey, Farber (1997) finds that displaced workers lose 12% of weekly earnings on average following displacement, but that the effect is much larger for workers age 55-64. The effect of job displacement on earnings is larger when there are weak labor market conditions relative to strong labor market conditions (Farber, 1997; Davis and Von Wachter, 2011) which is relevant for older workers displaced during the Great Recession.

There are three potentially important sources of bias in most of the work on job displacement. First, most work limits the sample to individuals who have positive earnings in each period. This may be important because difficulty finding new employment is likely to be a consequence of job displacement. In studies that use state-level administrative data such as Jacobson, LaLonde, and Sullivan (1993) and Couch and Placzek (2010), this choice is made because they are unable to tell whether 0 earnings represents unemployment, leaving the labor force, or moving to another state. Under the condition that more productive displaced workers are more likely to return to work than less productive workers, dropping individuals with no earnings is likely to cause the estimated effect of job displacement to be biased towards 0.

A second well known potential problem is that employers may selectively lay off their least productive workers during recessions (Gibbons and Katz, 1991). If this is the case, then comparing these workers to workers that are not displaced may tend to overestimate the effect of job displacement if these workers earnings would not have increased as much as non-displaced workers in the absence of job displacement.

Finally, analyzing the effect of job displacement on older workers is more challenging than for prime age workers because older workers may also retire following job displacment. Unlike the previous cases, which both clearly suggest the sign of the direction of resulting bias, it is not clear whether more or less productive workers are more likely to retire. On the one hand, more productive workers may face better labor market opportunities which may make them less likely to retire. On the other hand, more productive workers may have accumulated more retirement savings which may make them more likely to retire.

The effect of job displacement may be particularly severe for workers displaced during the Great Recession because of the particularly weak labor market conditions in the period immediately following the recession. From the official beginning of the recession in December 2007 to October 2009, four months after the official end of the recession, the unemployment rate doubled from 5.0% to 10.0% (U.S. Bureau of Labor Statistics, 2015b). And during the same period, the economy shed almost 8.4 million jobs (U.S. Bureau of Labor Statistics, 2015a). For workers ages 55 and over, the unemployment rate more than doubled from 3.2% to 6.9% (U.S. Bureau of Labor Statistics, 2015c).

There is recent work on the effect of job displacement during the Great Recession using the Displaced Workers Survey Farber (2015). I summarize some of the relevant results next. For all workers, the incidence of job loss was at its highest during the Great Recession compared to all other periods covered by the DWS (1981-present). Roughly, one in six workers report having lost a job. Compared to previous time periods, the rate of reemployment is very low with more workers being reemployed in part time jobs. For older workers, the job loss rate is slightly lower than for the population at large, but the difference is not as large as it was in earlier time periods. Historically, following displacement, older workers have been about equally likely to be unemployed and leave the labor force (around 25% with some variation over time). During the recession however, the unemployment rate jumped substantially relative to leaving the labor force (unemployment went to 40% while leaving labor force dropped slightly to 20%) (Farber, 2015). Interestingly, Farber (2015) finds that, for workers who find full time jobs following job displacement, the effect on earnings during the Great Recession has not been unusually large compared to other periods – about 12%. This provides some evidence that the main channel for a differential effect of job displacement during the Great Recession relative to other periods comes from either failure to find a new job or moving from a full time job to a part time job.

2.5.1 Data

The data comes from the Health and Retirement Study (HRS) (Health and Retirement Study, 2014). The HRS is a panel data set with interviews occurring every two years. The study follows participants ages 50 and older. Since its inception in 1992, the panel has added new participants who meet the age requirement six times, most recently adding the Early Baby Boomers (birth years 1948-1953) in 2004 and the Mid Baby Boomers (birth years 1954-1959) in 2010. The primary data source for the current chapter is the RAND HRS Version N longitudinal files which I also supplement with some of the base data. I use data from 2004, 2006, 2008, 2010, and 2012 which primarily covers the Early Baby Boomers who range in age from 50-56 in 2004 to 58-64 in 2012. In the context, of job displacement this dataset has been considered by Couch (1998), Chan and Stevens (1999), Chan and Stevens (2001), and Chan and Stevens (2004).

The RAND HRS data file contains 37,319 individual-level observations though many individuals are no longer in the dataset in the time frame being considered. Each of the yearly data files from 2004-2012 contains between 17,000 and just over 22,000 individuals. Merging all of these dataset leaves 12,984 individual observations. For this subset, the average age is 73 in 2012 implying that many of these individuals are not working at any point in the period of interest. I further subset the data to those who are coded as "Working for Pay" in 2006 leaving 5429 observations. Some observations have missing earnings data and others have imputed earnings data; I drop all of these observations. Following the job displacement literature, I also drop observations with 0 earnings in any periods leaving a final sample size of 1473 indviduals.

The HRS asks workers who are not employed at the same employer as in the previous survey the reason why they left their employer. Following the job displacement literature, I code individuals as being displaced if the reason they are not at the same employer is that (i) the business closed or (ii) they were laid off or let go. The latter category includes temporary workers, contract workers, layoffs from lack of work, downsizing, reorganization, change of political administration, and employer sickness or death. Other important causes of leaving a job that are not counted as being displaced are (i) poor health or disability, (ii) family care, (iii) better job, (iv) quit, (v) retired, and (vi) moving. I form the displaced group by counting individuals who were displaced in either 2008 or 2010. Using this definition, there are 160 individuals that are displaced which amounts to 10.9% of the sample.

Summary statistics are presented in Table 2.1. The outcome of interest is yearly earnings. Prior to the recession, average earnings levels are higher for non-displaced workers than displaced workers ($48,000 vs. $43,200). For non-displaced workers, average earnings levels remain essentially flat – in 2012, non-displaced workers earn $51,400 per year on average. For displaced workers, earnings fall dramatically following displacement. In 2010, average earnings for displaced workers are only $30,100 (30% lower than 2006 earnings); by 2012, average earnings have increased somewhat to $35,000 though this is still much lower than pre-displacement earnings.

There are only small differences in observable characteristics that may explain the differences in observed earnings for the displaced and non-displaced groups. 39% of non-displaced workers have a college degree compared to 33% of displaced workers. 85% of non-displaced workers are white compared to 81% of displaced workers. And, for both groups, 44% are male.

Larger differences can be seen with respect to labor force status. In 2006, 77% of non-displaced workers are employed full time compared to 79% of displaced workers. But by 2010, there are sharp differences. 68% of non-displaced workers are employed full time in 2010, but only 49% of displaced workers are employed full time. For non-displaced workers, the unemployment rate is 1.1% in 2010, but it is 16% for displaced workers in 2010. 2.1% of non-displaced workers are retired in 2010 compared to 8.7% of displaced workers. These differences in full time employment, unemployment rates, and retirement rates narrow somewhat by 2012 possibly accounting for the smaller earnings gap between displaced and non-displaced workers in 2012.

2.5.2 Baseline Results

In this section, I estimate the average effect of job displacement on the earnings of older workers. The results indicate that older workers lose 40% of their earnings due to job displacement. This effect is the same or somewhat larger in magnitude compared to estimates of the effect on prime age workers during the deep recession in the early 1980s (Jacobson, LaLonde, and Sullivan, 1993; Von Wachter, Song, and Manchester, 2009) which are the largest in the literature. This estimate is almost four times as large as the estimated effect of job displacement on all workers during the Great Recession (Farber, 2015).[9]

Let Y_{it} denote earnings for individual i in year t. Following the most common specifications in the literature, estimate the following model for individuals that have non-zero earnings in each period

$$\log(Y_{it}) = c_i + \gamma_t + \alpha D_{it} + X_{it}'\beta + \varepsilon_{it} \tag{2.7}$$

where c_i is an individual-specific fixed effect, γ_t is an aggregate time fixed effect, D_{it} is a binary variable indicating whether or not an individual is displaced, X_{it} is a vector of covariates, and ε_{it} is an error term. The coefficient of interest is α. I estimate the model

[9]The actual difference is probably not as large because the results in Farber (2015) come from workers who worked full time before and after job displacement. This is likely to be important empirically as only 50% of workers are reemployed at the time of their interview and 20% more are employed part time. The estimates in the current chapter would not include those that are not reemployed, but it would include those that are employed part time who, by construction, will tend to have lower earnings. Moreover, estimates from the DWS have tended to produce lower estimated effects of job displacement on earnings than estimates from other sources.

using data on earnings from 2006 and 2012 using a Correlated Random Effects approach.[10]

Table 2.2 provides the results for the correlated random effects model. When I only include a year fixed effect in addition to the displacement indicator (Model 1), earnings are estimated to be 41%[11] lower on average for displaced workers than non-displaced workers. Model 2 adds demographic, education, and location characteristics, and the estimated effect is very similar. The third model adds fixed effects for initial occupation and initial industry. The estimated effect increases to a 49% reduction in earnings. Model 4 includes time varying occupations as a covariate which allows the effect of job displacement to depend on the occupation of individuals following job displacement.[12] Conditioning on current occupation eliminates one of the channels through which job displacement may work – moving to lower paying occupations. Even when this channel is removed, job displacement is estimated to decrease earnings by 34%.

The last three models in Table 2.2 consider the younger subset of workers that are 64 or younger in 2012. The estimated effects are very similar for this group. When demographic characteristics (Model 5) and industry and occupation fixed effects are added (Model 6), job displacement is estimated to decrease earnings by 38% and 40%, respectively. Finally, when time varying occupations are included, the estimated effect is somewhat smaller at 28% and only borderline statistically significant (p-value=0.07).

A weakness of the previous specifications is that the aggregate time effect, γ_t, is commmon to all individuals. This means that the time trend is constrained by the model to be the same for individuals that may have very different observed and unobserved characteristics. This could potentially cause the effect of job displacement to be overestimated. For example, less educated workers may be more likely to be displaced than highly educated

[10]With two periods, estimates for variables that change over time are numerically identical to estimate from a Fixed Effects approach and also allow me to obtain estimates of the effects of variables that do not change over time though these estimates should be interpreted more cautiously. The original work on correlated random effects models is Mundlak (1978), Chamberlain (1982), and Chamberlain (1984)

[11]Estimates of the effect of job displacement on earnings as a percentage of earnings, which is what is reported in the text, are given by $\exp(\hat{\alpha}) - 1$ where $\hat{\alpha}$ is the estimated coefficient in the table.

[12]The reason I do not include time varying industry the sample sizes become too small due to missing values for industry in the latter period.

workers. Earnings are also likely to be increasing more over time for highly educated workers than for less educated workers. Since the trend in earnings for highly educated workers will be used to pin down γ_t in these specifications, the trend is likely to be overestimated for less educated workers; therefore, the size of the effect of job displacement would also be overestimated in this situation.

One could potentially mitigate this problem by interacting time fixed effects with observable characteristics and some variation of this approach is used in Jacobson, LaLonde, and Sullivan (1993) and Von Wachter, Song, and Manchester (2009). In Von Wachter, Song, and Manchester (2009), for example, interacting time fixed effects with industry dummy variables tends to somewhat mitigate the estimated effect of job displacement.

Instead of interacting observables and time fixed effects, I consider a nonparametric identifiying assumption consistent with the idea that individuals that differ in observable characteristics may have differing time trends in untreated outcomes

ASSUMPTION 15. *(Distributional Difference in Differences)*

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

This is a Difference in Differences assumption that says that the path of untreated potential outcomes for the treated group and for the control group is the same for workers that have the same observable characteristics. In other words, for observations with the same covariates, the distribution of the path of untreated outcomes is the same for displaced and non-displaced individuals. This assumption could be weakened to mean independence to identify the ATT, but it is useful for identifying the QTT (Callaway and Li, 2015) which will be required in the next section.[13] This assumption is consistent with a model such as

$$Y_{1t} = g_{1t}(X, c_i, \varepsilon_{it})$$

[13] In the job displacement literature, a very similar approach based on Difference in Differences with matching on the propensity score (Heckman, Ichimura, and Todd, 1997) is considered by Couch and Placzek (2010).

$$Y_{0t} = g_{0t}(X) + h_0(X, c_i, \varepsilon_{it})$$

where the model for treated potential outcomes is left completely unrestricted and the model for untreated potential oucomes depends on observable characteristics X, time invariant unobservables c_i that may be correlated with the covariates X, and time varying unobservables that are independent of (X, c_i) but may be serially correlated over time. The restriction on the model for untreated potential outcomes is that the path of untreated outcomes depends only on observable characteristics X but not on any unobservables. Assumption 15 identifies only the ATT and QTT but not the ATE or the QTE. Under this assumption,

$$ATT = \mathrm{E}\left[\frac{(D - p(X))\Delta Y_t}{p(1 - p(X))}\right]$$

where p is the unconditional probability that an individual is displaced and $p(x) = \mathrm{P}(D_t = 1|X = x)$ is the propensity score – the probability that an individual is displaced conditional on observed characteristics.

Using this method, the estimated effect of job displacement on older workers is 38% lower earnings on average. This estimate is only slightly smaller in magnitude than the estimates coming from the parametric model. This result provides some evidence that differences in trends for individuals with different covariates do not greatly affect the results.[14]

2.5.3 The Distributional Effects of Job Displacement

This section uses the techniques developed earlier in the chapter to understand the distributional effects of job displacement for older workers. I focus on estimating the QoTET and the ATT-CPO. The QoTET is useful for understanding heterogeneity in the effect of job displacement. It is partially identified in the current application. The ATT-CPO is point

[14] Couch and Placzek (2010) also find very similar results whether using a parametric model which is very similar to the current parametric model or a very similar semiparametric Difference in Differences technique.

identified and is useful for determining whether job displacement has had a relatively larger impact on workers who previously had high or low earnings.

There are three main findings using the methods developed in the current chapter. First, the bounds developed in the current chapter provide substantially more identifying power than bounds relying on only knowledge of the marginal distributions of potential outcomes. Those bounds are essentially uninformative in the current application. Second, the tighter bounds in the current chapter rule out: (i) the assumption of perfect positive dependence across treated and untreated potential outcomes and (ii) the assumption of Monotone Treatment Response (Manski, 1997). Finally, estimates of the ATT-CPO (See Figure 2.5) imply that workers that had higher earnings in the previous period experience larger earnings losses due to job displacement than workers with lower earnings in the previous period. But, as a fraction of earnings, the effect of job displacement is very similar for workers across all initial earnings levels.

Recall that the three key requirements to estimate these distributional treatment effect parameters are (i) access to panel data, (ii) identification of the counterfactual distribution of potential outcomes, and (iii) the Copula Stability Assumption. Thus, as a first step, I need to estimate the counterfactual distribution of potential outcomes – this is what I do next.

Step 1: Estimate the counterfactual distribution of untreated potential outcomes for the treated group

The first task to be accomplished is to estimate the counterfactual distribution of untreated potential outcomes for the treated group – in other words, the unobserved distribution of earnings for the group of displaced workers if they had not been displaced. Knowledge of this distribution, in combination with the distribution of treated potential outcomes for the treated group (which is observed), identifies the QTT. I use the Difference in Differences method of Callaway and Li (2015) though there are a variety of methods that could be used to estimate this counterfactual distribution.[15]

The key identifying assumption of Callaway and Li (2015) is Assumption 15. That paper also imposes a Copula Stability Assumption that is similar to the one in the current chapter though not exactly the same. However, that Copula Stability Assumption is also satisfied in the same types of models that satisfy the Copula Stability Assumption in the current chapter which suggests that this extra condition is not likely to add much empirical content. The counterfactual distribution of potential outcomes is point identified in this setup. Estimates of both marginal distributions are presented in Figure 2.2, and an estimate of the QTT is presented in Figure 2.3.

Step 2: Estimate parameters that depend on the joint distribution of treated and untreated potential outcomes

Next, I use the techniques presented earlier in the chapter to estimate some parameters that depend on the joint distribution of treated and untreated potential outcomes. First, I consider the QoTET. Figure 2.4 plots (i) bounds on the QoTET under no assumptions on the dependence between potential outcome distributions, (ii) the bounds developed in the current chapter, (iii) point estimates of the QoTET under the assumption of perfect positive dependence between treated and untreated potential outcomes, and (iv) point estimates of the QoTET under the assumption of Rank Invariance between untreated potential outcomes over time. There are several things to notice from the figure. First, when no information besides the identified marginal distributions is used, the bounds on the QoTET are very wide. For example, the median of the treatment effect is bounded to be between an earnings

[15] One idea would be to use the Change in Changes model (Athey and Imbens, 2006). Melly and Santangelo (2015) have recently extended this model to allow conditioning on covariates and this approach could be adapted to the current application. Another idea would be to impose selection on observables with a lag of earnings being a conditioning variable and use the method of Firpo (2007). The results are not sensitive to using the selection on observables method of Firpo (2007). I have not implemented the Athey and Imbens (2006) and Melly and Santangelo (2015) method to compare results.

losses of 79% and an earnings gain of 118%. The effect of displacement among those most affected by displacement, for example the 5th percentile of the QoTET is estimated to be between a 62% and 99% loss of earnings. The effect of displacement on those least affected by displacement, for example the 95th percentile of the QoTET is bounded between a 21% loss of earnings and a 1163% increase in earnings. From these bounds, one is not able to determine much. These bounds indicate that at least 19% of displaced workers have lower earnings than they would have had they not been displaced.

Next, under the Copula Stability Assumption, the bounds are indeed tighter. Earnings losses at the 5th percentile are between 90% and 99% which implies that some individuals lose almost all of their earnings due to displacement. The estimates of the QoTET also imply that at least 43% of individuals are worse off from being displaced. Interestingly, one can also conclude that at least 13% of individuals have higher earnings after being displaced than they would have had they not been displaced. This type of conclusion was not available without exploiting the Copula Stability Assumption and would imply that the assumption of Monotone Treatment Response (Manski, 1997) is not valid in the current case.

Figure 2.4 also plots the QoTET under several assumptions that would lead to point identification. First, it plots the QoTET under perfect positive dependence between Y_{1t} and Y_{0t}. I have argued that this is an especially strong assumption in this case. For example, it essentially restricts any previously high earnings individuals from moving into much lower paying positions following displacement. This identifying assumption implies the least amount of heterogeneity in the effect of being displaced. At the 5th percentile, individuals lose 82% from being displaced. At the 95th percentile, they lose 16%. At the median, they lose 32%, and this effect is largely constant across most of the interior quantiles. Of course, the no-assumptions bounds cannot rule out perfect positive dependence between Y_{1t} and Y_{0t}, but under the Copula Stability Assumption, perfect positive dependence is ruled out because the bounds imply more heterogeneity than occurs under perfect positive dependence.

Finally, I also plot the results for the case with perfect positive dependence between Y_{0t} and Y_{0t-1}. This assumption results in considerably more heterogeneity in the effect of job displacement than the assumption of perfect positive dependence between Y_{1t} and Y_{0t}. For example, at the 5th percentile, the estimated effect of job displacement is a loss of 97% of earnings. At the median, the estimated effect is 22% lower earnings per year. And at the 95th percentile, earnings are estimated to be 144% higher than they would have been without job displacement. Further, 65% of individuals are estimated to be worse off from job displacement and 35% are estimated to have higher earnings than they would have had they not been displaced. The reason that the bounds in the current chapter are close to the estimates of the QoTET under this point identifying assumption is that strong positive dependence, though not perfect positive dependence, is observed between Y_{0t-1} and $Y0t-2$ for the group of displaced workers (Spearman's Rho = 0.86).

2.6 Conclusion

This chapter has developed techniques to study distributional treatment effect parameters that depend on the joint distribution of potential outcomes. The results depend on three key ingredients: (i) access to at least three periods of panel data, (ii) identification of the marginal distribution untreated potential outcomes for the treated group and (iii) the Copula Stability Assumption which says that the dependence between untreated potential outcomes over time does not change over time. The last of these is the key idea that allows the researcher to exploit having access to panel data to learn about the joint distribution of potential outcomes. This type of idea may also be useful in other cases where the researcher has access to panel data.

Using these methods, I have studied the distributional effects of job displacement during the Great Recession for older workers. Using standard techniques, I find that older workers lose 40% of their yearly earnings following job displacement. Using the techniques developed in the current chapter, I find that this average effect masks substantial heterogeneity: some older workers lose a very large fraction of their earnings following job displacement though at least some workers have higher earnings following displacement than they would have had they not been displaced. Finally, I also find that workers with initially higher earnings experience larger earnings losses from job displacement than workers with initially lower earnings, but as a fraction of earnings, the average earnings loss is very similar across initial income levels.

2.7 Proofs

Proof of Lemma 21

Proof.

$$
\begin{aligned}
\mathrm{F}_{Y_{1t},Y_{0t}|D_t=1}(y_1,y_0) &= \mathrm{P}(Y_{1t} \leq y_1, Y_{0t} \leq y_0 | D_t = 1) \\
&= \mathrm{P}(Y_{1t} \leq y_1, \mathrm{F}_{Y_{0t}|D_t=1}(Y_{0t}) \leq \mathrm{F}_{Y_{0t}|D_t=1}(y_0) | D_t = 1) \\
&= \mathrm{P}(Y_{1t} \leq y_1, \mathrm{F}_{Y_{0t-1}|D_t=1}(Y_{0t-1}) \leq \mathrm{F}_{Y_{0t}|D_t=1}(y_0) | D_t = 1) \\
&= \mathrm{P}(Y_{1t} \leq y_1, Y_{0t-1} \leq \mathrm{F}^{-1}_{Y_{0t-1}|D_t=1}(\mathrm{F}_{Y_{0t}|D_t=1}(y_0)) | D_t = 1)
\end{aligned}
$$

where the third equality uses the assumption of perfect positive dependence. □

Proofs of Lemma 22, Lemma 23, Theorem 8, Theorem 9, and Theorem 10

Lemma 22 is just an application of the Fréchet-Hoeffding bounds to a conditional bivariate distribution.

Lemma 23 applies the sharp bounds on the difference between random variables with known marginal distributions but unknown copula of Williamson and Downs (1990) to the difference conditional on the previous outcome.

Theorem 8 and Theorem 9 follow from results in Fan and Park (2010, Section 5) and Fan, Guerre, and Zhu (2015) which derive sharp bounds on the unconditional distribution of the treatment effect when conditional marginal distributions are known. In those cases, the marginal distributions are conditional on observed covariates X; in the current chapter, the marginal distributions are conditional on a lag of the outcome Y_{0t-1}.

Theorem 10 holds because inverting sharp bounds on a distribution implies sharp bounds on the quantiles (Williamson and Downs, 1990; Fan and Park, 2010).

Proof of Theorem 11

Proof. To show the result in Theorem 11, it must be shown that $\mathrm{E}[Y_{0t}|Y_{0t-1}=y',D_t=1] = \mathrm{E}[\mathrm{F}^{-1}_{Y_{0t}|D_t=1}(\mathrm{F}_{Y_{0t-1}|D_t=1}(Y_{0t-1})) \mid Y_{0t-2} = \mathrm{F}^{-1}_{Y_{0t-2}|D_t=1}(\mathrm{F}_{Y_{0t-1}|D_t=1}(y'))]$.

$$
\begin{aligned}
&\mathrm{E}[Y_{0t}|Y_{0t-1}=y',D_t=1] \\
&\qquad = \int_{\mathcal{Y}} y_{0t} f_{Y_{0t}|Y_{0t-1},D_t=1}(y_{0t} \mid y')\, \mathrm{d}y_{0t} \\
&\qquad = \int_{\mathcal{Y}} y_{0t} \frac{f_{Y_{0t},Y_{0t-1},D_t=1}(y_{0t},y')}{f_{Y_{0t-1}|D_t=1}(y')}\, \mathrm{d}y_{0t} \\
&\qquad = \int_{\mathcal{Y}} y_{0t} c_{Y_{0t},Y_{0t-1}|D_t=1}(\mathrm{F}_{Y_{0t}|D_t=1}(y_{0t}),\mathrm{F}_{Y_{0t-1}|D_t=1}(y')) f_{Y_{0t}|D_t=1}(y_{0t})\, \mathrm{d}y_{0t} \\
&\qquad = \int_{\mathcal{Y}} y_{0t} c_{Y_{0t-1},Y_{0t-2}|D_t=1}(\mathrm{F}_{Y_{0t}|D_t=1}(y_{0t}),\mathrm{F}_{Y_{0t-1}|D_t=1}(y')) f_{Y_{0t}|D_t=1}(y_{0t})\, \mathrm{d}y_{0t} \\
&\qquad = \int_{\mathcal{Y}} y_{0t} f_{Y_{0t-1},Y_{0t-2}|D_t=1}(\mathrm{F}^{-1}_{Y_{0t-1}|D_t=1}(\mathrm{F}_{Y_{0t}|D_t=1}(y_{0t})),\mathrm{F}^{-1}_{Y_{0t-2}|D_t=1}(\mathrm{F}_{Y_{0t-1}|D_t=1}(y'))) \\
&\qquad\qquad \times \frac{f_{Y_{0t}|D_t=1}(y_{0t})}{f_{Y_{0t-1}|D_t=1}(\mathrm{F}^{-1}_{Y_{0t-1}|D_t=1}(\mathrm{F}_{Y_{0t}|D_t=1}(y_{0t}))) \times f_{Y_{0t-2}|D_t=1}(\mathrm{F}^{-1}_{Y_{0t-2}|D_t=1}(\mathrm{F}_{Y_{0t-1}|D_t=1}(y')))}\, \mathrm{d}y_{0t} \\
&\qquad = \int_{\mathcal{Y}} y_{0t} f_{Y_{0t-1}|Y_{0t-2},D_t=1}(\mathrm{F}^{-1}_{Y_{0t-1}|D_t=1}(\mathrm{F}_{Y_{0t}|D_t=1}(y_{0t}))|\mathrm{F}^{-1}_{Y_{0t-2}|D_t=1}(\mathrm{F}_{Y_{0t-1}|D_t=1}(y'))) \\
&\qquad\qquad \times \frac{f_{Y_{0t}|D_t=1}(y_{0t})}{f_{Y_{0t-1}|D_t=1}(\mathrm{F}^{-1}_{Y_{0t-1}|D_t=1}(\mathrm{F}_{Y_{0t}|D_t=1}(y_{0t})))}\, \mathrm{d}y_{0t}
\end{aligned}
$$

Next, make the substitution $u = \mathrm{F}^{-1}_{Y_{0t-1}|D_t=1}(\mathrm{F}_{Y_{0t}|D_t=1}(y_{0t}))$ which implies

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

and

$$\mathrm{d}y_{0t} = \frac{f_{Y_{0t-1}|D_t=1}(u)}{f_{Y_{0t}|D_t=1}(\mathrm{F}^{-1}_{Y_{0t}|D_t=1}(\mathrm{F}_{Y_{0t-1}|D_t=1}(u)))}$$

Plugging these back in implies

$$\begin{aligned}
&\mathrm{E}[Y_{0t}|Y_{0t-1} = y', D_t = 1] \\
&\qquad = \int_{\mathcal{Y}} \mathrm{F}^{-1}_{Y_{0t}|D_t=1}(\mathrm{F}_{Y_{0t-1}|D_t=1}(u)) f_{Y_{0t-1}|Y_{0t-2},D_t=1}(u|\,\mathrm{F}^{-1}_{Y_{0t-2}|D_t=1}(\mathrm{F}_{Y_{0t-1}|D_t=1}(y'))) \,\mathrm{d}u \\
&\qquad = \mathrm{E}[\mathrm{F}^{-1}_{Y_{0t}|D_t=1}(\mathrm{F}_{Y_{0t-1}|D_t=1}(Y_{0t-1})) \,\big|\, Y_{0t-2} = \mathrm{F}^{-1}_{Y_{0t-2}|D_t=1}(\mathrm{F}_{Y_{0t-1}|D_t=1}(y'))]
\end{aligned}$$

□

Proofs of Models Satisying the Conditional Copula Stability Assumption

Recall, the conditional copula is given by

$$\mathrm{C}_{Y_{0t},Y_{0t-1}|X}(u,v|x) = \mathrm{P}(\mathrm{F}_{Y_{0t}|X}(Y_{0t}|x) \leq u, \mathrm{F}_{Y_{0t-1}|X}(Y_{0t-1}|x) \leq v|X = x)$$

The main tactic for the proof is to show that $\mathrm{F}_{Y_{0t}|X}(Y_{0t}|x)$ does not depend on time. This implies that the Conditional Copula Stability Assumption will hold.

Model 1

: $Y_{0t} = g(X_i, c_i, v_{it})$. In this case,

$$\mathrm{P}(Y_{0t} \leq y|X = x, D_t = 1) = \mathrm{P}(g(x, c_i, v_{it}) \leq y|X = x, D_t = 1)$$

$$= \mathrm{E}_{c_i,v_{it}|X=x,D_t=1}[1\{g(x,c_i,v_{it}) \leq y\}]$$

which implies

$$\mathrm{F}_{Y_{0t}|X,D_t=1}(\tilde{Y}_{0t}|x) = \mathrm{E}_{c_i,v_{it}|X=x,D_t=1}[1\{g(x,c_i,v_{it}) \leq g(x,\tilde{c}_i,\tilde{v}_{it})\}]$$

which does not change over time because the distribution of v_{it} does not change over time.

Model 2:

$Y_{0t} = g(X_i,c_i,v_{it}) + h_t(X_i)$. In this case,

$$\begin{aligned} \mathrm{P}(Y_{0t} \leq y|X=x,D_t=1) &= \mathrm{P}(g(x,c_i,v_{it}) + h_t(x) \leq y|X=x,D_t=1) \\ &= \mathrm{E}_{c_i,v_{it}|X=x,D_t=1}[1\{g(x,c_i,v_{it}) + h_t(x) \leq y\}] \end{aligned}$$

which implies

$$\mathrm{F}_{Y_{0t}|X,D_t=1}(\tilde{Y}_{0t}|x) = \mathrm{E}_{c_i,v_{it}|X=x,D_t=1}[1\{g(x,c_i,v_{it}) \leq g(x,\tilde{c}_i,\tilde{v}_{it})\}]$$

which does not change over time because the distribution of v_{it} does not change over time.

Model 3:

$Y_{0t} = g(X_i,c_i,v_{it}) + h_t(X_i,v_{it})$. In this case,

$$\begin{aligned} \mathrm{P}(Y_{0t} \leq y|X=x,D_t=1) &= \mathrm{P}(g(x,c_i,v_{it}) + h_t(x,v_{it}) \leq y|X=x,D_t=1) \\ &= \mathrm{E}_{c_i,v_{it}|X=x,D_t=1}[1\{g(x,c_i,v_{it}) + h_t(x,v_{it}) \leq y\}] \end{aligned}$$

which implies

$$\mathrm{F}_{Y_{0t}|X,D_t=1}(\tilde{Y}_{0t}|x) = \mathrm{E}_{c_i,v_{it}|X=x,D_t=1}[1\{g(x,c_i,v_{it}) + h_t(x,v_{it}) \leq g(x,\tilde{c}_i,\tilde{v}_{it}) + h_t(x,\tilde{v}_{it}\}]$$

which does not satisfy the CSA because of the interaction of time and time varying unobservable v_{it}.

Model 4:

$Y_{0t} = g(X_i, c_i, v_{it}) + h_t(X_i, c_i)$. In this case,

$$\begin{aligned}
\mathrm{P}(Y_{0t} \leq y | X = x, D_t = 1) &= \mathrm{P}(g(x, c_i, v_{it}) + h_t(x, c_i) \leq y | X = x, D_t = 1) \\
&= \mathrm{E}_{c_i, v_{it} | X = x, D_t = 1}[1\{g(x, c_i, v_{it}) + h_t(x, c_i) \leq y\}]
\end{aligned}$$

which implies

$$\mathrm{F}_{Y_{0t}|X, D_t=1}(\tilde{Y}_{0t}|x) = \mathrm{E}_{c_i, v_{it}|X=x, D_t=1}[1\{g(x, c_i, v_{it}) + h_t(x, c_i) \leq g(x, \tilde{c}_i, \tilde{v}_{it}) + h_t(x, \tilde{c}_i\}]$$

which does not satisfy the CSA because of the interaction of time and time invariant unobservables.

Proof of Lemma 24

The proof of Lemma 24 is very similar to the proof of Theorem 11 and is omitted.

Proof of Theorem 12

Before proving the main result, I state the following lemmas

LEMMA 25. *(Uniform convergence of empirical distribution function)*

For any $\delta < \frac{1}{2}$

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

LEMMA 26. *(Uniform convergence of empirical quantiles)*

For X with compact support, continuous density bounded from above and bounded away from 0, and for any $\delta < \frac{1}{2}$

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

Proof. See Athey and Imbens (2006, Lemma A.3) $\square$

Step 1: (Accounting for 1st Step Estimation)

The first step is to show that estimating the first step estimates of unconditional distribution and quantile functions do not affect the asymptotic distribution of the ATT-CPO. As a first step, rewrite Equation 2.6 as

$$\sqrt{nh}\{(\hat{a}_1(y') - \tilde{a}_0(y')) - \text{ATT-CPO}(y') - \text{Bias}(y')\} + \sqrt{nh}(\tilde{a}_0(y') - \hat{a}_0(y'))$$

First, I show that the term $\sqrt{nh}(\tilde{a}_0(y') - \hat{a}_0(y')) = o_p(1)$ by showing the slightly more general result that

$$\begin{aligned}\sqrt{n_T h}\left(\hat{\delta}_0(y') - \tilde{\delta}_0(y')\right) &= \sqrt{n_T h}\left((\hat{\mathcal{X}}_0'\hat{\mathcal{K}}_0(y')\hat{\mathcal{X}}_0)^{-1}\hat{\mathcal{X}}_0'\hat{\mathcal{K}}_0(y')\hat{\mathcal{Z}} - (\mathcal{X}_0'\mathcal{K}_0(y')\mathcal{X}_0)^{-1}\mathcal{X}_0'\mathcal{K}_0(y')\mathcal{Z}\right) \\ &= o_p(1)\end{aligned}$$

Recall that $\delta_0(y') = [a_0(y'), b_0(y')]'$, $\hat{\mathcal{Z}}$ is an $n_T \times 1$ vector with its ith component given by $\hat{\mathrm{F}}^{-1}_{Y_{0t}|D_t=1}(\hat{\mathrm{F}}_{Y_{0t-1}|D_t=1}(Y_{it-1}))$, $\hat{\mathcal{X}}_0$ is an $n_T \times 2$ matrix with ith row given by $(1, Y_{it-2} - \hat{\mathrm{F}}^{-1}_{Y_{0t-2}|D_t=1}(\hat{\mathrm{F}}_{Y_{0t-1}|D_t=1}(y')))$, and $\hat{\mathcal{K}}_0(y')$ is an $n_T \times n_T$ diagonal matrix with ith diagonal element given by $K_h(Y_{it-2} - \hat{\mathrm{F}}^{-1}_{Y_{0t-2}|D_t=1}(\hat{\mathrm{F}}_{Y_{0t-1}|D_t=1}(y')))$. And the versions without hats are their population counterparts; for example, $\mathcal{Z}$ is an $n_T \times 1$ vector with ith component $\mathrm{F}^{-1}_{Y_{0t}|D_t=1}(\mathrm{F}_{Y_{0t-1}|D_t=1}(Y_{it-1}))$. Then,

$$
\begin{aligned}
&\sqrt{n_T h}\left(\hat{\delta}_0(y') - \tilde{\delta}_0(y')\right) \\
&\quad = \sqrt{n_T h}\big((\hat{\mathcal{X}}_0'\hat{\mathcal{K}}_0(y')\hat{\mathcal{X}}_0)^{-1}\hat{\mathcal{X}}_0'\hat{\mathcal{K}}_0(y')\hat{\mathcal{Z}} - (\mathcal{X}_0'\mathcal{K}_0(y')\mathcal{X}_0)^{-1}\mathcal{X}_0'\mathcal{K}_0(y')\mathcal{Z}\big) \\
&\quad = \Bigg\{\frac{1}{n_T}\sum_{i\in\mathcal{T}} K_h(Y_{it-2} - \hat{\mathrm{F}}^{-1}_{Y_{0t-2}|D_t=1}(\hat{\mathrm{F}}_{Y_{0t-1}|D_t=1}(y'))) \\
&\qquad\qquad \times \begin{pmatrix} 1 \\ Y_{it-2} - \hat{\mathrm{F}}^{-1}_{Y_{0t-2}|D_t=1}(\hat{\mathrm{F}}_{Y_{0t-1}|D_t=1}(y'))) \end{pmatrix} \left(1; \quad Y_{it-2} - \hat{\mathrm{F}}^{-1}_{Y_{0t-2}|D_t=1}(\hat{\mathrm{F}}_{Y_{0t-1}|D_t=1}(y'))\right)\Bigg\}^{-1} \\
&\qquad \times \Bigg\{\frac{1}{n_T}\sum_{i\in\mathcal{T}} K_h(Y_{it-2} - \hat{\mathrm{F}}^{-1}_{Y_{0t-2}|D_t=1}(\hat{\mathrm{F}}_{Y_{0t-1}|D_t=1}(y'))) \\
&\qquad\qquad \times \sqrt{n_T h}\begin{pmatrix} 1 \\ Y_{it-2} - \hat{\mathrm{F}}^{-1}_{Y_{0t-2}|D_t=1}(\hat{\mathrm{F}}_{Y_{0t-1}|D_t=1}(y'))) \end{pmatrix} \hat{\mathrm{F}}^{-1}_{Y_{0t}|D_t=1}(\hat{\mathrm{F}}_{Y_{0t-1}|D_t=1}(Y_{it-1}))\Bigg\} \\
&\qquad - \Bigg\{\frac{1}{n_T}\sum_{i\in\mathcal{T}} K_h(Y_{it-2} - \mathrm{F}^{-1}_{Y_{0t-2}|D_t=1}(\mathrm{F}_{Y_{0t-1}|D_t=1}(y'))) \\
&\qquad\qquad \times \begin{pmatrix} 1 \\ Y_{it-2} - \mathrm{F}^{-1}_{Y_{0t-2}|D_t=1}(\mathrm{F}_{Y_{0t-1}|D_t=1}(y'))) \end{pmatrix} \left(1; \quad Y_{it-2} - \mathrm{F}^{-1}_{Y_{0t-2}|D_t=1}(\mathrm{F}_{Y_{0t-1}|D_t=1}(y'))\right)\Bigg\}^{-1} \\
&\qquad\qquad \times \Bigg\{\frac{1}{n_T}\sum_{i\in\mathcal{T}} K_h(Y_{it-2} - \mathrm{F}^{-1}_{Y_{0t-2}|D_t=1}(\mathrm{F}_{Y_{0t-1}|D_t=1}(y'))) \\
&\qquad\qquad\qquad \times \sqrt{n_T h}\begin{pmatrix} 1 \\ Y_{it-2} - \mathrm{F}^{-1}_{Y_{0t-2}|D_t=1}(\mathrm{F}_{Y_{0t-1}|D_t=1}(y'))) \end{pmatrix} \mathrm{F}^{-1}_{Y_{0t}|D_t=1}(\mathrm{F}_{Y_{0t-1}|D_t=1}(Y_{it-1}))\Bigg\} \\
&\quad = \Bigg\{\frac{1}{n_T}\sum_{i\in\mathcal{T}} K_h(Y_{it-2} - \mathrm{F}^{-1}_{Y_{0t-2}|D_t=1}(\mathrm{F}_{Y_{0t-1}|D_t=1}(y'))) \\
&\qquad\qquad \times \begin{pmatrix} 1 \\ Y_{it-2} - \mathrm{F}^{-1}_{Y_{0t-2}|D_t=1}(\mathrm{F}_{Y_{0t-1}|D_t=1}(y'))) \end{pmatrix} \left(1; \quad Y_{it-2} - \mathrm{F}^{-1}_{Y_{0t-2}|D_t=1}(\mathrm{F}_{Y_{0t-1}|D_t=1}(y'))\right)\Bigg\}^{-1} \\
&\qquad \times \Bigg\{\frac{1}{n_T}\sum_{i\in\mathcal{T}} K_h(Y_{it-2} - \mathrm{F}^{-1}_{Y_{0t-2}|D_t=1}(\mathrm{F}_{Y_{0t-1}|D_t=1}(y')))
\end{aligned}
$$

$$\times \left(\begin{array}{c} \sqrt{n_T h}\left(\hat{\mathrm{F}}^{-1}_{Y_{0t}|D_t=1}(\hat{\mathrm{F}}_{Y_{0t-1}|D_t=1}(Y_{it-1})) - \mathrm{F}^{-1}_{Y_{0t}|D_t=1}(\mathrm{F}_{Y_{0t-1}|D_t=1}(Y_{it-1}))\right) \\ \sqrt{n_T h}\left(\hat{X}_i(y') \times \hat{\mathrm{F}}^{-1}_{Y_{0t}|D_t=1}(\hat{\mathrm{F}}_{Y_{0t-1}|D_t=1}(Y_{it-1})) - X_i(y') \times \mathrm{F}^{-1}_{Y_{0t}|D_t=1}(\mathrm{F}_{Y_{0t-1}|D_t=1}(Y_{it-1}))\right) \end{array} \right) \Bigg\}$$

$$+ o_p(1)$$

where the third equality holds by uniform convergence of the distribution functions, the quantile functions, continuity of the kernel function, continuity of the inverse function, and combining some terms from the previous equation. Next, I show that the two terms multiplied by $\sqrt{n_T h}$ in the final equation are $o_p(1)$. For all i,

$$\sqrt{n_T h}\left(\hat{\mathrm{F}}^{-1}_{Y_{0t}|D_t=1}(\hat{\mathrm{F}}_{Y_{0t-1}|D_t=1}(Y_{it-1})) - \mathrm{F}^{-1}_{Y_{0t}|D_t=1}(\mathrm{F}_{Y_{0t-1}|D_t=1}(Y_{it-1}))\right) \tag{2.8}$$

$$\leq \sup_z \sqrt{n_T h}\left|\hat{\mathrm{F}}^{-1}_{Y_{0t}|D_t=1}(\hat{\mathrm{F}}_{Y_{0t-1}|D_t=1}(z)) - \mathrm{F}^{-1}_{Y_{0t}|D_t=1}(\mathrm{F}_{Y_{0t-1}|D_t=1}(z))\right|$$

$$\leq \sup_z \sqrt{n_T h}\left|\hat{\mathrm{F}}^{-1}_{Y_{0t}|D_t=1}(\hat{\mathrm{F}}_{Y_{0t-1}|D_t=1}(z)) - \mathrm{F}^{-1}_{Y_{0t}|D_t=1}(\hat{\mathrm{F}}_{Y_{0t-1}|D_t=1}(z))\right| \tag{2.9}$$

$$+ \sup_z \sqrt{n_T h}\left|\mathrm{F}^{-1}_{Y_{0t}|D_t=1}(\hat{\mathrm{F}}_{Y_{0t-1}|D_t=1}(z)) - \mathrm{F}^{-1}_{Y_{0t}|D_t=1}(\mathrm{F}_{Y_{0t-1}|D_t=1}(z))\right| \tag{2.10}$$

$$\xrightarrow{p} 0$$

where the result follows since

$$\sup_z \sqrt{n_T h}\left|\hat{\mathrm{F}}^{-1}_{Y_{0t}|D_t=1}(\hat{\mathrm{F}}_{Y_{0t-1}|D_t=1}(z)) - \mathrm{F}^{-1}_{Y_{0t}|D_t=1}(\hat{\mathrm{F}}_{Y_{0t-1}|D_t=1}(z))\right|$$
$$= \sup_{q\in[0,1]} \sqrt{n_T h}\left|\hat{\mathrm{F}}^{-1}_{Y_{0t}|D_t=1}(q) - \mathrm{F}^{-1}_{Y_{0t}|D_t=1}(q)\right|$$

which converges to 0 by Lemma 26. And, by a Taylor Expansion,

$$\sup_z \sqrt{n_T h}\left|\mathrm{F}^{-1}_{Y_{0t}|D_t=1}(\hat{\mathrm{F}}_{Y_{0t-1}|D_t=1}(z)) - \mathrm{F}^{-1}_{Y_{0t}|D_t=1}(\mathrm{F}_{Y_{0t-1}|D_t=1}(z))\right|$$
$$= \sup_z \left| \frac{1}{f_{Y_{0t}|D_t=1}(\mathrm{F}^{-1}_{Y_{0t}|D_t=1}(\mathrm{F}_{Y_{0t-1}|D_t=1}(z)))} \times \sqrt{n_T h}\left(\hat{\mathrm{F}}_{Y_{0t-1}|D_t=1}(z) - \mathrm{F}_{Y_{0t-1}|D_t=1}(z)\right) \right|$$

which converges to 0 because $f_{Y_{0t}|D_t=1}(\cdot)$ is bounded away from 0 on its support and the term $\sqrt{n_T h}\left(\hat{\mathrm{F}}_{Y_{0t-1}|D_t=1}(z) - \mathrm{F}_{Y_{0t-1}|D_t=1}(z)\right)$ converges to 0 by Lemma 25.

Next, I show that $\sqrt{n_T h}\left(\hat{X}_i(y') \times \hat{\mathrm{F}}^{-1}_{Y_{0t}|D_t=1}(\hat{\mathrm{F}}_{Y_{0t-1}|D_t=1}(Y_{it-1})) - X_i(y') \times \mathrm{F}^{-1}_{Y_{0t}|D_t=1}(\mathrm{F}_{Y_{0t-1}|D_t=1}(Y_{it-1}))\right)$ converges in probability to 0.

$$\sqrt{n_T h}\left(\hat{X}_i(y') \times \hat{\mathrm{F}}^{-1}_{Y_{0t}|D_t=1}(\hat{\mathrm{F}}_{Y_{0t-1}|D_t=1}(Y_{it-1})) - X_i(y') \times \mathrm{F}^{-1}_{Y_{0t}|D_t=1}(\mathrm{F}_{Y_{0t-1}|D_t=1}(Y_{it-1}))\right)$$

$$\leq \sup_z \sqrt{n_T h}\left|\left(\hat{\mathrm{F}}^{-1}_{Y_{0t-2}|D_t=1}(\hat{\mathrm{F}}_{Y_{0t-1}|D_t=1}(y')) - \mathrm{F}^{-1}_{Y_{0t-2}|D_t=1}(\mathrm{F}_{Y_{0t-1}|D_t=1}(y'))\right) \times \hat{\mathrm{F}}^{-1}_{Y_{0t}|D_t=1}(\hat{\mathrm{F}}_{Y_{0t-1}|D_t=1}(z))\right| \tag{2.11}$$

$$+ \sup_z \sqrt{n_T h}\left|\mathrm{F}^{-1}_{Y_{0t-2}|D_t=1}(\mathrm{F}_{Y_{0t-1}|D_t=1}(y')) \times \left(\hat{\mathrm{F}}^{-1}_{Y_{0t}|D_t=1}(\hat{\mathrm{F}}_{Y_{0t-1}|D_t=1}(z)) - \mathrm{F}^{-1}_{Y_{0t}|D_t=1}(\mathrm{F}_{Y_{0t-1}|D_t=1}(z))\right)\right| \tag{2.12}$$

where the second line holds by adding and subtracting $\left(z - \mathrm{F}^{-1}_{Y_{0t-2}|D_t=1}(\mathrm{F}_{Y_{0t-1}|D_t=1}(y'))\right) \times \hat{\mathrm{F}}^{-1}_{Y_{0t}|D_t=1}(\hat{\mathrm{F}}_{Y_{0t-1}|D_t=1}(z))$ and by the triangle inequality. Equation 2.11 converges to 0 because $\hat{\mathrm{F}}^{-1}_{Y_{0t}|D_t=1}(\hat{\mathrm{F}}_{Y_{0t-1}|D_t=1}(z))$ is bounded for all z and one can show that $\hat{\mathrm{F}}^{-1}_{Y_{0t-2}|D_t=1}(\hat{\mathrm{F}}_{Y_{0t-1}|D_t=1}(y')) - \mathrm{F}^{-1}_{Y_{0t-2}|D_t=1}(\mathrm{F}_{Y_{0t-1}|D_t=1}(y'))$ converges to 0 by exactly the same sort of argument as for Equation 2.8.

Likewise, Equation 2.12 converges to 0 $\mathrm{F}^{-1}_{Y_{0t-2}|D_t=1}(\mathrm{F}_{Y_{0t-1}|D_t=1}(y'))$ is bounded and

$$\sup_z \left|\sqrt{n_T h}\left(\hat{\mathrm{F}}^{-1}_{Y_{0t}|D_t=1}(\hat{\mathrm{F}}_{Y_{0t-1}|D_t=1}(z)) - \mathrm{F}^{-1}_{Y_{0t}|D_t=1}(\mathrm{F}_{Y_{0t-1}|D_t=1}(z))\right)\right|$$

converges to 0; this is exactly the result coming from Equation 2.8.

Step 2: (Asymptotically Linear Representations)

Let $W_i = \mathrm{F}^{-1}_{Y_{0t}|D_t=1}(\mathrm{F}_{Y_{0t-1}|D_t=1}(Y_{it-1}))$ and $z = \mathrm{F}^{-1}_{Y_{0t-2}|D_t=1}(\mathrm{F}_{Y_{0t-1}|D_t=1}(y'))$. Then, standard results on estimating conditional distributions (see, for example, Fan and Gijbels, 1996; Li and Racine, 2007) implies

$$\sqrt{n_T h}\left(\hat{a}_1(y') - \mathrm{E}[Y_{1t}|Y_{0t-1} = y', D_t = 1] - \frac{\kappa_2}{2} f_{Y_{0t-1}|D_t=1}(y') g_1''(y') h_1^2\right)$$
$$= \frac{1}{f_{Y_{0t-1}|D_t=1}(y')} \sqrt{n_T h}\left(\frac{1}{n_T}\sum_{i\in\mathcal{T}} K_{h_1}(y', Y_{0it-1})\varepsilon_{it}\right) + o_p(1) \quad (2.13)$$

and

$$\sqrt{n_T h}\left(\tilde{a}_0(y') - \mathrm{E}[Y_{0t}|Y_{0t-1} = y', D_t = 1] - \frac{\kappa_2}{2} f_{Y_{0t-2}|D_t=1}(z) g''(z) h^2\right)$$
$$= \frac{1}{f_{Y_{0t-2}|D_t=1}(z)} \sqrt{n_T h}\left(\frac{1}{n_T}\sum_{i\in\mathcal{T}} K_h(z, Y_{0it-2}) u_{it}\right) + o_p(1) \quad (2.14)$$

Step 3: (Asymptotic Normality)

Combining Equation 2.13 and Equation 2.14 implies

$$\begin{aligned}
&\sqrt{nh}\left(\hat{a}_1(y') - \hat{a}_0(y') - \mathrm{E}[Y_{1t} - Y_{0t}|Y_{0t-1} = y', D_t = 1]\right. \\
&\qquad\qquad \left. - \frac{\kappa_2}{2}\left(f_{Y_{0t-1}|D_t=1}(y') g_1''(y') h_1^2 - f_{Y_{0t-2}|D_t=1}(z) g''(z) h^2\right)\right) \\
&= \frac{1}{f_{Y_{0t-1}|D_t=1}(y')} \sqrt{n_T h}\left(\frac{1}{n_T}\sum_{i\in\mathcal{T}} K_{h_1}(y', Y_{0it-1})\varepsilon_{it}\right) \\
&\qquad - \frac{1}{f_{Y_{0t-2}|D_t=1}(z)} \sqrt{n_T h}\left(\frac{1}{n_T}\sum_{i\in\mathcal{T}} K_h(z, Y_{0it-2} u_{it}\right) + o_p(1) \\
&= \begin{pmatrix} \frac{1}{f_{Y_{0t-1}|D_t=1}(y')} \\ -\frac{1}{f_{Y_{0t-2}|D_t=1}(z)} \end{pmatrix}^{\top} \begin{pmatrix} \sqrt{n_T h}\left(\frac{1}{n_T}\sum_{i\in\mathcal{T}} K_{h_1}(y', Y_{0it-1})\varepsilon_{it}\right) \\ \sqrt{n_T h}\left(\frac{1}{n_T}\sum_{i\in\mathcal{T}} K_h(z, Y_{0it-2}) u_{it}\right) \end{pmatrix} + o_p(1)
\end{aligned}$$

Since

$$\begin{aligned}
\operatorname{Var}\left(\sqrt{n_T h}\frac{1}{n_T}\sum_{i\in\mathcal{T}} K_h(z,Y_{0it-2})u_{it}\right) &= h\,\mathrm{E}\left[k^2\left(\frac{Y_{0it-2}-z}{h}\right)u_{it}^2\right] \\
&= \frac{1}{h}\,\mathrm{E}\left[\sigma_u^2(Y_{0it-2})k^2\left(\frac{Y_{0it-2}-z}{h}\right)\right] \\
&= \int \sigma_u^2(z+vh)k^2(v)\,f_{Y_{0t-2}|D_t=1}(z+vh)\ \mathrm{d}v \\
&= \sigma_u^2(z)f_{Y_{0t-2}|D_t=1}(z)\int k^2(v)\ \mathrm{d}v + o(1)
\end{aligned}$$

where $\sigma_u^2(z) = \mathrm{E}[u_{it}^2|Y_{0t-2}=z]$. Similarly, one can show that

$$\operatorname{Var}\left(\sqrt{n_T h}\frac{1}{n_T}\sum_{i\in\mathcal{T}} K_{h_1}(y',Y_{0it-1})\varepsilon_{it}\right) = \sigma_\varepsilon^2(y')f_{Y_{0t-1}|D_t=1}(y')\int k^2(v)\ \mathrm{d}v + o(1)$$

and

$$\begin{aligned}
&\operatorname{Cov}\left(\sqrt{n_T h}\frac{1}{n_T}\sum_{i\in\mathcal{T}} K_h(z,Y_{0it-2})u_{it},\sqrt{n_T h}\frac{1}{n_T}\sum_{i\in\mathcal{T}} K_{h_1}(y',Y_{0it-1})\varepsilon_{it}\right) \\
&\quad= \frac{1}{h}\,\mathrm{E}\left[k_h\left(\frac{Y_{0it-2}-z}{h}\right)u_{it}k_{h_1}\left(\frac{Y_{0it-1}-y'}{h_1}\right)\varepsilon_{it}\right] \\
&\quad= \frac{1}{h}\,\mathrm{E}\left[\sigma_{u\varepsilon}(z,y')k_h\left(\frac{Y_{0it-2}-z}{h}\right)k_{h_1}\left(\frac{Y_{0it-1}-y'}{h_1}\right)\right] \\
&\quad= \frac{1}{h}\int\int \sigma_{u\varepsilon}(z,y')k_h\left(\frac{\tilde{z}-z}{h}\right)k_{h_1}\left(\frac{y_{0t-1}-y'}{h_1}\right)f_{Z,Y_{0t-1}|D_t=1}(\tilde{z},y_{0t-1})\ \mathrm{d}\tilde{z}\ \mathrm{d}y_{0t-1} \\
&\quad= h_1\int\int \sigma_{u\varepsilon}(z,y')k_h(\tilde{v})\,k_{h_1}(v)\,f_{Z,Y_{0t-1}|D_t=1}(z+\tilde{v}h,y'+vh_1)\ \mathrm{d}\tilde{v}\ \mathrm{d}v \\
&\quad= O(h_1)
\end{aligned}$$

This implies

$$\begin{pmatrix} \sqrt{n_T h}\left(\frac{1}{n_T}\sum_{i\in\mathcal{T}} K_{h_1}(y', Y_{0it-1})\varepsilon_{it}\right) \\ \sqrt{n_T h}\left(\frac{1}{n_T}\sum_{i\in\mathcal{T}} K_h(z, Y_{0it-2})u_{it}\right) \end{pmatrix} \xrightarrow{d} N(0,\Omega)$$

where

$$\Omega = \begin{pmatrix} \sigma_\varepsilon^2(y') f_{Y_{0t-1}|D_t=1}(y')\kappa & 0 \\ 0 & \sigma_u^2(z) f_{Y_{0t-2}|D_t=1}(z)\kappa \end{pmatrix}$$

which implies the result that

$$\begin{aligned} \sqrt{nh}\Big(a_1(y') - a_0(y') - \mathrm{E}[Y_{1t} - Y_{0t}|Y_{0t-1} = y', D_t = 1] \\ -\frac{\kappa_2}{2}\left(f_{Y_{0t-1}|D_t=1}(y')g_1''(y')h_1^2 - f_{Y_{0t-2}|D_t=1}(z)g''(z)h^2\right)\Big) \\ \xrightarrow{d} N(0,V) \end{aligned}$$

where

$$\begin{aligned} V &= \frac{\sigma_\varepsilon^2(y')}{f_{Y_{0t-1}|D_t=1}(y')}\kappa + \frac{\sigma_u^2(z)}{f_{Y_{0t-2}|D_t=1}(z)}\kappa \\ &= \frac{\sigma_\varepsilon^2(y')}{f_{Y_{0t-1}|D_t=1}(y')}\kappa + \frac{\sigma_u^2(\mathrm{F}^{-1}_{Y_{0t}|D_t=1}(\mathrm{F}_{Y_{0t-1}|D_t=1}(y')))}{f_{Y_{0t-2}|D_t=1}(\mathrm{F}^{-1}_{Y_{0t-2}|D_t=1}(\mathrm{F}_{Y_{0t-1}|D_t=1}(y')))}\kappa \end{aligned}$$

Table 2.1: Summary Statistics

	Displaced		**Non-displaced**	
	mean	sd	mean	sd
Earnings $1000s				
Earnings 2006	43.2	39.9	48.0	42.7
Earnings 2008	42.6	39.9	50.4	57.5
Earnings 2010	30.1	39.3	48.4	44.4
Earnings 2012	35.0	42.6	51.4	55.4
Demographics				
Birth Year	1948.7	6.8	1947.8	6.8
% Male	0.44	0.50	0.44	0.50
% White	0.81	0.39	0.85	0.35
% Black	0.12	0.33	0.09	0.29
% No Degree	0.07	0.26	0.05	0.22
% HS Degree	0.59	0.49	0.56	0.50
% College Degree	0.33	0.47	0.39	0.49
Employment				
Works FT 2006	0.787	0.410	0.768	0.422
Works FT 2008	0.650	0.478	0.731	0.444
Works FT 2010	0.494	0.502	0.683	0.465
Works FT 2012	0.512	0.501	0.589	0.492
Unemployed 2006	0.000	0.000	0.000	0.000
Unemployed 2008	0.069	0.254	0.009	0.095
Unemployed 2010	0.163	0.370	0.011	0.103
Unemployed 2012	0.069	0.254	0.017	0.128
Retired 2006	0.000	0.000	0.000	0.000
Retired 2008	0.006	0.079	0.017	0.128
Retired 2010	0.087	0.283	0.021	0.145
Retired 2012	0.062	0.243	0.100	0.300

Table 2.2: Correlated Random Effects Estimates of the Effect of Job Displacement

Dependent Variable Log(Earnings)

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
Displaced	-0.533^*	-0.539^*	-0.677^*	-0.412^*	-0.474^*	-0.517^*	-0.335
	(0.126)	(0.112)	(0.129)	(0.162)	(0.118)	(0.136)	(0.185)
2012 Dummy	-0.053	-0.050	-0.005	-0.142	0.066	0.089^*	0.046
	(0.041)	(0.037)	(0.035)	(0.084)	(0.042)	(0.040)	(0.099)
Birth Year		0.051^*	0.047^*	0.049^*	0.004	0.004	0.010
		(0.003)	(0.003)	(0.005)	(0.005)	(0.005)	(0.011)
Female		-0.373^*	-0.360^*	-0.381^*	-0.369^*	-0.352^*	-0.396^*
		(0.036)	(0.042)	(0.071)	(0.041)	(0.048)	(0.085)
African American		-0.127^*	-0.067	-0.087	-0.190^*	-0.107	-0.172
		(0.061)	(0.062)	(0.102)	(0.068)	(0.070)	(0.116)
HS		0.349^*	0.103	0.454^*	0.395^*	0.238^*	0.478^*
		(0.081)	(0.084)	(0.148)	(0.096)	(0.100)	(0.191)
BA		0.717^*	0.291^*	0.522^*	0.818^*	0.503^*	0.608^*
		(0.081)	(0.091)	(0.157)	(0.096)	(0.106)	(0.195)
MA/MBA		0.990^*	0.638^*	0.845^*	0.964^*	0.797^*	0.872^*
		(0.088)	(0.101)	(0.172)	(0.104)	(0.118)	(0.229)
Law/MD/PhD		1.392^*	0.872^*	1.026^*	1.416^*	1.126^*	0.864^*
		(0.110)	(0.127)	(0.223)	(0.148)	(0.161)	(0.345)
Subsample					Young	Young	Young
Occupation FE	No	No	Yes	Yes	No	Yes	Yes
Time Varying Occ.	No	No	No	Yes	No	No	Yes
Industry FE	No	No	Yes	No	No	Yes	No
R^2	0.016	0.224	0.349	0.329	0.179	0.305	0.351
Adj. R^2	0.015	0.219	0.334	0.287	0.170	0.278	0.276
Num. obs.	3132	3132	2818	1298	1872	1684	720

$^*p < 0.05$

The dependent variable is the log of earnings in 2006 and 2012. The excluded group is white males without a HS degree. The regressions also include an additional dummy variables for Other Race, GED, Associate's Degree, and Other Education Level whose estimated coefficients are not presented for conciseness. The sample size decreases when Occupation and Industry Fixed Effects are included because these are not observed for all individuals in the sample. The last two columns limit the sample to individuals who are 64 or younger in 2012.

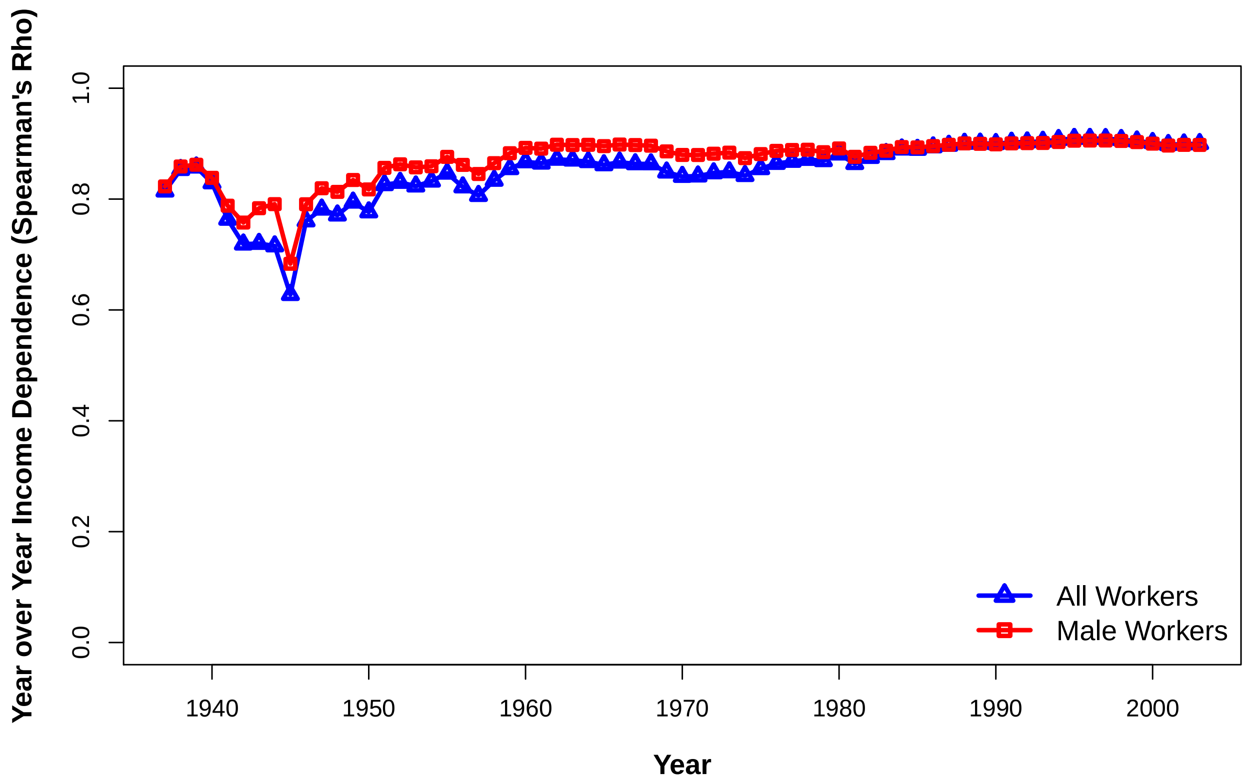


Figure 2.1: Rank Correlation of Year over Year Annual Income, 1937-2003
The data comes from Kopczuk, Saez, and Song (2010) and replicates part of Figure 4 in that paper.

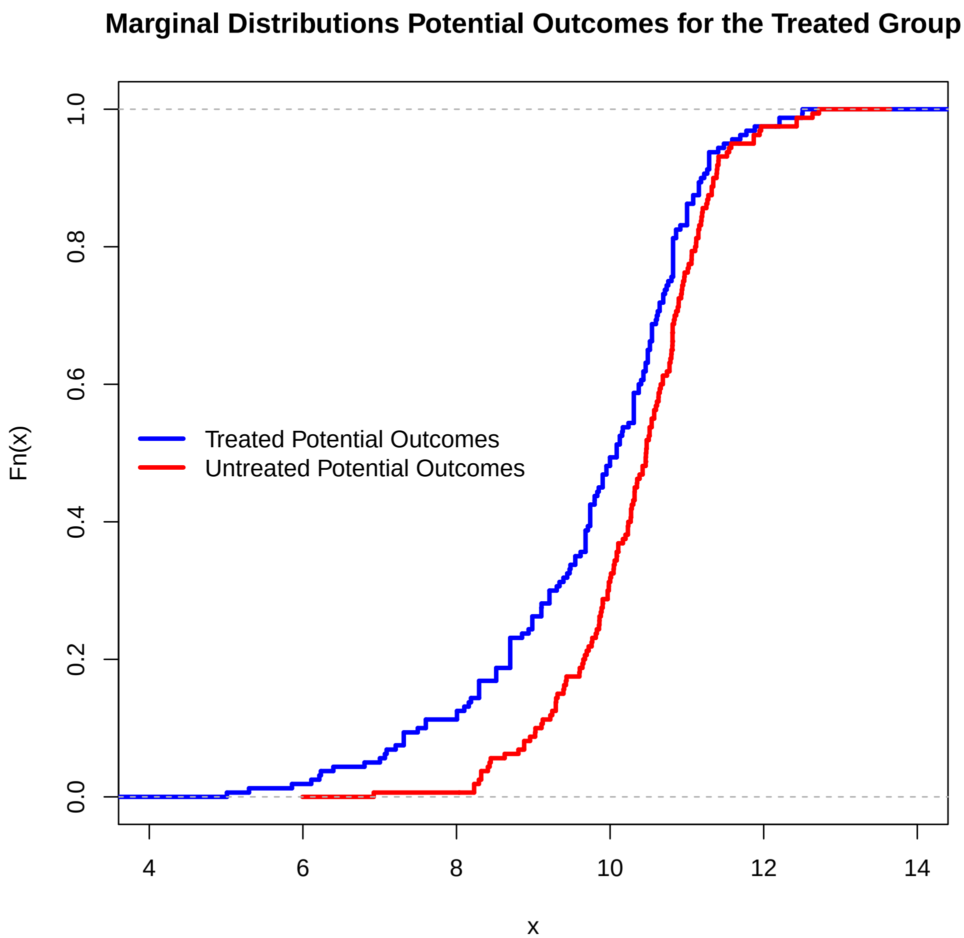


Figure 2.2: Marginal Distributions of Displaced and Non-displaced Potential Earnings for the Displaced Group of Workers

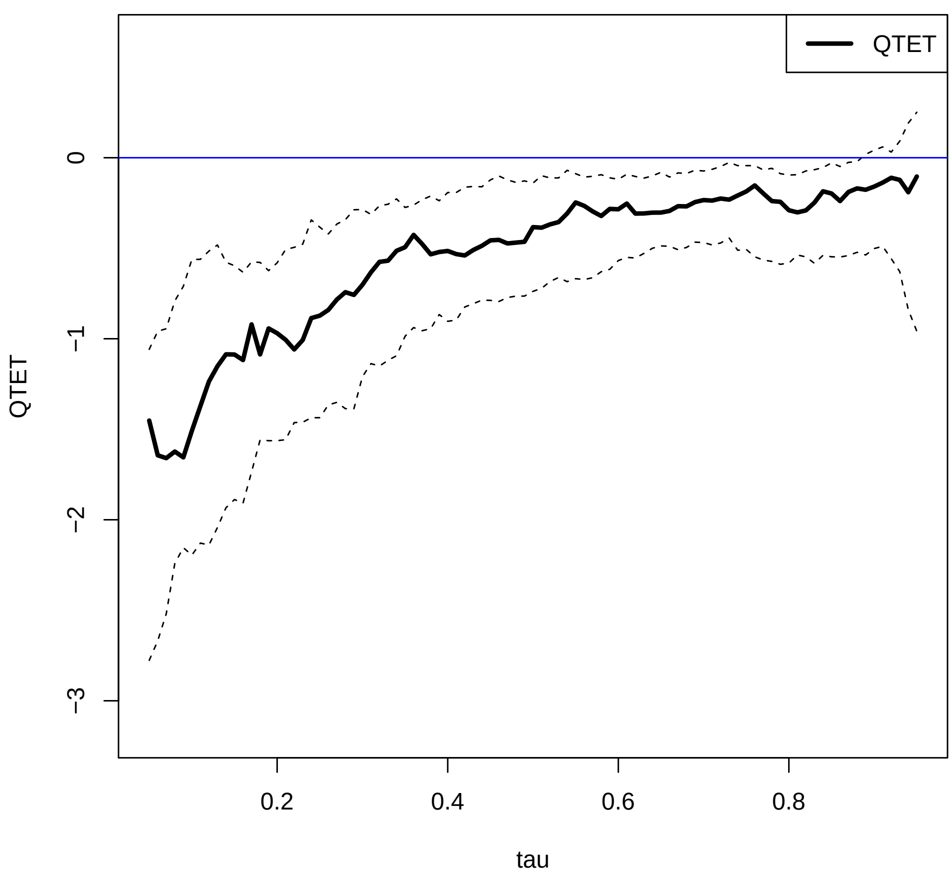


Figure 2.3: The Quantile Treatment Effect on the Treated estimated under the Distributional Difference in Differences Assumption.

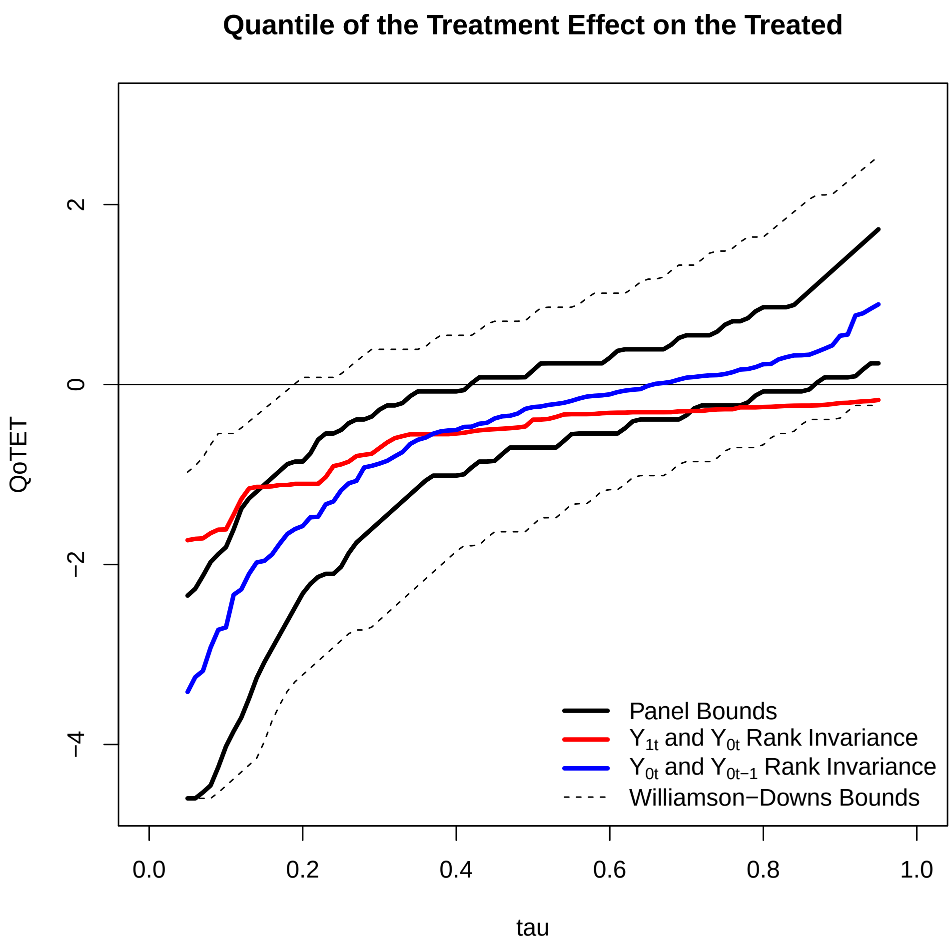


Figure 2.4: Bounds on the Quantile of the Treatment Effect

The 'Panel Bounds' are the estimates coming from the method in the current chapter. The 'Y_{1t} and Y_{0t} Rank Invariance' estimates come from employing the cross-sectional perfect positive dependence assumption. The 'Y_{0t} and Y_{0t-1} Rank Invariance' estimates come from applying the assumption of perfect positive dependence in non-displaced potential earnings over time. The 'Williamson-Downs Bounds' come from using only information about the marginal distributions of displaced and non-displaced potential earnings without applying any restrictions on their dependence.

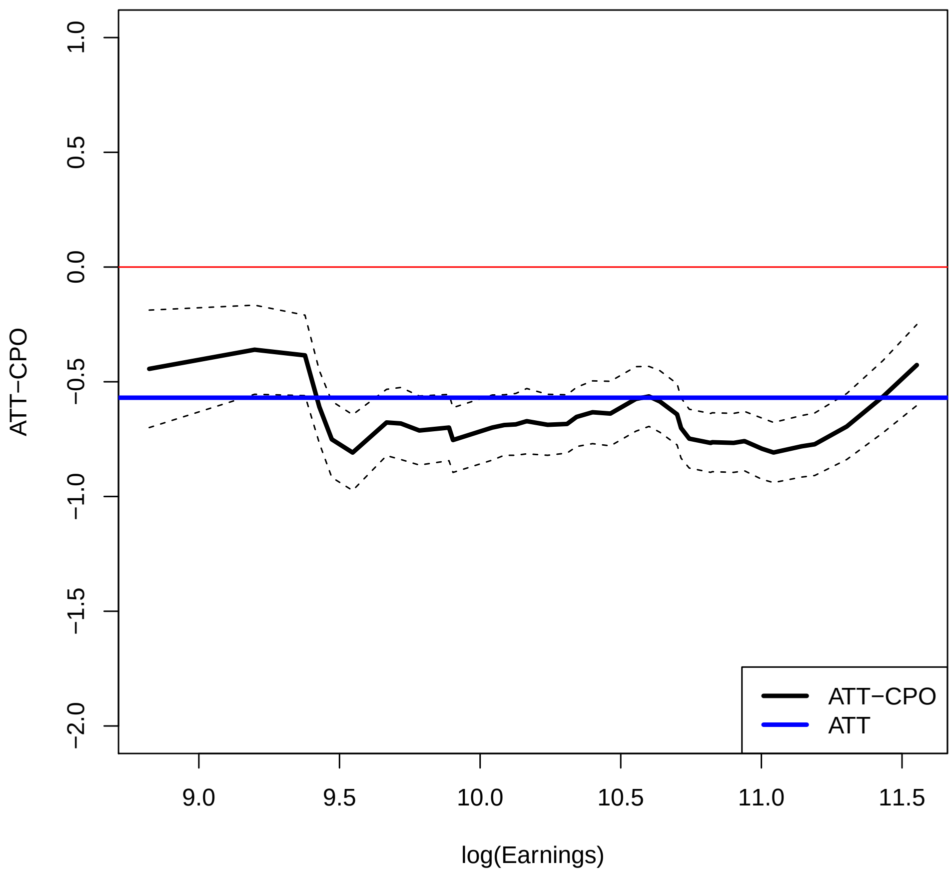


Figure 2.5: The ATT-CPO

Chapter 3

Quantile Treatment Effects in Difference in Differences Models under Dependence Restrictions and with only Two Time Periods[1]

3.1 Introduction

Researchers and policy makers are interested in evaluating the effect of participating in a program or experiencing a treatment but this is not a trivial task due to self-selection. Gronau (1974) and Heckman (1974) study the issue of self-selection in the context of labor market and Amemiya (1985) provides a comprehensive framework for this issue in his influential book *Advanced Econometrics*, which also includes his seminal paper Amemiya (1973) on the Tobit model. In the literature, the Average Treatment Effect (ATE) or Average Treatment Effect on the Treated (ATT) has received great attention. But there are cases where a researcher may be interested in studying the distributional effect of treatment. To give a few examples, labor economists may be interested in how a job training program impacts the lower tail of the earnings distribution; education economists may be interested in how smaller class sizes affect test scores at different points in the distribution of test scores; health economists may be interested in how smoking affects birthweight in the lower part of the distribution; and international trade economists may be interested in how joining the WTO affects bilateral trade between countries that otherwise have low trade flows. The current chapter considers identification and estimation of a particular distributional treatment effect parameter called the Quantile Treatment Effect on the Treated (QTT) under a Difference in Differences (DID) assumption when only two periods of panel or repeated cross sections data are available.

For a fixed $\tau \in (0,1)$, QTT(τ) is the horizontal distance between the τ-th quantile of treated potential outcomes for the treated group and the τ-th quantile of counterfactual

[1] This chapter is co-authored with Tong Li (Vanderbilt University) and Tatsushi Oka (National University of Singapore)

untreated potential outcomes for the treated group. For example QTT(0.5) is the difference between the median outcome for the treated group and the median of the counterfactual distribution of outcomes had the treated group not received treatment. A labor economist interested in studying the effect of a job training program on earnings at the bottom of the earnings distribution is likely to be interested in QTT(τ) for small values of τ.

The main identifying assumption is a distributional extension of the most common Mean DID Assumption. The key idea underlying a DID approach is that the unobserved *path* of untreated potential outcomes for the treated group is the same as the observed path of untreated potential outcomes for the untreated group. Mean DID requires this assumption to hold on average. The Distributional DID Assumption employed in the current chapter requires the distribution of the path to be the same for the treated and untreated groups.

Existing work on identifying the QTT under a Distributional DID Assumption either requires at least three periods of data (Callaway and Li, 2015) or results in partial identification (Fan and Yu, 2012). We consider a new set of conditions for point identification with only two periods of data. For applied researchers, it is important to have results when only two periods are available because it is not uncommon for researchers to have exactly two periods of data. For example, 25% of the papers employing DID assumptions considered by Bertrand, Duflo, and Mullainathan (2004) used exactly two periods of data. To give some specific examples, the Current Population Survery (CPS) Merged Outgoing Rotation Groups contains a 2-period panel (Madrian and Lefgren, 2000; Riddell and Song, 2011); the Displaced Workers Survey contains data on current wages and wages before displacement (Farber, 1997).

With panel data, the path of untreated potential outcomes for the treated group is observed; but repeated cross sections data presents a more challenging situation because, although untreated outcomes are observed at two different points in time, the path of these outcomes is not observed. Estimating the ATT with repeated cross sections imposes little additional challenge because the expectation of the difference is equal to the difference of the expectations, but the same approach does not work for the distribution. Instead, we impose the assumption of perfect positive dependence of untreated potential outcomes over time for the untreated group.[2] This assumption effectively turns the untreated group into a panel. For example, if there are the same number of untreated observations in each period, then the change in untreated potential outcomes is given by the difference between outcomes for individuals with the same rank across periods. Although this assumption is not likely to hold in practice, it appears to be "close" to holding. In our empirical example, the estimates of the QTT are very similar when we turn the panel dataset into repeated cross sections by de-linking individuals across time.

Alone, the Distributional DID Assumption is not strong enough to identify the QTT because the counterfactual distribution of untreated potential outcomes for the treated group is not identified under this assumption alone. Identifying the counterfactual distribution hinges on (i) knowing the distribution of the change in untreated potential outcomes for the treated group and (ii) knowing the dependence, or copula, between the change and initial level of untreated potential outcomes for the treated treated group. The Distributional DID Assumption handles the first identification challenge but not the second. The key innovation of the current chapter is to replace the unknown copula with the observed copula between the change and initial level of untreated potential outcomes for the untreated group. For example, if we observe in the untreated group that most of the increases in earnings over time go to individuals initially at the top of the earnings distribution, we assume that most of the increases in earnings that treated individuals *would have experienced* had they not been treated go to those that are at the top of the earnings distribution in the period before treatment. Importantly, this assumption allows for the marginal distribution of untreated outcomes in the period before treatment to differ for the treated group and the control group. This assumption handles the second identification challenge and implies

[2]This assumption is discussed in Heckman and Smith (1998) as a way to identify the joint distribution of potential outcomes in a panel data setup.

that the counterfactual distribution of untreated potential outcomes for the treated group is identified and, hence, the QTT is identified. Moreover, this assumption only requires that the researcher has access to two periods of data.

A second contribution of the chapter is to consider identification when the assumptions hold after conditioning on some covariates. A conditional DID Assumption allows the path of outcomes to be different for individuals with different observable characteristics (Heckman, Ichimura, Smith, and Todd, 1998; Heckman and Smith, 1999; Abadie, 2005). For example, the path of earnings might be quite different for highly educated individuals than for less educated individuals. If this is the case, our method also requires a conditional Copula Invariance Assumption. We show that under these assumptions, the QTT is still identified.

Estimating distributional treatment effect parameters is becoming more common in applied work. Recently, distributional treatment effects have been estimated in the context of welfare reform (Bitler, Gelbach, and Hoynes, 2006; Bitler, Gelbach, and Hoynes, 2008), conditional cash transfer programs in developing countries (Djebbari and Smith, 2008), head start (Bitler, Hoynes, and Domina, 2014), and the effect of Job Corps (Eren and Ozbeklik, 2014). One thing that each of the above papers have in common is that each uses experimental data. This chapter contributes to a growing literature on estimating quantile treatment effects with observational. Firpo (2007) considers quantile treatment effects under a selection on observables assumption. Abadie (2003) and Frölich and Melly (2013) consider quantile treatment effects when a researcher has access to an instrument that satisfies the exclusion restriction only after conditioning on some covariates. Melly and Santangelo (2015) extend the Change in Changes model (Athey and Imbens, 2006) to the case where the identifying assumptions hold conditional on covariates. Callaway and Li (2015) consider a distributional DID assumption that can hold conditional on covariates.

The QTT that we estimate is not conditional on covariates. The only role for covariates in our setting is to make the identifying assumptions more plausible. This setup is distinct from much of the literature on quantile regression and quantile treatment effects (Abadie, Angrist, and Imbens (2002) and Chernozhukov and Hansen (2005) though exceptions include Abadie (2003), Firpo (2007), and Frölich and Melly (2013)). In the job training example, at lower quantiles, the QTT is the difference in outcomes for those with (unconditionally) low earnings. On the other hand, if, for example the covariates include education, the conditional QTT at lower quantiles could actually represent individuals that have medium or high earnings relative to the overall population but have low earnings compared to individuals with the same characteristics. A secondary advantage of studying unconditional QTTs is that our estimators converge at the parametric $\sqrt{n}$-rate without making any parametric assumptions which would not be possible for conditional QTTs unless all covariates were discrete.

Given the point-identification result for the counterfactual distribution, we propose a two-step estimation procedure based on empirical distributions. In the first step, we estimate empirical distributions of observed outcomes for the treated and untreated groups separately. In the second step, the first-step estimates are used in the estimation of the the distribution of the counterfactual outcome by empirical distribution. The QTT is estimated by inverting the estimated counterfactual distribution of potential outcome for the treated group. The proposed estimator is shown to converge in distribution to a Gaussian random process at the parametric rate through empirical process techniques, while the limiting process is not nuisance parameter free because of estimation error at the first step and our identification strategy. To obtain an accurate approximation for the limiting process in a finite sample, we consider the exchangeable bootstrap proposed by Præstgaard and Wellner (1993) and show its first order validity.

The chapter is organized as follows. In Section 2, we provide identification results for the QTT including the cases where the assumptions hold conditional on covariates and when only repeated cross sections data is available. In Section 3, we discuss an estimation procedure and provide asymptotic results. We also introduce a resampling procedure and give a theoretical result on its consistency. In Section 4 we present an empirical application on the effect of a job training program. Section 5 concludes.

3.2 Identification

This section considers the main identification results in the current chapter. We consider identification of the QTT when the identifying assumptions hold unconditionally and when the identifying assumptions hold only after conditioning on some observed covariates X. We also consider additional requirements for identification when only repeated cross sections are available. Through the chapter, we use $\text{supp}(V)$ and $\text{supp}(V|W)$ to denote the support of V and the support of V conditional on W, respectively, for some random variables V and W.

3.2.1 Identification without Covariates

We consider a DID framework, in which all individuals in the sample receive no treatment before period $t-1$ while a fraction of individuals receive the treatment after period t. Let D_{is} be a treatment indicator that takes the value one if individual i is treated at a period s and zero otherwise. For each individual i, there is a pair of potential outcomes $(Y_{is}(0), Y_{is}(1))$ in period s, where $Y_{is}(0)$ and $Y_{is}(1)$ denote potential outcomes in the treated and untreated state in period s, respectively. Every individual experiences either treated or untreated status but not both, and thus the pair of potential outcomes is not observable.

We suppose that researchers can access panel data, which consist of outcomes and treatment statuses for each individual over some periods including both before and after the implementation of the program of interest. We consider the case with two-periods panel data and so that the data consists of observations $\{(Y_{i,t-1}, Y_{it}, D_{it})\}_{i=1}^{n}$ with n denoting the sample size, where the observed outcomes are given by

$$Y_{i,t-1} := Y_{i,t-1}(0) \quad \text{and} \quad Y_{it} := (1-D_{it})Y_{it}(0) + D_{it}Y_{it}(1). \tag{3.1}$$

Throughout the chapter, we assume independent and identically distributed (i.i.d.) observations within treatment and control group as stated below.

Assumption A1 (Random sampling). The data consists of observations $\{(Y_{i,t-1}, Y_{it}, D_{it})\}_{i=1}^{n}$ from the structure in (3.1) and the potential outcomes $\big(Y_{i,t-1}(0), Y_{i,t-1}(1)\big)$ and $\big(Y_{it}(0), Y_{it}(1)\big)$ are cross-sectionally i.i.d. conditional on treatment status D_{it}.

This assumption allows for the possibility that the marginal or joint distributions of potential outcomes can be different between treatment and control groups. Under this conditional random sampling assumption, we use $F_{Y_s(0)|D_t=d}$ and $F_{Y_s(1)|D_t=d}$ to denote marginal distributions of the potential outcomes conditional on treatment status in period s for $d = 0,1$, and let $F_{Y_s|D_t=d}$ be the marginal distribution of the outcome Y_{is} in period s conditional on $D_{it} = d$ for $s = t, t-1$ and for $d = 0,1$. Also we define $\mathcal{I}_d := \big\{i \in \{1,\ldots,n\} : D_{it} = d\big\}$ to denote a set of indices for individuals with $D_{it} = d$ for $d = 0,1$ and let n_d be the number of observations with $D_{it} = d$.

Our primary goal is to identify distributional features of treatment effects through conditional distributions of observed outcomes given D_t. The issue of identification of treatment effects arises because the pair of potential outcomes is unobservable for each individual and thus marginal distributions of potential outcomes are not necessarily identified from data. For instance, the conditional distribution $F_{Y_t(1)|D_t=1}$ of a potential outcome $Y_{it}(1)$ given $D_{it} = 1$ can be identified by $F_{Y_t|D_t=1}$, the observed distribution of outcomes for the treated group. But the other conditional distribution $F_{Y_t(0)|D_t=1}$, the counterfactual distribution of untreated potential outcomes for the treated group, cannot be identified generally from the sample. Thus, for identifying distributional features of treatment effects, we need to make additional restrictions.

As a measure of treatment effects, this chapter considers the identification and estimation of the Quantile Treatment Effect on the Treated (QTT) at $\tau \in \mathcal{T} \subset (0,1)$, defined by

$$\Delta^{QTT}(\tau) := F^{-1}_{Y_t(1)|D_t=1}(\tau) - F^{-1}_{Y_t(0)|D_t=1}(\tau),$$

where the τth quantile $F^{-1}_{Y_t(j)|D_t=1}(\tau)$ conditional on $D_t = 1$, given by

$$F^{-1}_{Y_t(j)|D_t=1}(\tau) := \inf\left\{y \in \mathbb{R} : F_{Y_t(j)|D_t=1}(y) \geq \tau\right\}, \quad j = 0, 1.$$

As discussed in the preceding paragraph, we can identify the distribution $F_{Y_t(1)|D_t=1}$ from observables. For identifying the QTT, it remains to establish the identification of the other distribution $F_{Y_t(0)|D_t=1}$. To this end, we need to make three additional restrictions.

The first condition restricts a time-difference of potential outcomes, $Y_{it}(0) - Y_{i,t-1}(0)$, without treatment such that its distribution does not dependent on the treatment status D_{it} in period t.

Assumption A2 (Distributional Difference in Differences).

$$\Pr\left\{\Delta Y_{it}(0) \leq \Delta y | D_{it} = 1\right\} = \Pr\left\{\Delta Y_{it}(0) \leq \Delta y | D_{it} = 0\right\},$$

for all $\Delta y \in \operatorname{supp}(\Delta Y_{it}(0))$, where $\Delta Y_{it}(0) := Y_{it}(0) - Y_{i,t-1}(0)$.

This assumption ensures that potential outcomes without treatment are comparable between treatment and control groups after taking a time-difference. An analogous condition employed under the mean DID framework is the “parallel trends” assumption:

$$E[\Delta Y_{it}(0)|D_{it} = 1] = E[\Delta Y_{it}(0)|D_{it} = 0],$$

which is necessary for identifying the average treatment effect on the treated (ATT). The distributional restriction in Assumption A2 replaces this standard mean restriction. If multiple pre-treatment periods in sample are available, then this assumption can be tested under a strict stationary assumption.

For the treated group, we can identify (i) the distribution $F_{Y_{t-1}(0)|D_t=1}$ of untreated potential outcomes in period $t-1$ from observed outcomes and (ii) the distribution $F_{\Delta Y_t(0)|D_t=1}$ of the change in untreated potential outcomes through Assumption A2 (the distributional DID). When these two distributions are identified, the average untreated potential outcome (and hence, the ATT) is identified. Without imposing Assumption A3 (the copula invariance), however, the QTT is not identified as many possible distributions of untreated potential outcomes in period t are observationally equivalent. For example, the distribution $F_{Y_t(0)|D_t=1}$ of untreated potential outcomes in period t will be highly unequal if the change in untreated potential outcomes and the initial untreated outcome are strongly positively dependent. On the other hand, the distribution of untreated potential outcomes in period t will be less unequal if the change and initial level of untreated potential outcomes are independent or negatively dependent.

The next condition imposes a restriction on the joint distribution $F_{\Delta Y_t(0),Y_{t-1}(0)|D_t}$ of $\Delta Y_{it}(0)$ and $Y_{i,t-1}(0)$ conditional on D_{it} through the conditional copula $C_{\Delta Y_t(0),Y_{t-1}(0)|D_t}$ of $\Delta Y_{it}(0)$ and $Y_{i,t-1}(0)$ given D_{it}. By Sklar's theorem, we have

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

for $(\Delta y, y) \in \text{supp}(\Delta Y_{it}(0), Y_{i,t-1}(0)|D_t)$. The following condition requires an invariance of the conditional copula with respect to the conditional variable D_{it}.

Assumption A3 (Copula Invariance).

$$C_{\Delta Y_t(0),Y_{t-1}(0)|D_t=1}(u,v) = C_{\Delta Y_t(0),Y_{t-1}(0)|D_t=0}(u,v),$$

for every $(u,v) \in [0,1]^2$.

Given a realized value of some random variable, the marginal distribution evaluated at this value can be interpreted as a ranking normalized to the unit interval. The conditional copula function captures some rank dependency between two variables $\Delta Y_{it}(0)$ and $Y_{i,t-1}(0)$ conditional on D_{it} and Assumption A3 requires that the dependency of ranking of these random variables $\Delta Y_{it}(0)$ and $Y_{i,t-1}(0)$ are the same for the treated and control groups. As in Assumption A1, however, this assumption does not rule out the possibility that the joint distribution of $\Delta Y_{it}(0)$ and $Y_{i,t-1}(0)$ conditional on D_{it} varies between the treatment and control group.

The Copula Invariance Assumption recovers the missing dependence required to uniquely identify the counterfactual distribution of untreated potential outcomes for the treated group. It does so by replacing the unknown copula for the treated group with the known copula from the untreated group. Intuitively, if, for example, we observe that observations in the control group at the top of the distribution of initial outcomes tend to experience the largest increases in outcomes over time, the Copula Invariance Assumption implies that this would also occur for the control group. The Distributional Difference in Differences Assumption implies that the distribution of the change in outcomes is the same for the two groups. But the initial distribution of outcomes can be different for the two groups.

As an additional identifying restriction, we assume continuity conditions on distributions of potential outcomes and its time-difference as below.

Assumption A4 (Continuous distributions). Each random variable of $\Delta Y_{it}(0)$ and $Y_{i,t-1}(0)$ has a continuous distribution conditional on $D_{it} = d$ for $d = 0,1$ and a random variable $Y_{it}(1)$ also has a continuous distribution conditional on $D_{it} = 1$. Each distribution has a compact support with densities uniformly bounded away from 0 and ∞ over the support.

The continuity of marginal distributions conditional on treatment status guarantees that conditional copulas in Assumption A3 are unique and facilitate the identification analysis. More precisely, we obtain identification by employing the Rosenblatt transform, which is the distribution transform studied by Rosenblatt (1952). Also Assumption A4 imposes a compact support assumption as in Athey and Imbens (2006) in order to avoid technical difficulties in the rest of analysis, while this condition is not used for our identification analysis and can be replaced by other conditions for the rest of the results.

Given the random sample in Assumption A1, the additional conditions in Assumption A2-A4 deliver the point-identification of the counterfactual distribution as in the following theorem.

THEOREM 13. *Suppose that Assumption A1-A4 hold. Then,*

$$F_{Y_t(0)|D_t=1}(y) = \Pr\left\{\Delta Y_{it} + F^{-1}_{Y_{t-1}|D_t=1} \circ F_{Y_{t-1}|D_t=0}(Y_{i,t-1}) \leq y | D_{it} = 0\right\},$$

for $y \in \operatorname{supp}(Y_{it}(0)|D_{it} = 1)$*, where* $\Delta Y_{it} := Y_{it} - Y_{i,t-1}$.

The above theorem shows that the counterfactual distribution of interest can be identified from observed outcomes of untreated individuals. This implies that that treated and untreated groups must be similar in the distributional sense of not only marginal distribution but also some dependency over periods, and thus Assumption A2 and A3 play a crucial role for the identification as shown in its proof. An immediate consequence is the identification of the QTT since the other distribution $F_{Y_t(1)|D_t=1}$ is identified by the distribution $F_{Y_t|D_t=1}$ of observed outcomes.

As an extension of the result above, we establish the identification of the counterfactual distribution when available data is two-periods repeated cross sections, rather than panel data. Even when the data generating process satisfies Assumption A1-A4, the change in outcome over time, $Y_{it} - Y_{i,t-1}$, is unobserved because each individual in sample is observed only at one period. To deal with the identification issue due to the data, we consider a restriction of the rank invariance, which enables us to recover individual outcome in period t by using the rank of outcome in period $t-1$ as formalized by the following theorem.

COROLLARY 1. *Consider the repeated cross sections* $\{(Y_{is}, D_{is})\}_{i=1}^{n^{(s)}}$ *in period s with* $n^{(s)}$ *being the sample size for* $s = t-1, t$. *Suppose that the data generating process for the repeated cross satisfy Assumption A1-A4 hold. Additionally if the conditional copula of* $(Y_{i,t-1}(0), Y_{i,t}(0))$ *given* $D_{it} = 1$ *satisfies the rank invariance over time:*

$$C_{Y_{t-1}(0), Y_t(0)|D_t=1}(u, v) = \min\{u, v\},$$

for $(u, v) \in [0, 1]^2$*, then,*

$$F_{Y_t(0)|D_t=1}(y) = \Pr\left\{(F^{-1}_{Y_t|D_t=0} + F^{-1}_{Y_{t-1}|D_t=1}) \circ F_{Y_{t-1}|D_t=0}(Y_{i,t-1}) - Y_{i,t-1} \leq y | D_{it} = 0\right\},$$

for $y \in \text{supp}(Y_{it}(0)|D_{it} = 1)$.

3.2.2 Identification with Covariates

In observational studies, treated and untreated groups may differ significantly, while their differences may become negligible or at least be mitigated after controlling for some observed characteristics. In this subsection, we consider identifying restrictions conditional on some covariates X_i, which may include all time-varying variables, such as X_{it} and $X_{i,t-1}$, and also time-invariant variables Z_i. By using our argument in the preceding section, we can establish the point-identification of the Conditional Quantile Treatment Effect on Treated (CQTT), given by

$$\Delta^{CQTT}(\tau, x) := F^{-1}_{Y_t(1)|X=x,D_t=1}(\tau) - F^{-1}_{Y_t(0)|X=x,D_t=1}(\tau),$$

for $\tau \in \mathcal{T}$, where $F^{-1}_{Y_t(j)|X=x,D_t=1}(\tau)$ denotes the τth conditional quantile of $Y_t(j)$ given $(X, D_t) = (x, 1)$ for $j = 0, 1$. For identifying the QTT, we extend our argument together with a propensity score re-weighting similar to the approach taken in Hirano, Imbens, and

Ridder (2003), Abadie (2005), and Firpo (2007).

To establish identification of CQTT and QTT while taking into account for observational difference across individuals, we make the conditional version of assumptions parallel to Assumption A1-A4 as follows.

Assumptions:

A1'. The data consists of observations $\{(Y_{i,t-1}, Y_{it}, D_{it}, X_i)\}_{i=1}^{n}$ satisfying the relation in (3.1) and the potential outcomes $\{(Y_{is}(0), Y_{is}(1)) : s = t-1, t\}$ and covariates X_i are cross-sectionally i.i.d. conditional on treatment status D_{it}.

A2'. The time-difference of potential outcome $\Delta Y_{it}(0)$ is independent of the treatment status D_{it} conditional on covariates X_i.

A3'. The conditional copula $C_{\Delta Y_t(0), Y_{t-1}(0)|X,D_t}$ satisfies

$$C_{\Delta Y_t(0), Y_{t-1}(0)|X,D_t=1}(u,v) = C_{\Delta Y_t(0), Y_{t-1}(0)|X,D_t=0}(u,v),$$

for every $(u,v) \in [0,1]^2$.

A4'. Random variables $\Delta Y_{it}(0)$ and $Y_{i,t-1}(0)$ respectively have continuous marginal distribution $F_{\Delta Y_t(0)|X,D_t}$ and $F_{Y_{t-1}(0)|X,D_t}$ conditional on (X_i, D_{it}) and a random variable $Y_{it}(1)$ also has a continuous distribution $F_{Y_t(1)|X,D_t}$ conditional on X_i and $D_{it} = 1$. Each distribution has a compact support with densities uniformly bounded away from 0 and ∞ over the support.

A5'. $p := \Pr\{D_{it} = 1\} > 0$ and $p(X_i) := \Pr\{D_{it} = 1|X_i\} < 1$.

Assumption A1' requires a random sample even with covariates. Assumption A2' and A3' impose the conditional version of the Distributional DID Assumption and Copula Invariance Assumption, respectively. Continuity of conditional distributions are assumed in Assumption A4'. Assumption A5' is used only for identifying the QTT and is a standard assumption in the literature requiring that the sample includes treated individuals and for each X, there are some individuals that are untreated.

For identifying the CQTT or QTT, it suffices to identify counterfactual distributions $F_{Y_t(0)|X,D_t=1}$ and $F_{Y_t(0)|D_t=1}$, respectively, because the conditional distributions $F_{Y_t(1)|X,D_t=1}$ and $F_{Y_t(1)|D_t=1}$ can be identified by the conditional distributions of observed outcome $F_{Y_t|X,D_t=1}$ and $F_{Y_t|D_t=1}$, respectively. The next theorem presents an identification result, which states that the counterfactual distribution $F_{Y_t(0)|X,D_t=1}$ of untreated potential outcomes $Y_{it}(0)$ conditional on X_i and $D_{it}=1$ can be recovered by utilizing distributions $F_{Y_{t-1}|X,D_t=1}$ and $F_{Y_{t-1}|X,D_t=0}$ of the observed outcome Y_{t-1} conditional on $(X_i, D_{it}=1)$ and $(X_i, D_{it}=0)$, respectively.

THEOREM 14. *Suppose that Assumption A1'-A4' hold. (a) Then, for* $y \in \operatorname{supp}(Y_{it}(0)|X, D_{it}=1)$,

$$F_{Y_t(0)|X,D_t=1}(y) = \Pr\left\{\Delta Y_{it} + F^{-1}_{Y_{t-1}|X,D_t=1} \circ F_{Y_{t-1}|X,D_t=0}(Y_{i,t-1}) \leq y | D_{it} = 0\right\}.$$

(b) Additionally, if Assumption A5' holds, then

$$F_{Y_t(0)|D_t=1}(y) = E\left[\omega(X_i) 1\{\Delta Y_{it} + F^{-1}_{Y_{t-1}|X,D_t=1} \circ F_{Y_{t-1}|X,D_t=0}(Y_{i,t-1}) \leq y\} \middle| D_{it} = 0\right],$$

for $y \in \operatorname{supp}(Y_{it}(0)|D_{it}=1)$, *where* $\omega(X_i) := (1-p)p(X_i)/\big(1-p(X_i)\big)p$

The identification result presented above is similar in spirit to the one in Theorem 13. The main differences are (i) it requires estimating conditional on covariates distributions and quantile functions and (ii) it requires estimating the propensity score in a first step and taking an expectation with weights based on the propensity score.

3.3 Estimation and Inference

As in the previous section, the counterfactual distribution is identified by empirical distributions of observed outcomes conditional on treatment status. In this section, we first explain an estimation procedure based on conditional empirical distributions and then provide asymptotic results for the proposed estimator using the functional delta method. We develop uniform inference results using techniques from the literature on empirical processes (see, for example, van der Vaart and Wellner, 1996).

3.3.1 Estimation

We estimate the conditional distribution $F_{Y_s|D_t=d}$ of observed outcome Y_{is} given treatment status $D_{it} = d$ by using the corresponding empirical distribution $\hat{F}_{Y_s|D_t=d}$, given by

$$\hat{F}_{Y_s|D_t=d}(y) := n_d^{-1} \sum_{i \in \mathcal{I}_d} 1\{Y_{is} \leq y\}.$$

We denote an estimator for $F_{Y_t(1)|D_t=1}(y)$ by $\hat{F}_{Y_t(1)|D_t=1}(y)$, which is given by the empirical distribution $\hat{F}_{Y_t|D_t=1}(y)$ because $F_{Y_t(1)|D_t=1}(y) = F_{Y_t|D_t=1}(y)$. For estimation of the counterfactual distribution provided in Theorem 13, we obtain estimated quantiles $\hat{F}^{-1}_{Y_{t-1}|D_t=1}$ from the empirical distribution $\hat{F}_{Y_{t-1}|D_t=1}$ and then set

$$\hat{F}_{Y_t(0)|D_t=1}(y) := n_0^{-1} \sum_{i \in \mathcal{I}_0} 1\{\Delta Y_{it} + \hat{F}^{-1}_{Y_{t-1}|D_t=1} \circ \hat{F}_{Y_{t-1}|D_t=0}(Y_{i,t-1}) \leq y\}, \tag{3.2}$$

for $y \in \mathbb{R}$. We use estimated distribution functions $\hat{F}_{Y_t(1)|D_t=1}$ and $\hat{F}_{Y_t(0)|D_t=1}$ to obtain quantiles for each distribution. Then, the estimator $\hat{\Delta}^{QTT}$ of the QTT is given by

$$\hat{\Delta}^{QTT}(\tau) := \hat{F}^{-1}_{Y_t(1)|D_t=1}(\tau) - \hat{F}^{-1}_{Y_t(1)|D_t=1}(\tau),$$

for $\tau \in \mathcal{T}$.

3.3.2 Asymptotic Results

We provide a functional central limit theorem for the QTT estimator over $\mathcal{T}$, where $\mathcal{T}$ is assumed to be a compact subset strictly within the unit interval. We begin with a preliminary result on weak convergence of empirical distributions, which facilitates the use of the functional delta method with Hadamard differentiable maps. In what follows, we denote by $\mathcal{Y}_{s|d} := \text{supp}(Y_{is}|D_{it} = d)$ and $\mathcal{Y}_{s|1}(j) := \text{supp}(Y_{is}(j)|D_{it} = 1)$ for $s = t, t-1$, $d = 0, 1$ and $j = 0, 1$.

We define empirical processes as

$$\hat{G}_{Y_s|D_t=d}(y) := \sqrt{n_d}\big(\hat{F}_{Y_s|D_t=d}(y) - F_{Y_s|D_t=d}(y)\big), \quad y \in \mathcal{Y}_{s|d},$$

for $(s,d) = (t-1,0), (t-1,1), (t,1)$. Let $\tilde{Y}_{it} := \Delta Y_{it} + F^{-1}_{Y_{t-1}|D_t=1} \circ F_{Y_{t-1}|D_t=0}(Y_{i,t-1})$ and define an empirical processes as

$$\tilde{G}_{Y_t(0)|D_t=1}(y) := \sqrt{n_0}\big(\tilde{F}_{Y_t(0)|D_t=1}(y) - F_{Y_t(0)|D_t=1}(y)\big), \quad y \in \mathcal{Y}_{t|1}(0),$$

where $\tilde{F}_{Y_t(0)|D_t=1}(y) := n_0^{-1} \sum_{i \in \mathcal{I}_0} 1\{\tilde{Y}_{it} \leq y\}$. We make an additional assumption.

Assumption A6. (a) A pair of random variables $(\Delta Y_{it}, Y_{i,t-1})$ is continuously distributed conditional on $D_t = 0$ over a compact support. with a distribution $F_{\Delta Y_t, Y_{t-1}|D_t=0}$ and a density $f_{\Delta Y_t, Y_{t-1}|D_t=0}$. A random variable ΔY_{it} is continuously distributed conditional on $Y_{i,t-1}$ and $D_t = 0$ with a uniformly continuous density $f_{\Delta Y_t|Y_{t-1}, D_t=0}$ over a compact support. (b) The sample sizes n_0 and n_1 go to ∞ as $n \to \infty$, while $r_j := \lim_{n\to\infty} \sqrt{n/n_j} \in [0, \infty)$ for $j = 0, 1$.

The following lemma provides a functional central limit theorem for the empirical processes above. We define $\mathbb{S} := \ell^\infty\big(\mathcal{Y}_{t|1}(0)\big) \times \ell^\infty(\mathcal{Y}_{t-1|0}) \times \ell^\infty(\mathcal{Y}_{t|1}) \times \ell^\infty(\mathcal{Y}_{t-1|1})$, where $\ell^\infty(S)$ denotes the space of all uniformly bounded functions on some set S, equipped with supremum norm $\|\cdot\|_\infty$.

LEMMA 27. *Suppose that Assumption A1-A6 hold. Then,*

$$\big(\tilde{G}_{Y_t(0)|D_t=1}, \hat{G}_{Y_{t-1}|D_t=0}, \hat{G}_{Y_t|D_t=1}, \hat{G}_{Y_{t-1}|D_t=1}\big) \rightsquigarrow \big(\mathbb{V}^{(0)}, \mathbb{W}^{(0)}, \mathbb{V}^{(1)}, \mathbb{W}^{(1)}\big),$$

in the space $\mathbb{S}$. *Here,* $\big(\mathbb{V}^{(0)}, \mathbb{W}^{(0)}, \mathbb{V}^{(1)}, \mathbb{W}^{(1)}\big)$ *is a tight Gaussian process with mean zero and covariance kernel* $\operatorname{diag}\{\Sigma_0(\cdot,\cdot), \Sigma_1(\cdot,\cdot)\}$ *defined on* $\mathbb{S}$, *where* $\Sigma_j(\cdot,\cdot)$ *is the* 2×2 *positive definite, covariance kernel of* $(\mathbb{V}^{(j)}, \mathbb{W}^{(j)})$ *for* $j=0,1$, *given by*

$$\Sigma_0(y_1,y_2) := \operatorname{Cov}_0\big(1\{\tilde{Y}_{it} \le y_1\}, 1\{Y_{i,t-1} \le y_2\}\big) \text{ and } \Sigma_1(y_3,y_4) := \operatorname{Cov}_1\big(1\{Y_{it} \le y_3\}, 1\{Y_{i,t-1} \le y_4\}\big),$$

with Cov_j *being the covariance function conditional on* $D_{it} = j$ *and* $(y_1,y_2,y_3,y_4) \in \mathbb{S}$.

Using the result in this lemma, we first obtain the joint limiting process for the estimator $(\hat{F}_{Y_t(0)|D_t=1}, \hat{F}_{Y_t(1)|D_t=1})$ of the potential outcome distributions. It is straightforward from the above result to obtain the limit process for the distribution $\hat{F}_{Y_t(1)|D_t=1}$, which is identified directly from data; whereas, the one for the counterfactual distribution $\hat{F}_{Y_t(0)|D_t=1}$ needs several steps. Since the estimator for the counterfactual distribution in (3.2) can be considered as a process indexed by functions depending on estimated distributions, we use recent results for empirical processes in van der Vaart and Wellner (2007) with some modifications in order to obtain the limiting process as formalized by the following proposition.

PROPOSITION 2. *For each* $j=0,1$, *define* $\hat{Z}_j(y) := \sqrt{n}\big(\hat{F}_{Y_t(j)|D_t=1}(y) - F_{Y_t(j)|D_t=1}(y)\big)$ *for* $y \in \mathcal{Y}_{t|1}(j)$. *Suppose that Assumption A1-A6 hold. Then,*

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

in the metric space $\ell^\infty\big(\mathcal{Y}_{t|1}(0)\big) \times \ell^\infty\big(\mathcal{Y}_{t|1}(1)\big)$. *Here,* $(\mathbb{Z}_0, \mathbb{Z}_1)$ *is a tight zero-mean Gaus-*

sian process with a.s. uniformly continuous paths on $\mathcal{Y}_{t|1}(0)\times\mathcal{Y}_{t|1}(1)$, *given by* $\mathbb{Z}^{(1)} = r_1\mathbb{V}^{(1)}$ *and* $\mathbb{Z}^{(0)} := r_0\mathbb{V}^{(0)} + \kappa(\mathbb{W}^{(0)}, \mathbb{W}^{(1)})$, *where the map* $\kappa : \ell^\infty(\mathcal{Y}_{t-1|0})\times\ell^\infty(\mathcal{Y}_{t-1|1}) \mapsto \ell^\infty(\mathcal{Y}_{t|1}(0))$ *is given by*

$$\kappa(W_0, W_1) := r_0 \int W_0(v)K(y,v)dv - r_1 \int W_1 \circ F^{-1}_{Y_{t-1}|D_t=1} \circ F_{Y_{t-1}|D_t=0}(v)K(y,v)dv,$$

for $(W_0, W_1) \in \ell^\infty(\mathcal{Y}_{t-1|0})\times\ell^\infty(\mathcal{Y}_{t-1|1})$ *with*

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

for $(y,v) \in \mathrm{supp}(\Delta Y_{it}, Y_{t-1}|D_{it} = 0)$.

This proposition shows that the limiting process $\mathbb{Z}^{(0)}$ for the counterfactual distribution has an extra term depending on the map κ, which reflects our identification argument of the counterfactual distribution of interest as well as the contribution of estimation errors from empirical distributions. Thus the limiting distribution is not nuisance parameter free, and a bootstrap procedure can facilitate statistical inference in practice as shown in the next subsection.

Next we present the limiting process of the QTT estimators over a range of quantile $\mathcal{T}$. Proposition 2 together with the functional delta method delivers the following theorem.

THEOREM 15. *Suppose that Assumption A1-A6 hold. If* $F_{Y_t(0)|D_t=1}$ *admits a positive continuous density* $f_{Y_t(0)|D_t=1}$ *on an interval* $[a,b]$ *containing an* ε*-enlargement of the set* $\{F^{-1}_{Y_t(0)|D_t=1}(\tau) : \tau \in \mathcal{T}\} \subset \mathcal{Y}_{t|1}(0)$ *with* $\mathcal{T} \subset (0,1)$, *then*

$$\sqrt{n}\big(\hat{\Delta}^{QTT}(\tau) - \Delta^{QTT}(\tau)\big) \rightsquigarrow \bar{\mathbb{Z}}^{(1)}(\tau) - \bar{\mathbb{Z}}^{(0)}(\tau),$$

where $\left(\bar{\mathbb{Z}}^{(0)}(\tau), \bar{\mathbb{Z}}^{(1)}(\tau)\right)$ *is a stochastic process in the metric space* $(\ell^\infty(\mathcal{T}))^2$*, given by*

$$\bar{\mathbb{Z}}^{(j)}(\tau) := \frac{\mathbb{Z}^{(j)}\left(F^{-1}_{Y_t(j)|D_t=1}(\tau)\right)}{f_{Y_t(j)|D_t=1}\left(F^{-1}_{Y_t(j)|D_t=1}(\tau)\right)},$$

for $j = 0, 1$.

We could also consider other plug-in estimators of Hadamard differentiable functionals, such as Lorenz curve and Gini coefficient using a similar argument to obtain the limit process.

3.3.3 Bootstrap

The limiting processes presented in the preceding section depend on unknown nuisance parameters, some of which require nonparametric estimation and may complicate inference in finite samples. To deal with the issue of nonpivotal limit processes, we consider a resampling method called the exchangeable bootstrap (see Præstgaard and Wellner (1993) and van der Vaart and Wellner (1996)). This resampling scheme consistently estimates limit laws of relevant empirical distributions and thus with the functional delta method consistently estimates the limit process of the QTT estimator.

For the resampling scheme, we introduce a vector of random weights $(W_{j1}, \ldots, W_{jn_j})$ for $j = 0, 1$. To establish the validity of the bootstrap, we assume that the random weights satisfy the following conditions.

Assumption B. For each $j \in \{0, 1\}$ and for each n_j, let $(W_{j1}, \ldots, W_{jn_j})$ be an n_j-dimensional vector of exchangeable, nonnegative random variables, The vectors $(W_{01}, \ldots, W_{0n_0})$ and $(W_{11}, \ldots, W_{1n_1})$ are independent of the original sample as well as each other. The vectors of random weights, depending on the size of each group, satisfy the following conditions:

$$\sup_{n_j} E|W_{j1}|^{2+\varepsilon} < \infty, \;\; \bar{W}_{n_j} := n_j^{-1} \sum_{i \in \mathcal{I}_j} W_{ji} \to^p 1, \;\; n_j^{-1} \sum_{i \in \mathcal{I}_j} (W_{ji} - \bar{W}_{n_j})^2 \to^p 1,$$

for $j = 0, 1$,

As van der Vaart and Wellner (1996) explain, this resampling scheme encompass a variety of bootstrap methods, such as the empirical bootstrap, subsampling, wild bootstrap and so on. This condition is employed in Chernozhukov, Fernández-Val, and Melly (2013) for inference of counterfactual distributions.

Given the random weights, we define the weighted bootstrap empirical distribution as

$$\hat{F}^*_{Y_s|D_t=d}(y) := n_d^{-1} \sum_{i \in \mathcal{I}_d} W_{di} 1\{Y_{is} \leq y\},$$

for $s = t-1, t$ and $d = 0, 1$. As in the previous subsection, the bootstrap distribution of the treated potential outcome $\hat{F}^*_{Y_t(1)|D_t=1}$ is given by $\hat{F}^*_{Y_t|D_t=1}$, while the bootstrap version of the counterfactual distribution is given by

$$\hat{F}^*_{Y_t(0)|D_t=1}(y) := n_0^{-1} \sum_{i \in \mathcal{I}_0} W_{0i} 1\{\Delta Y_{it} + \hat{F}^{*-1}_{Y_{t-1}|D_t=1} \circ \hat{F}^*_{Y_{t-1}|D_t=0}(Y_{i,t-1}) \leq y\},$$

for $y \in \mathbb{R}$, where $\hat{F}^{*-1}_{Y_{t-1}|D_t=1}$ is the bootstrap version of the quantile function obtained through the bootstrap empirical distribution $\hat{F}^*_{Y_{t-1}|D_t=1}$. The bootstrap version of the QTT process is given by

$$\hat{\Delta}^{QTT*}(\tau) := \hat{F}^{*-1}_{Y_t(1)|D_t=1}(\tau) - \hat{F}^{*-1}_{Y_t(1)|D_t=1}(\tau),$$

for $\tau \in \mathcal{T}$, where $\hat{F}^{*-1}_{Y_t(j)|D_t=1}(\tau)$ is the τth quantile of the bootstrap empirical distribution $\hat{F}^*_{Y_t(j)|D_t=1}$ of potential outcomes for $j = 0, 1$.

For the validity of the resampling method explained above, we need to introduce the notion of conditional weak convergence in probability, following van der Vaart and Wellner (1996). For some normed space $\mathbb{D}$, let $BL_1(\mathbb{D})$ denoted the space of all Lipschitz continuous functions from $\mathbb{D}$ to $[-1,1]$. Given the original sample D_n with n being the sample size, consider a random element $B_n^* := g(D_n, W_n)$ as a function of the original sample and the random weight vector W_n generating the bootstrap draw. The bootstrap law of B_n^* is said to consistently estimate the law of some tight random element B or $B_n \rightsquigarrow^p B$ if

$$\sup_{h \in BL_1(\mathbb{D})} \left| E_{W_n}[h(B_n^*)] - E[h(B)] \right| \to^p 0,$$

where E_{W_n} is the conditional expectation with respect to W_n given the original sample D_n.

To state a preliminary result, we define empirical processes indexed by $\mathcal{Y}_{s|d}$ as

$$\hat{G}^*_{Y_s|D_t=d} := \sqrt{n_d}\big(\hat{F}^*_{Y_s|D_t=d} - \hat{F}_{Y_s|D_t=d}\big),$$

for $(s,d) \in \{(t-1,0),(t-1,1)(t,1)\}$. Also, define an empirical process indexed by $\mathcal{Y}_{t|1}(0)$ as

$$\tilde{G}_{Y_t(0)|D_t=1} := \sqrt{n_0}\big(\tilde{F}^*_{Y_t(0)|D_t=1} - \tilde{F}_{Y_t(0)|D_t=1}\big),$$

where the empirical distribution is given by

$$\tilde{F}^*_{Y_t(0)|D_t=1}(y) := n_0^{-1} \sum_{i \in \mathcal{I}_0} W_{0i} 1\big\{\Delta Y_{it} + F^{-1}_{Y_{t-1}|D_t=1} \circ F_{Y_{t-1}|D_t=0}(Y_{i,t-1}) \leq y\big\}.$$

The following lemma shows that a set of the empirical processes defined above consistently estimates the tight random element defined in Lemma 27.

LEMMA 28. *Suppose that Assumption A1-A6 and B hold. Then,*

$$\big(\tilde{G}^*_{Y_t(0)|D_t=1}, \hat{G}^*_{Y_t|D_t=1}, \hat{G}^*_{Y_{t-1}|D_t=0}, \hat{G}^*_{Y_{t-1}|D_t=1}\big) \rightsquigarrow^p \big(\mathbb{V}^{(0)}, \mathbb{V}^{(1)}, \mathbb{W}^{(0)}, \mathbb{W}^{(1)}\big),$$

in $\mathbb{S}$, where the limit processes defined in Lemma 27.

Using this lemma, we first show that the exchangeable bootstrap provides a way to consistently estimate limit process of a pair of marginal distributions of potential outcomes. Subsequently we argue that the limit process of the QTT estimator can be estimated, using the functional delta method for a Hadmard differentiable map. The result is summarized in the following theorem.

THEOREM 16. *For each $j = 0, 1$, define $\hat{Z}_j^*(y) := \sqrt{n}\big(\hat{F}^*_{Y_t(j)|D_t=1}(y) - \hat{F}_{Y_t(j)|D_t=1}(y)\big)$ for $y \in \mathcal{Y}_{t|1}(j)$. Suppose that Assumption A1-A6 and B hold. Then, $(\hat{Z}_0^*, \hat{Z}_1^*) \rightsquigarrow^p (\mathbb{Z}_0, \mathbb{Z}_1)$, and thus the exchangeable bootstrap procedure consistently estimates the law of the limit stochastic process of the QTT:*

$$\sqrt{n}\big(\hat{\Delta}^{QTT*}(\tau) - \hat{\Delta}^{QTT}(\tau)\big) \rightsquigarrow^p \bar{\mathbb{Z}}^{(1)}(\tau) - \bar{\mathbb{Z}}^{(0)}(\tau).$$

Throughout this subsection, we have considered the validity of the exchangeable bootstrap in the case where the identifying assumptions hold without conditioning on covariates. As discussed in Section 3.1, however, when key assumptions in this chapter hold conditional on some covariates, the above estimation and bootstrap results can be extended by using conditional distributions given the covariates. Although additional conditions are necessary to deal with the curse of dimensionality, we can take an approach based on non-parametric methods in the spirit of Chernozhukov, Fernández-Val, and Melly (2013).

3.4 Empirical Application

This section considers the effect of job training on the quantiles of earnings. We use a dataset from LaLonde (1986) containing (i) a group of job training applicants that were randomly assigned to the training program, (ii) a group of job training applicants that were randomly assigned out of the training program, and (iii) an observational control group from the Panel Study of Income Dynamics (PSID). This setup allows us to use our method on the treated group and observational control group and compare those results to those obtained from the experimental data. The same dataset and approach have been used to estimate the QTT in Firpo (2007) and Callaway and Li (2015). This dataset has also been used to study the average effect of job training in LaLonde (1986), Heckman and Hotz (1989), Dehejia and Wahba (1999), and Smith and Todd (2005) among others.

3.4.1 Data

The data comes from the National Supported Work Demonstration in the mid-1970s. The program provided extensive training to workers who were unemployed or worked very few hours prior to the program. We use the all-male subset considered in Dehejia and Wahba (1999). More detailed descriptions of the data can be found in Hollister, Kemper, and Maynard (1984), LaLonde (1986), and Smith and Todd (2005).

In the dataset, there are 185 individuals that are randomly assigned to participate in the job training program; there are 260 that are randomly assigned to the control group; and there are 2490 individuals in the PSID sample. The outcome of interest in real annual earnings which are observed in 1974, 1975, and 1978. Individuals participate in the job training program between 1975 and 1978 which implies that earnings in 1974 and 1975 are pre-treatment; earnings in 1978 are post-treatment.

We also observe some characteristics about each individual: age, years of education, race, marital status, and unemployment status in 1974 and 1975. Summary Statistics are provided in Table 3.1. Noticeably, the experimental group – both those randomly assigned to treatment and those randomly assigned to the untreated group – have quite different average outcomes in each period and quite different observable characterstics relative to the PSID control group. In 1975, the period before treatment, average earnings for the treated and untreated experimental groups are close to $1500. But, for the PSID sample, average earnings are $19,100. Members of the experimental group are younger, less educated, much more likely to be black or Hispanic, much less likely to be married, much less likely to have a high school degree, and much more likely to be unemployed before the job training program. For the treated group, average earnings increase from \$1500 to over \$6400 between 1975 and 1978. But, it appears unlikely that all of that increase is due to the job training program. Average earnings for the experimental control group increase from \$1300 to \$4600 over the same period. And average earnings for the observational control group increase from \$19,100 to \$21,600.

3.4.2 Results

The theoretical identification results rely on the outcome of interest being continuously distributed. But there is a mass of workers that have 0 earnings in each period particularly for the experimental group. For this reason, we focus on estimating the QTT only for $\tau \geq .25$ where .25 corresponds roughly to the fraction of individuals that participate in the job training program with 0 earnings in the final period.

Panel Results

Figure 3.1 shows the estimates of the QTT using our method with a 90% confidence interval. For $\tau \in [0.25, 0.8]$, the estimates of the QTT are mostly flat with estimates of the QTT between roughly \$500 and \$1850 though most estimates are very close to \$1000. In the same range for τ, the estimates of the QTT are borderline statistically significant. Most interestingly, for this range of τ, the estimates of the QTT are remarkably close to the estimate of the QTT from the experimental data. However, for $\tau > .8$, the estimates of the QTT diverge from the experimental results. With the experimental data, the QTT continues to increase as τ increases. Our estimates of the QTT decrease and become negative for larger values of τ. Only for $\tau = .95$ does the estimated QTT from the experimental data fall outside of the 90% confidence interval for our estimate of the QTT.

Figure 3.2 compares point estimates of the QTT using the current method to point estimates using the three period method of Callaway and Li (2015). The problem of observations with 0 earnings is more severe for the three period method. For this reason, we follow Callaway and Li (2015) and only estimate the model using for $\tau \in [.75, .95]$. Our estimates are very similar to the unconditional estimates using the three period method. Both move further away from the experimental results as τ increases. For the three period method holding conditionally on covariates, the estimates share a similar shape to the experimental results but tend to be larger.

Repeated Cross Section Results

Figure 3.3 provides point estimates of the QTT with the experimental data, using our panel data method, and using our method for repeated cross sectional data. The repeated cross sections estimates come from the panel data but throw away the information on individual's identities. Most notably from this figure, the repeated cross section estimates are remarkably similar to the estimates when panel data is available. It appears that, in this case, the assumption of rank invariance in untreated potential outcomes over time does not have much effect on the resulting estimates.

Finally, Figure 3.4 compares the results of our repeated cross sections estimator to other methods available with repeated cross sections – Quantile DID (QDID) and Change in Changes (CIC). Our method performs better than these alternative methods. For QDID, the estimates of the QTT are too low at $\tau = .25$, but they quickly increase and are too large for $\tau > .5$. For example, for $\tau = .75$, the QDID estimate of the QTT is roughly \$4800 while the experimental QTT is roughly \$2400. The results are worse for the CIC model. For $\tau > .4$, the estimates of the QTT are too large. For example, for $\tau = .75$, the CIC estimate of the QTT is roughly \$9600 while the experimental QTT is roughly \$2400.

3.5 Conclusion

This chapter has considered identifying the Quantile Treatment Effect on the Treated under a Distributional DID Assumption when only two periods of data – panel or repeated cross sections – are available. We have developed uniform confidence intervals for the QTT and shown the validity of a bootstrap procedure for computing the confidence interval. Finally, we used our method to estimate the QTT of a job training program on earnings and found that the results of using our method compare favorably to results coming from existing methods.

Methodologically, the key innovation is to recover the unknown dependence between the change and initial level of untreated potential outcomes for the treated group from the observed dependence from the untreated group. Combining this condition with a distributional extension of the most common mean DID assumption results in point identification of the counterfactual distribution of untreated potential potential outcomes for the treated group; and, therefore, to identification of the QTT. There are many examples in finance, auction models, and duration models where identification depends on an unknown copula. The idea of replacing an unknown copula with one observed for another group may prove to be a fruitful line of research in those cases.

In Appendix, we use $\|\cdot\|$ to denote the Euclidean norm for vectors.

Proof of Theorem 13. For notational simplicity in the proof, we suppress a subscript i for an individual, which does not affect the result because of random sampling conditional on treatment status under Assumption A1. We can write $F_{Y_t(0)|D_t=1}(y) = \Pr\{\Delta Y_t(0) + Y_{t-1}(0) \leq y|D_t = 1\}$ for every $y \in \mathrm{supp}(Y_t(0)|D_t = 1)$. Define $U_d := F_{\Delta Y_t(0)|D_t=d}(\Delta Y_t(0))$ and $V_d := F_{Y_t(0)|D_t=d}(Y_t(0))$ for $d = 0,1$. Then under Assumption A4, we have

$$\Delta Y_t(0) = F^{-1}_{\Delta Y_t(0)|D_t=d}(U_d) \text{ and } Y_{t-1}(0) = F^{-1}_{Y_{t-1}(0)|D_t=d}(V_d), \tag{3.3}$$

almost surely, see Rosenblatt (1952). It follows that

$$F_{Y_t(0)|D_t=1}(y) = \Pr\left\{F^{-1}_{\Delta Y_t(0)|D_t=1}(U_1) + F^{-1}_{Y_{t-1}(0)|D_t=1}(V_1) \leq y \middle| D_t = 1\right\}.$$

For each $d = 0,1$, the joint distribution of (U_d, V_d) conditional on $D_t = d$ is given by a conditional copula $C_{\Delta Y_t(0),Y_{t-1}(0)|D_t=d}$, which is invariant with respect to the conditional variable D_t under Assumption A3. This implies that the conditional distribution of (U_1, V_1) given $D_t = 1$ is the same as that of (U_0, V_0) given $D_t = 0$. Thus we have

$$F_{Y_t(0)|D_t=1}(y) = \Pr\left\{F^{-1}_{\Delta Y_t(0)|D_t=1}(U_0) + F^{-1}_{Y_{t-1}(0)|D_t=1}(V_0) \leq y \middle| D_t = 0\right\}. \tag{3.4}$$

Under Assumption A2, $F^{-1}_{\Delta Y_t(0)|D_t=1}(\cdot) = F^{-1}_{\Delta Y_t(0)|D_t=0}(\cdot)$, which with (3.3) implies that

$$F^{-1}_{\Delta Y_t(0)|D_t=1}(U_0) = \Delta Y_t(0) \text{ and } F^{-1}_{Y_{t-1}(0)|D_t=1}(V_0) = F^{-1}_{Y_{t-1}(0)|D_t=1} \circ F_{Y_{t-1}(0)|D_t=0}(Y_t(0)),$$

almost surely. This together with (3.4) yields the desired result. $\square$

Proof of Corollary 1. Given that the data generating process satisfies Assumption A1-A4, the result in Theorem 13 holds and we have

$$F_{Y_t(0)|D_t=1}(y) = \Pr\left\{Y_{it}(0) - Y_{i,t-1}(0) + F^{-1}_{Y_{t-1}|D_t=1} \circ F_{Y_{t-1}|D_t=0}(Y_{i,t-1}) \leq y|D_{it} = 0\right\}, \quad (3.5)$$

for $y \in \text{supp}(Y_{it}(0)|D_{it} = 1)$. Because of the repeated cross section, we cannot identify the term $Y_{it}(0) - Y_{i,t-1}(0)$ from the observed outcomes of the untreated group. Under the rank invariance assumption, however, we have

$$F_{Y_t(0)|D_t=0}(Y_{it}(0)) = F_{Y_{t-1}(0)|D_t=0}(Y_{i,t-1}(0)),$$

where the distributions $F_{Y_t(0)|D_t=0}$ and $F_{Y_{t-1}(0)|D_t=0}$ of potential outcomes, can be identified by the distributions $F_{Y_t|D_t=0}$ and $F_{Y_{t-1}|D_t=0}$ of observed outcomes, respectively. It follows that

$$Y_{it}(0) = F^{-1}_{Y_t|D_t=0} \circ F_{Y_{t-1}|D_t=0}(Y_{i,t-1}).$$

This together with (3.5) leads to the desired result. □

Proof of Theorem 14. For the same reason as in Theorem 13, we suppress a subscript *i*.

(a) Applying a similar argument used to show Theorem 13 with conditional variables, we can show that, under Assumption A1'-A4',

$$F_{Y_t(0)|X,D_t=1}(y) = \Pr\left\{\tilde{Y}_t \leq y \middle| X, D_t = 0\right\}, \quad (3.6)$$

where $\tilde{Y}_t := \Delta Y_t + F^{-1}_{Y_{t-1}|X,D_t=1} \circ F_{Y_{t-1}|X,D_t=0}(Y_{t-1})$.

(b) By Bayes' theorem, we have

$$\frac{dF_{X|D=1}}{dF_{X|D=0}} \cdot \frac{p}{1-p} = \frac{p(X)}{1-p(X)}.$$

This together with (3.6) implies that

$$F_{Y_t(0)|D_t=1}(y) = E\left[\frac{1-p}{p}\frac{p(X)}{1-p(X)}E\left[1\left\{\tilde{Y}_t \leq y\right\}|X, D_t = 0\right]\Big|D_t = 0\right].$$

Hence the desired result follows from the law of iterated expectation. $\square$

To derive the limiting distribution of the estimator for the QTT, we present two technical lemmas concerning the Hadamard differentiability. We introduce a setup and notations used in the these lemmas. Let $F_0 := (G_0, H_0)$ with G_0 and H_0 being distribution functions having a compact support $\mathcal{V} \subset \mathbb{R}$ and a density function g_0 and h_0, respectively. Consider a pair of continuous random variables (V_1, V_2) taking values on $\mathcal{V}\times\mathcal{V}$ with the joint distribution $F_{V_1V_2}$ having a density $f_{V_1V_2}$ as well as the marginal distributions F_{V_j} having a density f_{V_j} for $j = 0, 1$. We suppose that the conditional distribution $F_{V_1|V_2}$ has a continuous density function $f_{V_1|V_2}$ uniformly bounded away from 0 and ∞.

LEMMA 29. *Let $\mathbb{D} := (C(\mathcal{V}))^2$ and define the map $\psi : \mathbb{D}_\psi \subset \mathbb{D} \mapsto \ell^\infty(\mathcal{V})$, given by*

$$\psi(F) := G^{-1} \circ H,$$

for $F := (G, H) \in \mathbb{D}_\psi$, where $\mathbb{D}_\psi := \mathbb{E}\times\mathbb{E}$ with $\mathbb{E}$ denoting the set of all distributions functions having a strictly positive, bounded density. Then, the map ψ is Hadamard differentiable at F_0 tangentially to $\mathbb{D}$. Its derivative at F_0 in $\gamma := (\gamma_1, \gamma_2) \in \mathbb{D}$ is given by

$$\psi'_{F_0}(\gamma) = \frac{\gamma_2 - \gamma_1 \circ G_0^{-1} \circ H_0}{g_0 \circ G_0^{-1} \circ H_0}.$$

Proof. To prove the assertion, we first represent ψ as a composition map. Let $\mathbb{D}_{\psi_2} := \mathbb{E}^{-} \times C(\mathcal{V})$, where $\mathbb{E}^{-}$ denotes the set of generalized inverse of all functions in $\mathbb{E}$. Define the maps $\psi_1 : \mathbb{D}_{\psi} \mapsto \mathbb{D}_{\psi_2}$ and $\psi_2 : \mathbb{D}_{\psi_2}, \mapsto \ell^{\infty}(\mathcal{V})$, given by

$$\psi_1(\Gamma) := (\Gamma_1^{-1}, \Gamma_2) \quad \text{and} \quad \psi_2(\Lambda) := \Lambda_1 \circ \Lambda_2,$$

for $\Gamma := (\Gamma_1, \Gamma_2) \in \mathbb{D}_{\psi}$ and $\Lambda := (\Lambda_1, \Lambda_2) \in \mathbb{D}_{\psi_2}$. Then we can write $\psi = \psi_2 \circ \psi_1$.

For the map ψ_1, Lemma 3.9.23(ii) of van der Vaart and Wellner (1996) implies that if Γ has a derivative denoted by Γ', then the map ψ_1 is Hadamard differentiable at Γ tangentially to $\mathbb{D}$. Its derivative at Γ in $\gamma := (\gamma_1, \gamma_2) \in \mathbb{D}$ is given by

$$\psi'_{1,\Gamma}(\gamma) := \big(-(\gamma_1/\Gamma'_1) \circ \Gamma_1^{-1}, \gamma_2 \big).$$

In terms of the map ψ_2, Lemma 3.9.27 of van der Vaart and Wellner (1996) implies that ψ_2 is Hadamard differentiable at Λ tangentially to $C([0,1]) \times \ell^{\infty}(\mathcal{V})$. Its derivative at Λ in $\lambda := (\lambda_1, \lambda_2) \in C([0,1]) \times \ell^{\infty}(\mathcal{V})$ is given by

$$\psi'_{2,\Lambda}(\lambda) := \lambda_1 \circ \Lambda_2 + \Lambda'_{1,\Lambda_2} \lambda_2.$$

Lemma 3.9.3 of van der Vaart and Wellner (1996) with Hadamard derivatives of the maps ψ_1 and ψ_2 yields that $\psi'_{F_0}(\gamma) = \psi'_{2,(G_0^{-1}, H_0)} \circ \psi'_{1,F_0}(\gamma)$ for $\gamma \in \mathbb{D}$, where

$$\psi'_{1,F_0}(\gamma) = \big(-(\gamma_1/g_0) \circ G_0^{-1}, \gamma_2 \big),$$

and

$$\psi'_{2,(G_0^{-1}, H_0)}(\lambda) = \lambda_1 \circ H_0 + \frac{\lambda_2}{g_0 \circ G_0^{-1} \circ H_0},$$

because $\partial G_0^{-1}(\tau)/\partial \tau = 1/g_0 \circ G_0^{-1}(\tau)$. Hence the desired result follows. $\square$

LEMMA 30. *Let $\mathbb{D} := (C(\mathcal{V}))^2$ and let $\mathcal{W}$ be a compact subset of $\mathbb{R}$. Define the map $\phi : \mathbb{D}_\phi \subset \mathbb{D} \mapsto \ell^\infty(\mathcal{W})$, given by*

$$\phi(F)(w) := \Pr\{V_1 + G^{-1} \circ H(V_2) \le w\},$$

for $F := (G,H) \in \mathbb{D}_\phi$ and for $w \in \mathcal{W}$, where $\mathbb{D}_\psi := \mathbb{E} \times \mathbb{E}$ with $\mathbb{E}$ being the set of all distributions functions having a strictly positive, bounded density. Then, the map ϕ is Hadamard differentiable at F_0 tangentially to $\mathbb{D}$. Its derivative at F_0 in $\gamma := (\gamma_1, \gamma_2) \in \mathbb{D}$ is given by

$$\phi'_{F_0}(\gamma)(w) := \int \big(\gamma_2(v_2) - \gamma_1 \circ G_0^{-1} \circ H_0(v_2)\big) \frac{f_{V_1 V_2}(w - G_0^{-1} \circ H_0(v_2), v_2)}{g_0 \circ G_0^{-1} \circ H_0(v_2)} dv_2.$$

Proof. To prove the assertion, we represent ϕ as a composition map. Define $\psi : \mathbb{D}_\phi \to \mathbb{D}_\pi$ as in the proceeding lemma, where $\mathbb{D}_\pi$ denotes the set of all functions $F^{-1} \circ G$ for $(F^{-1}, G) \in \mathbb{E}^- \times \mathbb{E}$ with $\mathbb{E}^-$ and $\mathbb{E}$ defined in the proof of the proceeding lemma. Define the map $\pi : \mathbb{D}_\pi \mapsto \ell^\infty(\mathcal{W})$, given by

$$\pi(\Xi)(w) := \int F_{V_1|V_2}\big(w - \Xi(v_2)|v_2\big) dF_{V_2}(v_2),$$

for $w \in \mathcal{W}$. Since we can write $\phi(F)(w) = \int F_{V_1|V_2}\big(w - G^{-1} \circ H(v_2)|v_2\big) dF_{V_2}(v_2)$ for $F \in \mathbb{D}$ and $w \in \mathcal{W}$, we can show that $\phi = \pi \circ \psi$.

We wish to show that π has a Hadamard derivative at $\Xi \in \mathbb{D}_\pi$ tangentially to $\mathbb{D}$ with derivative at Ξ in $\xi \in \mathbb{D}$

$$\pi'_\Xi(\xi)(w) = \int \xi(v_2) f_{V_1|V_2}(w - \Xi(v_2)|v_2) dF_{V_2}(v_2). \tag{3.7}$$

Consider any sequence $t_k > 0$ and $\Xi_k \in \mathbb{D}_\pi$ for $k \in \mathbb{N}$ such that $t_k \searrow 0$ and $\xi_k := (\Xi_k -$

$\Xi)/t_k \to \xi$ in $\mathbb{D}$ as $k \to \infty$. We have

$$F_{V_1|V_2}(w - \Xi_k(v_2)|v_2) - F_{V_1|V_2}(w - \Xi(v_2)|v_2) = t_k \xi_k(v_2) \int_0^1 f_{V_1|V_2}(w - \Xi(v_2) - rt_k \xi_k(v_2)|v_2)dr.$$

It follows that

$$\begin{aligned} \frac{\pi(\Xi_k) - \pi(\Xi)}{t_k} - \pi'_\Xi(\xi) &= \int \big(\xi_k(v_2) - \xi(v_2)\big) f_{V_1|V_2}(\cdot - \Xi(v_2)|v_2) dF_{V_2}(v_2) \\ &\quad + \int \xi_k(v_2) D_k(\cdot, v_2) dF_{v_2}(v_2), \end{aligned}$$

where $D_k(w, v_2) := \int_0^1 \left\{ f_{V_1|V_2}(w - \Xi(v_2) - rt_k \xi_k(v_2)|v_2) - f_{V_1|V_2}(w - \Xi(v_2)|v_2) \right\} dr$. Since $f_{V_1|V_2}$ is uniformly continuous and ξ_k is uniformly bounded, $\lim_{k\to\infty} \|D_k\|_{\mathcal{W}\times\mathcal{V}} = 0$ and thus the second term on the above display converges to 0 as $k \to \infty$. Also the first term on the above display converges to zero because $f_{V_2|V_1}$ is uniformly bounded and $\|\xi_k - \xi\|_\infty \to 0$ as $k \to \infty$. Thus the map π has the Hadamard derivative as stated.

Lemma 3.9.3 of van der Vaart and Wellner (1996) shows that $\phi'_{F_0}(\gamma) = \pi'_{G_0^{-1}\circ H_0} \circ \psi'_{F_0}(h)$, which together with the Hadamard derivative of π in (3.7) and the one of ψ in Lemma 29 yields

$$\phi'_{F_0}(\gamma)(w) = \int \frac{\gamma_2(v_2) - \gamma_1 \circ G_0^{-1} \circ H_0(v_2)}{g_0 \circ G_0^{-1} \circ H_0(v_2)} f_{V_1|V_2}\big(w - G_0^{-1} \circ H_0(v_2)\big|v_2\big) dF_{V_2}(v_2).$$

Hence the desired result follows. $\square$

Define $\bar{V} := V_1 + G_0^{-1} \circ H_0(V_2)$. We additionally assume that $\bar{V}$ is distributed over a compact space $\mathcal{V}$ with a distribution $F_{\bar{V}}$ and a continuous density $f_{\bar{V}}$ uniformly bounded away from 0 and ∞. We consider random sample $\{(V_{1i}, V_{2i})\}_{i=1}^n$ of n independent copies of (V_1, V_2) and let $\bar{V}_i := V_{1i} + G_0^{-1} \circ H_0(V_{2i})$. We set $F_n := (G_n, H_n)$ to denote a random element of $(\ell^\infty(\mathcal{V}))^2$ as a consistent estimator for F_0. For $F = (G, H) \in (C(\mathcal{V}))^2$ and $w \in \mathcal{W}$, define a functional taking values at $\ell^\infty(\mathcal{W})$:

$$\phi_n(F)(w) := n^{-1} \sum_{i=1}^{n} 1\{V_{1i} + G^{-1} \circ H(V_{2i}) \leq w\},$$

and the empirical process indexed by $F \in (\ell^\infty(\mathcal{V}))^2$:

$$\nu_n(F) := \sqrt{n}\big(\phi_n(F) - \phi(F)\big).$$

The lemma is proven, along the line of Theorem 2.3 of van der Vaart and Wellner (2007) with some modification.

LEMMA 31. *Suppose that $\sqrt{n}(F_n - F_0)$ converges in distribution to a tight, random element with values in $(\ell^\infty(\mathcal{V}))^2$. Then,*

$$\sup_{w \in \mathcal{W}} \big| \nu_n(F_n)(w) - \nu_n(F_0)(w) \big| = o_p(1).$$

Proof. Because the set $\mathcal{W}$ is compact, it suffices to show that $|\nu_n(F_n) - \nu_n(F_0)|(w) = o_p(1)$ for every $w \in \mathcal{W}$. Let $w \in \mathcal{W}$ and $\varepsilon > 0$ be fixed. Suppose that $\sqrt{n}(F_n - F_0)$ converges in distribution to a tight random element. Then, by the functional delta method and Hadamard differentiability, Lemma 29 implies that

$$\xi_n := \sqrt{n}\big(G_n^{-1} \circ H_n - G_0^{-1} \circ H_0\big) \rightsquigarrow Z, \tag{3.8}$$

in $\ell^\infty(\mathcal{V})$ for some tight random element Z. Then there exists a compact set $S \subset \ell^\infty(\mathcal{V})$ such that $\Pr\{Z \notin S\} \leq \varepsilon/2$, and also $\limsup_{n\to\infty} \Pr\{\xi_n \notin S^{\delta/2}\} \leq \varepsilon/2$ for any $\delta > 0$, where $S^{\delta/2}$ is the $\delta/2$-enlargement set of S. Because S is compact, there exist a finite set $\{\mu_j\}_{j=1}^J \subset S$ with $J = J(\delta)$ such that $\sup_{\mu \in S} \min_{1\leq j\leq J} \|\mu - \mu_j\|_\infty < \delta$. It follows that, for any $\delta > 0$,

$$\Pr\Big\{\min_{1\leq j\leq J} \|\xi_n - \mu_j\|_\infty \geq \delta\Big\} \leq \Pr\big\{\xi_n \notin S^{\delta/2}\big\} \leq \varepsilon/2, \tag{3.9}$$

for a sufficiently large n. In the view of the compactness of $\mathcal{V}$, for every $\eta > 0$, there is a finite set $\{v_k\}_{k=1}^K \subset \mathcal{V}$ with $K = K(\eta)$ such that $\sup_{v\in\mathcal{V}} \min_{1\le k\le K} |v - v_k| < \eta$. Define the map $\Pi_\delta : \mathcal{V} \mapsto \{v_k\}_{k=1}^K$ such that $|v - \Pi_\eta(v)| \le \eta$ for every $v \in \mathcal{V}$. By Theorem 1.5.7 of van der Vaart and Wellner (1996) with (3.8), there exists $\eta > 0$ such that

$$\lim_{n\to\infty} \Pr\big\{\|\xi_n - \xi_n \circ \Pi_\eta\|_\infty > \varepsilon/2\big\} < \varepsilon/2. \tag{3.10}$$

It follows from (3.9) and (3.10) that

$$\limsup_{n\to\infty} \Pr\Big\{\min_{1\le j\le J} \|\xi_n - \mu_j \circ \Pi_\eta\|_\infty > \varepsilon\Big\} < \varepsilon,$$

which yields that, given a set $\mathcal{M}_{j,\eta}(\delta) := \{\mu \in \ell^\infty(\mathcal{V}) : \|\mu - \mu_j \circ \Pi_\eta\|_\infty \le \delta\}$, we have

$$\Pr\Big\{\big|\{\nu_n(F_n) - \nu_n(F_0)\}(w)\big| \ge \varepsilon\Big\} \le \varepsilon + \Pr\Big\{\max_{1\le j\le J} \sup_{\mu\in\mathcal{M}_{j,\eta}(\delta)} \big|\{\nu_n(F_0 + n^{-1/2}\mu) - \nu_n(F_0)\}(w)\big| \ge \varepsilon\Big\}.$$

Since J is finite, it suffices to show that, for every $j = 1, \ldots, J$,

$$\sup_{\mu\in\mathcal{M}_{j,\eta}(\delta)} \big|\{\nu_n(F_0 + n^{-1/2}\mu) - \nu_n(F_0)\}(w)\big| = o_p(1).$$

An application of the triangle inequality yields that, for every $\mu \in \mathcal{M}_{j,\eta}(\delta)$,

$$\begin{aligned}\sup_{\mu\in\mathcal{M}_{j,\eta}(\delta)} \big|\{\nu_n(F_0 + n^{-1/2}\mu) - \nu_n(F_0)\}(w)\big| \quad &\le \quad \sup_{\mu\in\mathcal{M}_{j,\eta}(\delta)} \big|\{\nu_n(F_0 + n^{-1/2}\mu) - \nu_n(F_0 + n^{-1/2}\mu_j)\}(w)\big| \\ &\quad + \big|\{\nu_n(F_0 + n^{-1/2}\mu_j) - \nu_n(F_0)\}(w)\big|.\end{aligned}$$

Let $\bar{\mu}_{jk} := |\mu_j \circ \Pi_\eta(v_k)|$ for $j = 1, \ldots, J$ and $k = 1, \ldots, K$. For a collection of functions

$$\Big\{1\big\{\bar{V}_i + n^{-1/2}\mu(V_{2i}) \le w\big\} - 1\big\{\bar{V}_i + n^{-1/2}\mu_j \circ \Pi_\eta(V_{2i}) \le w\big\} : \mu \in \mathcal{M}_{j,\eta}(\delta)\Big\},$$

we can form an envelop function $I_{i,j}^{(1)}(\delta) := 1\big\{\max_{1\le k\le K} |\bar{V}_i + n^{-1/2}\bar{\mu}_{jk} - w| \le n^{-1/2}\delta\big\}$.

We can write

$$\sup_{\mu \in \mathcal{M}_{j,\eta}(\delta)} \left| \{ \nu_n(F_0 + n^{-1/2}\mu) - \nu_n(F_0 + n^{-1/2}\mu_j) \}(w) \right| \leq n^{-1/2} \sum_{i=1}^{n} I_{i,j}^{(1)}(\delta) + \sqrt{n} E[I_{1,j}^{(1)}(\delta)] \quad (3.11)$$

The second term on the right-hand of (3.11) become arbitrarily small for a sufficiently small $\delta > 0$, because we have

$$\sqrt{n} E[I_{1,j}^{(1)}(\delta)] \leq \sqrt{n} \max_{1 \leq k \leq K} \int_{-n^{-1/2}\delta}^{n^{-1/2}\delta} f_{\bar{V}}(s - \bar{\mu}_{jk} + w) ds \leq \delta C_1, \quad (3.12)$$

for some constant C_1. Applying the Markov inequality for the first term on the right-hand of (3.11), we obtain

$$\Pr\left\{ n^{-1/2} \sum_{i=1}^{n} I_{i,j}(\delta) \geq \varepsilon \right\} \leq \varepsilon^{-1} \sqrt{n} E[I_{1,j}(\delta)],$$

where the right-hand side becomes arbitrarily small for a sufficiently small δ due to (3.12). It follows that the right-hand side of (3.11) converges to 0 in probability for a sufficiently small δ.

We have

$$\left| 1\left\{ \bar{V}_i + n^{-1/2} \mu_j \circ \Pi_\eta(V_{2i}) \leq w \right\} - 1\left\{ \bar{V}_i \leq w \right\} \right| \leq I_{i,j}^{(2)}(\delta)$$

for every $i = 1, \ldots, n$, where $I_{i,j}^{(2)}(\delta) := 1\left\{ |\bar{V}_i - w| \leq n^{-1/2} \max_{1 \leq k \leq K} \bar{\mu}_{jk} \right\}$. Using the Markov inequality, we can show that

$$\Pr\left\{ \left| \{ \nu_n(F_0 + n^{-1/2}\mu_j) - \nu_n(F_0) \}(w) \right| \geq \varepsilon \right\} \leq \varepsilon^{-2} E[I_{i,j}^{(2)}(\delta)],$$

where the right-hand side goes to zero as $n \to \infty$ because $E\left[I_{i,j}^{(2)}(\delta) \right] \leq C_2 n^{-1/2} \max_{1 \leq k \leq K} \bar{\mu}_{jk}$ for some constant C_2. Hence the proof is completed. □

LEMMA 32. *Suppose that $\sqrt{n}(F_n - F_0)$ converges in distribution to a tight, random element in $\ell^\infty(\mathcal{V})$. Then,*

$$\sqrt{n}\big(\phi_n(F_n) - \phi(F_0)\big) = \nu_n(F_0) + \phi'_{F_0}\big(\sqrt{n}(F_n - F_0)\big) + o_p(1),$$

where ϕ'_{F_0} is the Hadamard derivative given in Lemma 30.

Proof. By definition, we can write

$$\sqrt{n}\big(\phi_n(F_n) - \phi(F_0)\big) = \nu_n(F_n) + \sqrt{n}\big(\phi(F_n) - \phi(F_0)\big).$$

First Lemma 31 show that

$$\nu_n(F_n) = \nu_n(F_0) + o_p(1).$$

uniformly in $w \in \mathcal{W}$. Since the map ϕ is Hadmard differentiable, the functional delta method in Theorem 3.9.4 of van der Vaart and Wellner (1996) with Lemma 30 implies that

$$\sqrt{n}\big(\phi(F_n) - \phi(F_0)\big) = \phi'_{F_0}\big(\sqrt{n}(F_n - F_0)\big) + o_p(1).$$

Hence the desired result follows. □

Proof of Lemma 27. The result follows from the functional central limit theorem for empirical distribution functions. See Chapter 2 of van der Vaart and Wellner (1996) for instance. □

Proof of Proposition 2. We can show that $\sqrt{n}\big((\hat{F}_{Y_{t-1}|D_t=0}, \hat{F}_{Y_{t-1}|D_t=1}) - (F_{Y_{t-1}|D_t=0}, F_{Y_{t-1}|D_t=1})\big)$

has a tight limit asymptotically. It follows from Lemma 32 that

$$\sqrt{n}\big(\hat{F}_{Y_t(0)|D_t=1} - F_{Y_t(0)|D_t=1}\big) = r_0 \tilde{G}_{Y_t(0)|D_t=0} + \kappa\big(\hat{G}_{Y_{t-1}|D_t=0}, \hat{G}_{Y_{t-1}|D_t=1}\big) + o_p(1),$$

uniformly in $y \in \mathrm{supp} Y_t(0)|D_t = 1$. Hence the extended continuous mapping theorem with Lemma 27 yields the desired result. □

Proof of Theorem 15. When $\hat{F}_{Y_t(j)|D_t=1}(y)$ is weakly increasing in y, we can show that the corresponding quantile function $\hat{F}^{-1}_{Y_t(j)|D_t=1}(\tau)$ is Hadamard differentiable. It follows from the functional delta method that

$$\sqrt{n}\big(\hat{F}^{-1}_{Y_t(j)|D_t=1}(\tau) - F^{-1}_{Y_t(j)|D_t=1}(\tau)\big) \rightsquigarrow \big(\mathbb{Z}(j)/f_{Y_t(j)|D_t=1}\big) \circ F^{-1}_{Y_t(j)|D_t=1}(\tau),$$

as a stochastic process indexed by $\tau \in \mathcal{T}$ and $j \in \{0,1\}$. Hence the desired result holds. □

We now prove a technical lemma, which is a bootstrap version of Lemma 32.

LEMMA 33. *Suppose that the assumption in Lemma 32 and also assume that $\sqrt{n}(F_n^* - F_0)$ converges in distribution to a tight random element unconditional on the original sample. Then,*

$$\sqrt{n}\big(\phi_n^*(F_n^*) - \phi_n(F_n)\big) = \sqrt{n}\big(\phi_n^*(F_0) - \phi_n(F_0)\big) + \phi'_{F_0}\big(\sqrt{n}(F_n^* - F_n)\big) + o_p(1),$$

Proof. Let $\tilde{\nu}_n^*(F) := \sqrt{n}\big(\phi_n^*(F) - \phi(F)\big)$ for $F \in (\ell^\infty(\mathcal{V}))^2$. We have

$$\sqrt{n}\big(\phi_n^*(F_n^*) - \phi(F_0)\big) = \tilde{\nu}_n^*(F_n) + \sqrt{n}\big(\phi(F_n^*) - \phi(F_0)\big).$$

By a similar argument used to prove Lemma 31, we can show that, uniformly in $w \in \mathcal{W}$,

$$\tilde{\nu}_n^*(F_n) = \tilde{\nu}_n^*(F_0) + o_p(1). \tag{3.13}$$

Also ϕ is the Hadamard differentiable function ϕ and $\sqrt{n}(F_n^* - F_0)$ converges in distribution to a tight random element unconditional on the original sample. Thus the functional delta method implies that

$$\sqrt{n}\big(\phi(F_n^*) - \phi(F_0)\big) = \phi'_{F_0}\big(\sqrt{n}(F_n^* - F_0)\big) + o_p(1). \tag{3.14}$$

It follows from (3.13) and (3.14) that

$$\sqrt{n}\big(\phi_n^*(F_n^*) - \phi(F_0)\big) = \tilde{\nu}_n^*(F_0) + \phi'_{F_0}\big(\sqrt{n}(F_n^* - F_0)\big) + o_p(1),$$

which together with Lemma 32 yields the desired result because ϕ'_{F_0} is a linear map. □

Proof of Lemma 28. The result follows from Theorem 3.6.13 of van der Vaart and Wellner (1996). Thus we omit the detail. □

Proof of Theorem 16. First we wish to show that $\hat{Z}^* \rightsquigarrow^p \mathbb{Z}$, where $\hat{Z}^* := (\hat{Z}_0^*, \hat{Z}_1^*)'$ and $\mathbb{Z} := (\mathbb{Z}_0, \mathbb{Z}_1)'$. By the triangle inequality, we obtain

$$\sup_{h \in BL_1} \big|E_M[h(\hat{Z}^*)] - E[h(\mathbb{Z})]\big| \leq \sup_{h \in BL_1} \big|E_M[h(\hat{Z}^*)] - E_M[h(\tilde{Z}^*)]\big| \tag{3.15}$$

$$+ \sup_{h \in BL_1} \big|E_M[h(\tilde{Z}^*)] - E[h(\mathbb{Z})]\big|, \tag{3.16}$$

where $\tilde{Z}^* := \big(r_0 \tilde{G}^*_{Y_t(0)|D_t=0} + \kappa(\hat{G}^*_{Y_{t-1}|D_t=0}, \hat{G}^*_{Y_{t-1}|D_t=1}), \hat{Z}_1^*\big)'$. It suffices to show that (3.15) and (3.16) converge in probability to zero, separately.

We consider (3.15). Because $\left|E_M[h(\hat{Z}^*)] - E_M[h(\tilde{Z}^*)]\right| \leq E_M\left|h(\hat{Z}^*) - h(\tilde{Z}^*)\right|$, we have

$$E\Big[\sup_{h\in BL_1}\left|E_M[h(\hat{Z}^*)] - E_M[h(\tilde{Z}^*)]\right|\Big] \leq E\Big[\sup_{h\in BL_1}\left|h(\hat{Z}^*) - h(\tilde{Z}^*)\right|\Big]. \tag{3.17}$$

Let $\varepsilon > 0$ be fixed and define $I^*_{n,\varepsilon} := 1\{\|\hat{Z}^* - \tilde{Z}^*\|_\infty > \varepsilon\}$. Lemma 28 and 33 imply that $\lim_{n\to\infty} E[I^*_{n,\varepsilon}] \leq \varepsilon$, while $\sup_{h\in BL_1}\left|h(\hat{Z}^*) - h(\tilde{Z}^*)\right| \leq 2$. It follows that

$$E\Big[\sup_{h\in BL_1}\left|h(\hat{Z}^*) - h(\tilde{Z}^*)\right|\cdot I^*_{n,\varepsilon}\Big] \leq 2\varepsilon. \tag{3.18}$$

Also we can show that

$$E\Big[\sup_{h\in BL_1}\left|h(\hat{Z}^*) - h(\tilde{Z}^*)\right|\cdot (1 - I^*_{n,\varepsilon})\Big] \leq \varepsilon, \tag{3.19}$$

because $\sup_{h\in BL_1}\left|h(\hat{Z}^*) - h(\tilde{Z}^*)\right| \leq \|\hat{Z}^*(y) - \tilde{Z}^*(y)\|_\infty$. It follows from (3.18) and (3.19) that the right-hand side of (3.17) is bounded by 3ε. Since ε is arbitrary, an application of the Markov inequality yields the convergence of (3.15) to 0 in probability.

Consider (3.16). Using Lemma 28 together with the continuous mapping theorem, we can show that (3.16) converges to 0 in probability. Hence we obtain the desired result.

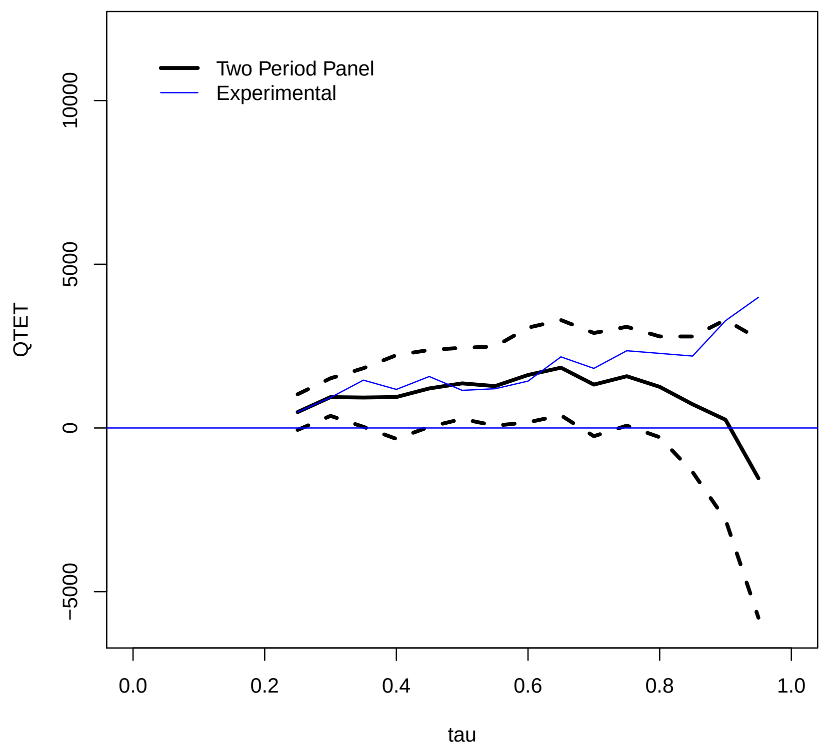
We now consider validity of exchangeable bootstrap for the QTT. Theorem 3.9.11 of van der Vaart and Wellner (1996) shows that the functional delta method can apply for Hadamard differentiable maps under resampling. Since the map from distribution to quantile is Hadamard differentiable, the desired result follows. $\square$

Table 3.1: Summary Statistics

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

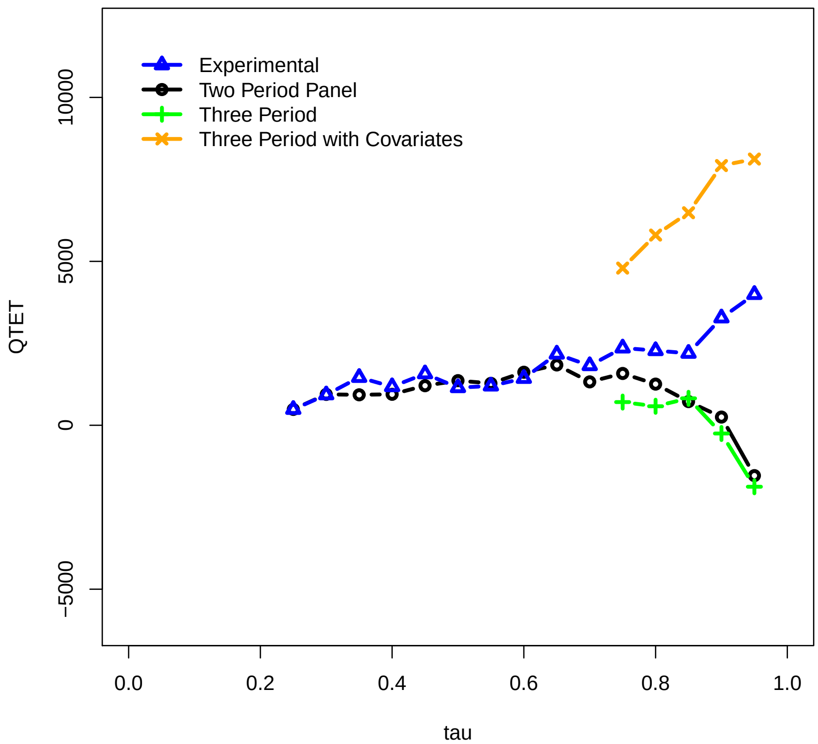
RE are real earnings in thousands of dollars. The Treated and Randomized statistics come from the experimental data. The observational statistics come from the PSID control group. “nd” is the normalized difference for each variable between the Treated group and the Randomized/Observational group.

Figure 3.1: QTT Estimates



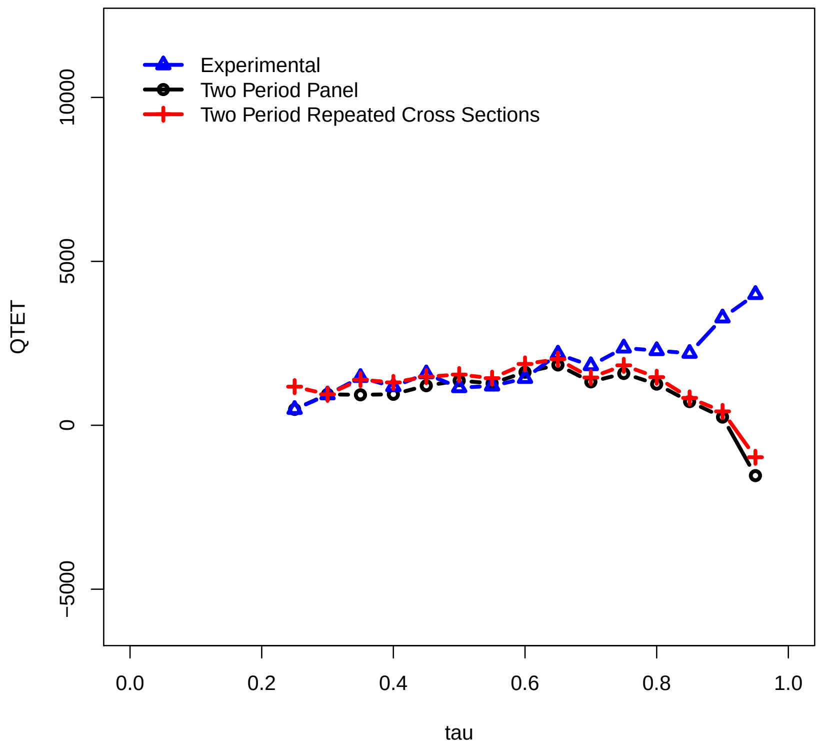
The QTT is estimated using the two-period, panel method of the current chapter. The dashed line is the 90% confidence interval for the QTT. The blue line is the QTT estimated using the experimental data.

Figure 3.2: QTT Point Estimates using Two-Period and Three-Period Model



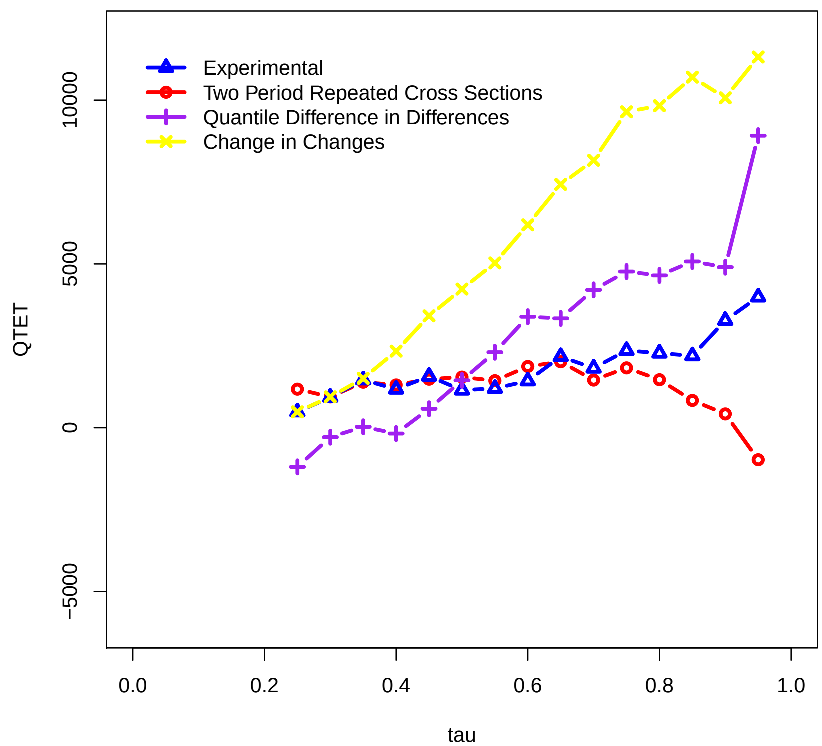
The figure provides estimates of the QTT using the method in the current chapter, using the experimental data, and using the the three period method of Callaway and Li (2015) under an unconditional Distributional Difference in Differences Assumption (green line) and under a conditional Distributional Difference in Differences Assumption (orange line).

Figure 3.3: QTT Estimates with Panel Data and Repeated Cross Sections Data



The figure provides estimates of the QTT using panel data and using repeated cross sections data and compares these estimates to the estimated QTT from the experimental data. The estimates using repeated cross sections data come from the same dataset as the panel, but "throw away" the panel structure of the dataset.

Figure 3.4: QTT Estimates from QDID and CIC



The figure provides estimates of the QTT coming from the Quantile Difference in Differences approach and Change in Changes (Athey and Imbens, 2006) approach and compares these to estimates coming from the repeated cross sections data estimates using our approach and the estimates of the QTT using the experimental data.

BIBLIOGRAPHY

[1] Abadie, Alberto. "Semiparametric instrumental variable estimation of treatment response models". *Journal of econometrics* 113.2 (2003), pp. 231–263.

[2] Abadie, Alberto. "Semiparametric Difference-in-Differences Estimators". *The Review of Economic Studies* 72.1 (2005), pp. 1–19.

[3] 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.

[4] Abbring, Jaap and James Heckman. "Econometric Evaluation of Social Programs, Part III: Distributional Treatment Effects, Dynamic Treatment Effects, Dynamic Discrete Choice, and General Equilibrium Policy Evaluation". *Handbook of Econometrics* 6 (2007), pp. 5145–5303.

[5] 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.

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

[7] Amemiya, Takeshi. "Regression Analysis when the Dependent Variable is Truncated Normal". *Econometrica* 41.6 (Nov. 1973), pp. 997–1016. URL: https://ideas.repec.org/a/ecm/emetrp/v41y1973i6p997-1016.html.

[8] Amemiya, Takeshi. *Advanced econometrics*. Harvard university press, 1985.

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

[10] Athey, Susan and Guido Imbens. "Identification and Inference in Nonlinear Difference-in-Differences Models". *Econometrica* 74.2 (2006), pp. 431–497.

[11] Bertrand, Marianne, Esther Duflo, and Sendhil Mullainathan. "How Much Should We Trust Differences-in-Differences Estimates?" *The Quarterly Journal of Economics* 119.1 (2004), pp. 249–275.

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

[13] Bitler, Marianne P, Jonah B Gelbach, and Hilary W Hoynes. "Welfare Reform and Children's Living Arrangements". *Journal of Human Resources* 41.1 (2006), pp. 1–27.

[14] Bitler, Marianne P, Jonah B Gelbach, and Hilary W Hoynes. "Distributional Impacts of the Self-Sufficiency Project". *Journal of Public Economics* 92.3 (2008), pp. 748–765.

[15] Bitler, Marianne P, Hilary W Hoynes, and Thurston Domina. “Experimental evidence on distributional effects of head start”. Worker Paper, National Bureau of Economic Research. 2014.

[16] Bonhomme, Stéphane and Ulrich Sauder. “Recovering Distributions in Difference-in-Differences models: A Comparison of Selective and Comprehensive Schooling”. *Review of Economics and Statistics* 93.2 (2011), pp. 479–494.

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

[18] Callaway, Brantly and Tong Li. “Quantile Treatment Effects in Difference in Differences Models with Panel Data” (2015).

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

[20] Card, David. “The impact of the Mariel boatlift on the Miami labor market”. *Industrial & Labor Relations Review* 43.2 (1990), pp. 245–257.

[21] 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.

[22] 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.

[23] Carneiro, Pedro, Karsten Hansen, and James Heckman. “Estimating distributions of treatment effects with an application to the returns to schooling and measurement of the effects of uncertainty on college choice”. *International Economic Review* 44.2 (2003), pp. 361–422.

[24] 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.

[25] Carroll, Robert, David Joulfaian, and Mark Rider. “Income Mobility: The Recent American Experience”. Working Paper, Andrew Young School of Policy Studies Research Paper Series. 2007.

[26] Chamberlain, Gary. “Multivariate Regression Models for Panel Data”. *Journal of Econometrics* 18.1 (1982), pp. 5–46.

[27] Chamberlain, Gary. “Panel Data”. *Handbook of Econometrics* 2 (1984), pp. 1248–1318.

[28] Chan, Sewin and Ann Huff Stevens. “Employment and retirement following a late-career job loss”. *American Economic Review* (1999), pp. 211–216.

[29] Chan, Sewin and Ann Huff Stevens. “The Effects of Job Loss on Older Workers: Employment, Earnings and Wealth”. *Ensuring Health and Income Security for an Aging Workforce* (2001), pp. 189–211.

[30] Chan, Sewin and Ann Huff Stevens. “How Does Job Loss Affect the Timing of Retirement?” *The BE Journal of Economic Analysis & Policy* 3.1 (2004), pp. 1–26.

[31] Chen, Heng. “Within-group Estimators for Fixed Effects Quantile Models with Large N and Large T”. Working Paper, Bank of Canada. 2015.

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

[33] Chernozhukov, Victor, Iván Fernández-Val, and Blaise Melly. “Inference on counterfactual distributions”. *Econometrica* 81.6 (2013), pp. 2205–2268.

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

[35] Chetty, Raj, Nathaniel Hendren, Patrick Kline, and Emmanuel Saez. “Where is the land of Opportunity? The Geography of Intergenerational Mobility in the United States”. *The Quarterly Journal of Economics* 129.4 (2014), pp. 1553–1623.

[36] Couch, Kenneth A. “Late life job displacement”. *The Gerontologist* 38.1 (1998), pp. 7–17.

[37] Couch, Kenneth A and Dana W Placzek. “Earnings losses of displaced workers revisited”. *The American Economic Review* (2010), pp. 572–589.

[38] Davis, Steven J and Till Von Wachter. “Recessions and the Costs of Job Loss”. *Brookings Papers on Economic Activity* 43.2 (Fall) (2011), pp. 1–72.

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

[40] Djebbari, Habiba and Jeffrey Smith. “Heterogeneous Impacts in PROGRESA”. *Journal of Econometrics* 145.1 (2008), pp. 64–80.

[41] Doksum, Kjell. “Empirical probability plots and statistical inference for nonlinear models in the two-sample case”. *The annals of statistics* (1974), pp. 267–277.

[42] Dong, Yingying and Shu Shen. “Testing for Rank Invariance or Similarity in Program Evaluation: The Effect of Training on Earnings Revisited”. Working Paper. 2015.

[43] Eren, Ozkan and Serkan Ozbeklik. “Who Benefits from Job Corps? A Distributional Analysis of an Active Labor Market Program”. *Journal of Applied Econometrics* 29.4 (2014), pp. 586–611.

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

[45] Fan, Jianqing and Irene Gijbels. *Local Polynomial Modelling and its Applications: Monographs on Statistics and Applied Probability 66*. Vol. 66. CRC Press, 1996.

[46] Fan, Yanqin, Emmanuel Guerre, and Dongming Zhu. “Partial Identification and Confidence Sets for Functionals of the Joint Distribution of Potential Outcomes” (2015).

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

[48] Fan, Yanqin and Sang Soo Park. "Sharp bounds on the distribution of treatment effects and their statistical inference". *Econometric Theory* 26.03 (2010), pp. 931–951.

[49] Fan, Yanqin and Sang Soo Park. "Confidence intervals for the quantile of treatment effects in randomized experiments". *Journal of Econometrics* 167.2 (2012), pp. 330–344.

[50] 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.

[51] Farber, Henry S. "The Changing Face of Job Loss in the United States, 1981-1995". *Brookings Papers on Economic Activity. Microeconomics* (1997), pp. 55–142.

[52] Farber, Henry S. "Job Loss in the Great Recession and its Aftermath: U.S. Evidence from the Displaced Workers Survey". Working Paper. 2015.

[53] Firpo, Sergio. "Efficient Semiparametric Estimation of Quantile Treatment Effects". *Econometrica* 75.1 (2007), pp. 259–276.

[54] Firpo, Sergio, Nicole M Fortin, and Thomas Lemieux. "Unconditional Quantile Regressions". *Econometrica* (2009), pp. 953–973.

[55] Firpo, Sergio and Geert Ridder. "Bounds on functionals of the distribution of treatment effects" (2008).

[56] Frandsen, Brigham R and Lars J Lefgren. "A Test for Rank Similarity and Partial Identification of the Distribution of Treatment Effects". Working Paper. 2015.

[57] Frank, Maurice, Roger Nelsen, and Berthold Schweizer. "Best-possible bounds for the distribution of a sum-a problem of Kolmogorov". *Probability Theory and Related Fields* 74.2 (1987), pp. 199–211.

[58] Fréchet, Maurice. "Sur Les Tableaux de Correlation Dont Les Marges Sont Donnees". *Annals University Lyon: Series A* 14 (1951), pp. 53–77.

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

[60] Gibbons, Robert and Lawrence F Katz. "Layoffs and Lemons". *Journal of Labor Economics* 9 (1991), pp. 351–380.

[61] Gottschalk, Peter. "Inequality, Income Growth, and Mobility: The Basic Facts". *The Journal of Economic Perspectives* (1997), pp. 21–40.

[62] 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.

[63] Gronau, R. “Wage Comparisons-A Selectivity Bias”. *Journal of Political Economy* 82 (1974), pp. 1119–1143. URL: http://ideas.repec.org/a/ecm/emetrp/v42y1974i4p679-94.html.

[64] Hall, Peter, Rodney CL Wolff, and Qiwei Yao. “Methods for estimating a conditional distribution function”. *Journal of the American Statistical Association* 94.445 (1999), pp. 154–163.

[65] Hansen, Bruce E. “Nonparametric estimation of smooth conditional distributions”. Working Paper: Department of Economics, University of Wisconsin. 2004.

[66] Health and Retirement Study. *RAND HRS Data file public use dataset*. Produced and distributed by the University of Michigan with funding from the National Institute on Aging (grant number NIA U01AG009740). Ann Arbor, MI, 2014.

[67] Heckman, James J. “Shadow Prices, Market Wages, and Labor Supply”. *Econometrica* 42.4 (July 1974), pp. 679–94. URL: http://ideas.repec.org/a/ecm/emetrp/v42y1974i4p679-94.html.

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

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

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

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

[72] 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.

[73] Heckman, James and Jeffrey Smith. “Evaluating the Welfare State James J. Heckman and Jeffrey Smith”. *Econometrics and Economic Theory in the 20th Century: The Ragnar Frisch Centennial Symposium*. 31. Cambridge University Press. 1998, pp. 241–318.

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

[75] 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.

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

[77] 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.

[78] 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.

[79] Hoeffding, Wassily. "Massstabinvariante korrelationstheorie". *Schriften des Mathematischen Instituts und des Institutes Fur Angewandte Mathematik der Universitat Berlin* 5 (1940), pp. 179–233.

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

[81] Hungerford, Thomas L. "US Income Mobility in the Seventies and Eighties". *Review of Income and Wealth* 39.4 (1993), pp. 403–417.

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

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

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

[85] Jacobson, Louis S, Robert J LaLonde, and Daniel G Sullivan. "Earnings losses of displaced workers". *The American Economic Review* (1993), pp. 685–709.

[86] Joe, Harry. *Multivariate models and multivariate dependence concepts*. CRC Press, 1997.

[87] Joe, Harry. *Dependence modeling with copulas. No. 134 in monographs on statistics and applied probability*. CRC Press, Boca Raton, FL, 2015.

[88] Kim, Ju Hyun. "Identifying the Distribution of Treatment Effects under Support Restrictions". Working Paper. 2014.

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

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

[91] Koenker, Roger. *Quantile Regression*. 38. Cambridge university press, 2005.

[92] Kopczuk, Wojciech, Emmanuel Saez, and Jae Song. "Earnings inequality and mobility in the United States: evidence from social security data since 1937". *The Quarterly Journal of Economics* 125.1 (2010), pp. 91–128.

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

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

[95] Lee, Alan. *U-statistics. Theory and Practice*. Marcel Dekker, Inc., New York, 1990.

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

[97] Li, Qi and Jeffrey Scott Racine. *Nonparametric Econometrics: Theory and Practice*. Princeton University Press, 2007.

[98] Madrian, Brigitte C and Lars John Lefgren. "An Approach to Longitudinally Matching Current Population Survey (CPS) Respondents". *Journal of Economic and Social Measurement* 26 (2000), pp. 31–62.

[99] Makarov, GD. "Estimates for the Distribution Function of a Sum of Two Random Variables When the Marginal Distributions are Fixed". *Theory of Probability & its Applications* 26.4 (1982), pp. 803–806.

[100] Manski, Charles. "Monotone Treatment Response". *Econometrica* (1997), pp. 1311–1334.

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

[102] 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.

[103] Mundlak, Yair. "On the Pooling of Time Series and Cross Section Data". *Econometrica: Journal of the Econometric Society* (1978), pp. 69–85.

[104] Nelsen, Roger. *An introduction to copulas*. Springer, 2007.

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

[106] Præstgaard, Jens and Jon A Wellner. "Exchangeably weighted bootstraps of the general empirical process". *The Annals of Probability* (1993), pp. 2053–2086.

[107] Riddell, W Craig and Xueda Song. "The Impact of Education on Unemployment Incidence and Re-employment Success: Evidence from the US Labour Market". *Labour Economics* 18.4 (2011), pp. 453–463.

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

[109] Rosenblatt, Murray. "Remarks on a multivariate transformation". *The Annals of Mathematical Statistics* (1952), pp. 470–472.

[110] Rüschendorf, Ludger. “Random variables with maximum sums”. *Advances in Applied Probability* (1982), pp. 623–632.

[111] Sen, Amartya. *On economic inequality*. Clarendon Press, 1997.

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

[113] Smith, Jeffrey and Petra Todd. “Does matching overcome LaLonde’s critique of nonexperimental estimators?” *Journal of econometrics* 125.1 (2005), pp. 305–353.

[114] Stevens, Ann Huff. “Persistent Effects of Job Displacement: The Importance of Multiple Job Losses”. *Journal of Labor Economics* (1997), pp. 165–188.

[115] Thuysbaert, Bram. “Distributional Comparisons in Difference in Differences Models”. Working Paper, University of Leuven. 2007.

[116] U.S. Bureau of Labor Statistics. *All Employees: Total Nonfarm Payrolls [PAYEMS]*. Retrieved from FRED, Federal Reserve Bank of St. Louis on October 6, 2015. 2015.

[117] U.S. Bureau of Labor Statistics. *Civilian Unemployment Rate [UNRATE]*. Retrieved from FRED, Federal Reserve Bank of St. Louis on October 6, 2015. 2015.

[118] U.S. Bureau of Labor Statistics. *Unemployment Rate: 55 years and over [LNS14024230]*. Retrieved from FRED, Federal Reserve Bank of St. Louis on October 6, 2015. 2015.

[119] Van der Vaart, Aad. *Asymptotic statistics*. Vol. 3. Cambridge university press, 2000.

[120] van der Vaart, Aad W. and Jon A. Wellner. *Weak Convergence and Empirical Processes with Applications to Statistics*. Springer, 1996.

[121] van der Vaart, Aad W. and Jon A. Wellner. “Empirical processes indexed by estimated functions.” English. *Asymptotics: particles, processes and inverse problems. Festschrift for Piet Groeneboom.* Institute of Mathematical Statistics, 2007, pp. 234–252. ISBN: 978-0-940600-71-3/pbk.

[122] Von Wachter, Till, Jae Song, and Joyce Manchester. “Long-term earnings losses due to mass layoffs during the 1982 recession: An analysis using US administrative data from 1974 to 2004”. *unpublished; New York: Columbia University* (2009).

[123] Williamson, Robert and Tom Downs. “Probabilistic Arithmetic. I. Numerical Methods for Calculating Convolutions and Dependency Bounds”. *International Journal of Approximate Reasoning* 4 (1990), pp. 89–158.