# Bounds on Distributional Treatment Effect Parameters using Panel Data with an Application on Job Displacement*

Brantly Callaway†

This Version: August 2019
First Version: October 2015

**Abstract**

This paper develops new techniques to bound distributional treatment effect parameters that depend on the joint distribution of potential outcomes – an object not identified by standard identifying assumptions such as selection on observables or even when treatment is randomly assigned. I show that panel data and an additional assumption on the dependence between untreated potential outcomes for the treated group over time (i) provide more identifying power for distributional treatment effect parameters than existing bounds and (ii) provide a more plausible set of conditions than existing methods that obtain point identification. I apply these bounds to study heterogeneity in the effect of job displacement during the Great Recession. Using standard techniques, I find that workers who were displaced during the Great Recession lost on average 34% of their earnings relative to their counterfactual earnings had they not been displaced. Using the methods developed in the current paper, I also show that the average effect masks substantial heterogeneity across workers.

*Keywords:* Joint Distribution of Potential Outcomes, Distribution of the Treatment Effect, Quantile of the Treatment Effect, Copula Stability Assumption, Panel Data, Job Displacement

*JEL Codes:* C14, C31, C33, J63

*This is an updated version of my job market paper "Job Displacement during the Great Recession: Tight Bounds on Distributional Treatment Effect Parameters using Panel Data." I would like to thank Alper Arslan, Afrouz Azadikhah-Jahromi, Colin Cameron, Bill Collins, Scott Cunningham, Xu Cheng, Dakshina De Silva, Frank Diebold, Frank DiTraglia, Juan Carlos Escanciano, Dalia Ghanem, Collin Green, Federico Gutierrez, Jorgen Hansen, Atsushi Inoue, Catherine Maclean, Pedro Sant'Anna, Shu Shen, Artyom Shneyerov, Doug Webber, and seminar participants at Baylor University, the Bureau of Labor Statistics, Concordia University, the Consumer Financial Protection Bureau, Indiana University, Lancaster University, Mathematica Policy Research, Temple University, the University of California-Davis, the University of California-Merced, and the University of Pennsylvania for helpful comments. I especially thank my advisor Tong Li for his detailed comments and support on this project. Code for the methods developed in the paper is available in the `R` package `csabounds`.

†Department of Economics, University of Mississippi, 225 Holman Hall, University, MS 38677. Email: bmcallaw@olemiss.edu

# 1 Introduction

One of the key contributions of the modern treatment effects literature is to explicitly acknowledge that the effect of participating in a treatment can differ across different individuals, even individuals with identical observable characteristics (see, for example, Heckman, Smith, and Clements (1997), Heckman (2001), and Imbens and Wooldridge (2009)). Despite allowing for heterogeneous effects, most work in the treatment effects literature focuses on summary measures like the average treatment effect (ATE) or average treatment effect for the treated (ATT). This paper considers bounds on the distribution of *individual-level* treatment effects in the case where a researcher has access to panel data. Individual-level treatment effect heterogeneity poses a particularly difficult challenge. Unlike summary parameters such as the ATE, which only depend on the marginal distributions of treated and untreated "potential" outcomes, the distribution of the treatment effect depends on the *joint* distribution of treated and untreated potential outcomes.

The 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 treated and untreated potential outcomes are identified, the copula – which "couples" the marginal distributions into the joint distribution and captures the dependence between the marginal distributions – is not identified. The fundamental reason why the joint distribution of treated and untreated potential outcomes is not identified is that, for each individual, either a treated potential outcome or an untreated potential outcome is observed (but not both).

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.

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 (2010), and Fan and Park (2012)). In the case of job displacement, these bounds rule out that the effect of job displacement is the same across all individuals (which is an important finding), but they are less informative about other parameters of interest. For example, the 10th percentile of the effect of job displacement (which is a measure of the effect of job displacement for those most negatively effected by job displacement) is bounded between 26% lower earnings and 96% lower earnings. These bounds are also consistent with anywhere between 0% and 81% of workers having higher earnings following displacement than they would have had if they had not been displaced.

Another approach is to assume that the dependence is known. The leading choice is rank invariance.[1] This assumption says that individuals at a given rank in the distribution of treated potential outcomes would have the same rank in the distribution of untreated potential outcomes. This is a very strong assumption. In the case of job displacement, it imposes severe restrictions on how heterogeneous the effect of job displacement can be; for example, it prohibits any workers at the top of the distribution of non-displaced earnings from becoming unemployed 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.

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. Panel data presents a unique opportunity to observe, at least for some individuals, both their treated

[1] Rank invariance is also sometimes called perfect positive dependence, comonotonicity, or rank permanence in the literature. This assumption was first implicitly made in the earliest work on estimating the distributional effects of treatment (Doksum (1974) and 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 rank invariance (Bitler, Gelbach, and Hoynes (2006), Dong and Shen (2018), and Frandsen and Lefgren (2018))

and untreated potential outcomes 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.

Even though panel data appears to be useful for identifying the joint distribution of potential outcomes, there are still some challenges. Let $Y_{1t}$ denote treated potential outcomes in the last period, $Y_{0t}$ denote untreated potential outcomes in the last period, and $Y_{0t-1}$ denote untreated potential outcomes in the previous period. Under the condition that no one is treated until the last period, then the joint distribution of $(Y_{1t}, Y_{0t-1})$ is identified for the treated group – this is the joint distribution of treated potential outcomes in the last period and untreated potential outcomes in the prior period which is observed for treated individuals. Standard identifying assumptions may be used to identify the marginal distribution $Y_{0t}$ for the treated group. But tighter bounds on the distribution of the treatment effect hinge on obtaining restrictions on the joint distribution of $(Y_{1t}, Y_{0t})$ – the joint distribution of treated and untreated potential outcomes in the last period. Panel data alone does not provide these restrictions.

The main assumption in the current paper is that the dependence (or copula) of untreated potential outcomes over time does not change over time. I call this assumption the Copula Stability Assumption. This assumption combined with the panel data setup mentioned above leads to identification of the joint distribution of $(Y_{0t}, Y_{0t-1})$ – the joint distribution of untreated potential outcomes in the last two periods for the treated group. With this joint distribution in hand, 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)). 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 (2019)). But that approach is not possible in the current context because the dependence between treated and untreated potential outcomes is never observed – even in previous periods. Instead panel data is informative about the dependence between untreated potential outcomes over time, which leads to bounds instead of point identification here. To utilize the Copula Stability Assumption requires that at least three periods of panel data are available. In order to assess the validity of the

Copula Stability Assumption, I consider what additional conditions are required for the Copula Stability Assumption to hold in models with time invariant unobserved heterogeneity and panel data. I also show that the copula of earnings over time is stable over time and discuss how to "pre-test" the Copula Stability Assumption in cases where a researcher has access to more than three periods of panel data.

To see how the Copula Stability Assumption leads to tighter bounds, consider the following 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) rank invariance is observed between $Y_{1t}$ and $Y_{0t-1}$ or (B) rank invariance is observed between $Y_{0t-1}$ and $Y_{0t-2}$. The first case is very similar to the leading idea for point identification – rank invariance across treated and untreated potential outcomes – though it also involves an additional time dimension. The second case turns out to correspond exactly to the leading assumption for point identification with panel data – rank invariance of untreated potential outcomes over time (see the discussion in the next section). 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 falls exactly into this category. Neither type of rank invariance is observed; nonetheless, there is strong positive dependence which results in substantially tighter bounds.

The approach developed in the current paper is related to other work on tighter bounds on distributional treatment effect parameters. Fan and Park (2009), Fan and Park (2010), Fan, Guerre, and Zhu (2017), and Firpo and Ridder (2019) bound parameters that depend on the joint distribution when covariates are available. I discuss how this approach can be combined with the approach considered in the current paper to obtain even tighter bounds. 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 (2018) combines this assumption with the statistical bounds approach. MTR would imply that earnings for displaced workers cannot be larger than earnings would have been had they not been displaced. Frandsen and Lefgren (2017) obtain tighter bounds on the joint distribution of potential outcomes by ruling out negative dependence between the potential outcomes. Masten and Poirier (2019) construct breakdown frontiers which are informative about how robust results are to deviations from rank invariance assumptions. There is also 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. A common alternative approach to studying treatment effect heterogeneity is to see how the average treatment effect differs across groups which are defined by their observable characteristics (see the discussion in Bitler, Gelbach, and Hoynes (2017) for more details about this approach and its relationship to treatment effect heterogeneity due to unobservables). The current paper is also related to work on bounds under relatively weak assumptions in other contexts (for example, Manski (1990), Manski and Pepper (2000), Blundell, Gosling, Ichimura, and Meghir (2007), Kline and Santos (2013), Gechter (2016), and Kline and Tartari (2016)).

I propose estimators of the bounds on the main parameters of interest. These estimators depend on a number of first-step estimators of conditional distribution functions. I propose estimating these conditional distribution functions using quantile regression and distribution regression (similar approaches are taken in Melly and Santangelo (2015) and Wüthrich (2019)). I also provide the limiting distribution of the estimators of the bounds and propose using the numerical bootstrap (Hong and Li (2018)) to conduct inference. These asymptotic results are related to work on inference on partially identified parameters that depend on the joint distribution of potential outcomes (Fan and Park (2010) and Fan and Wu (2010)) and build on recent results on Hadamard directionally differentiable functions (Fang and Santos (2019), Hong and Li (2018), and Masten and Poirier (2019)). As an intermediate step, I develop some new results on distribution regression estimators with generated regressors.

I apply the methods developed in the paper to study heterogeneous effects of job displacement. Using standard techniques, I find that annual earnings of displaced workers were, on average, 34% lower in 2011 than they would have been had the worker not been displaced. Then, using the methods developed in the paper, I construct bounds on distributional treatment effect parameters that exploit having access to panel data. These bounds are substantially tighter than existing bounds and provide a credible alternative to point identifying assumptions that are not likely to hold in the current application. I estimate that the 10th percentile of earnings losses due to displacement is bounded between 46% lower earnings and 89% lower earnings. I also find that at least 10% 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 existing approaches.

# 2 Parameters of Interest

### Notation

The notation used throughout the paper 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 $D = 1$ for individuals that participate in the treatment and $D = 0$ for individuals who do not participate in the treatment (to minimize notation, a subscript $i$ representing each individual is omitted throughout the paper except in a few cases to increase clarity). The paper considers the case where panel data is available. The baseline case considered in the paper is one where there are exactly three time periods though the results could be extended to the case with more time periods. Throughout the paper, I use $s$ to represent a generic time period and $t$, $t-1$, and $t-2$ to represent particular time periods. Each individual has potential outcomes in the treated and untreated states in each time period which are given by $Y_{1s}$ and $Y_{0s}$, respectively. For each individual, only one of these potential outcomes is observed at each time period; I denote an individual's observed outcome in a particular time period by $Y_s$. For individuals that are treated in period $s$, $Y_{1s}$ is observed, but $Y_{0s}$ is not observed. For individuals that are untreated in period $s$, $Y_{0s}$ is observed but $Y_{1s}$ is unobserved. I make the following assumption

**Assumption 1.** *The observed data consists of $n$ observations of $\{Y_{idt}, Y_{0it-1}, Y_{0it-2}, X_i, D_i\}$ which are independently and identically distributed.*

Assumption 1 covers available data in the baseline case considered in the paper. In particular, Assumption 1 says that the researcher observes outcomes in three periods. The researcher may also observe a vector of covariates $X$ which, following much of the treatment effects literature (e.g., Heckman, Ichimura, and Todd (1997) and Abadie (2005)), I assume are time invariant. Assumption 1 also says that individuals are first treated in the last period which implies that untreated potential outcomes are observed for both the treated group and the untreated group in periods $t-1$ and $t-2$. That is,

$$Y_t = DY_{1t} + (1-D)Y_{0t}, \quad Y_{t-1} = Y_{0t-1}, \quad \text{and} \quad Y_{t-2} = Y_{0t-2}$$

Assumption 1 can be relaxed if additional periods are available or if treatment can occur in other periods besides the last one, but it represents a baseline case for tighter bounds and corresponds to the data used to study job displacement.

The next assumption is the starting point for the main identification results in the paper.[2]

**Assumption 2.** $\mathrm{F}_{Y_{1t}|D=1}$ *and* $\mathrm{F}_{Y_{0t}|D=1}$ *are identified.*

Assumption 2 says that the marginal distribution of treated potential outcomes for the treated group, $\mathrm{F}_{Y_{1t}|D=1}$, and the marginal distribution of untreated potential outcomes for the treated group, $\mathrm{F}_{Y_{0t}|D=1}$, are identified. The first is not a strong assumption – it is given by the distribution of observed outcomes for the treated group, $\mathrm{F}_{Y_t|D=1}$. The second is a stronger assumption. This counterfactual distribution would be identified if, for example, treatment were randomly assigned. However, in cases with observational data, like job displacement, it requires some identifying assumption. But there are many methods available to identify this counterfactual distribution. At any rate, the goal of the current paper is to go beyond the more standard objective of identifying this counterfactual distribution and learn about the joint distribution; thus, at this point, Assumption 2 considers the more standard problem of identifying the counterfactual marginal distribution to be solved.[3]

[2] All of the results in the paper continue to go through after conditioning on covariates $X$. Throughout most of this section, I omit conditioning on covariates to keep the notation simpler and focus on main ideas; however, bounds on parameters of interest can be tightened when there are available covariates by combining the results in the current paper with existing results on tightening bounds in the presence of covariates (Fan and Park (2009), Fan and Park (2010), Fan, Guerre, and Zhu (2017), and Firpo and Ridder (2019)). This is straightforward to do in practice (see the discussion in Remark 2 below).

[3] There are some cases where the identifying assumption for $F_{Y_{0t}|D=1}$ may provide additional structure that could potentially

Assumption 2 implies that parameters that depend on the marginal distributions of treated and untreated potential outcomes for the treated group are identified. These include the Average Treatment Effect on the Treated (ATT)[4]

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

and the Quantile Treatment Effect on the Treated (QTT)

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

for $\tau \in (0, 1)$ and where $\mathrm{F}^{-1}_X(\tau) = \inf\{x : \mathrm{F}_X(x) \geq \tau\}$. But distributional parameters that depend on the joint distribution of potential outcomes are not identified and these may be of considerable interest. For job displacement, I focus primarily on the Distribution of the Treatment Effect for the Treated (DoTT) and closely related parameters that are simple functionals of the DoTT; Heckman, Smith, and Clements (1997) and Firpo and Ridder (2019) discuss other parameters in this class that may be of interest in other applications.

The DoTT is the fraction of individuals that experience a treatment effect less than some value $\delta$. It is given by

$$DoTT(\delta) = \mathrm{P}(Y_{1t} - Y_{0t} \leq \delta|D = 1)$$

One can estimate the DoTT for different values of $\delta$ and plot them. An alternative approach, and the one that seems more useful for studying job displacement is to invert the DoTT to obtain the

tighten the bounds on parameters that depend on the joint distribution of treated and untreated potential outcomes. See Footnote 15 for more discussion of this point. I thank an anonymous referee for pointing this out.

[4]All the parameters mentioned in this section 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 because only for the treated group does one observe both treated and untreated potential outcomes, albeit at different points in time. Using the techniques presented in the current paper can still lead to bounds on parameters for the entire population by combining the bounds for the treated group presented in the current paper 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 paper.

Quantile of the Treatment Effect on the Treated (QoTT) which is given by

$$QoTT(\tau) = \inf\{\delta : DoTT(\delta) \geq \tau\}$$

To give some examples, in the context of job displacement, $QoTT(0.05)$ is the 5th percentile of the individual level effect of job displacement – these are the workers who experience some of the largest negative effects of job displacement. $QoTT(0.5)$ is the median effect of job displacement. And $QoTT(0.95)$ is the effect of job displacement for workers who have close to the highest earnings relative to what they would have if they had not been displaced. Also, for some $0 < \tau_L < \tau_U < 1$, I use $QoTT(\tau_U) - QoTT(\tau_L)$ as a measure of treatment effect heterogeneity. An example is the difference between the 95th percentile of earnings following job displacement and the 5th percentile of earnings following job displacement.

Another interesting parameter for job displacement is the fraction of workers that have higher earnings following job displacement than they would have had if they had not been displaced. Let $\beta$ denote this probability; it is given by

$$\beta = 1 - DoTT(0)$$

One can also consider the fraction of workers who are much worse off due to job displacement by considering $DoTT(\delta^*)$ for some large negative value $\delta^*$.

## 2.1 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, by Assumption 2, both the marginal distribution of treated potential outcomes for the treated group $\mathrm{F}_{Y_{1t}|D=1}$ and the marginal distribution of untreated potential outcomes for the treated group $\mathrm{F}_{Y_{0t}|D=1}$ 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

$$\mathrm{F}_{Y_{1t},Y_{0t}|D=1}(y_1, y_0) = C_{Y_{1t},Y_{0t}|D=1}\left(\mathrm{F}_{Y_{1t}|D=1}(y_1), \mathrm{F}_{Y_{0t}|D=1}(y_0)\right) \tag{2.1}$$

where $C_{Y_{1t},Y_{0t}|D=1}(\cdot,\cdot) : [0,1]^2 \to [0,1]$. 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 (Hoeffding (1940) and Fréchet (1951)). These bounds arise from considering two extreme cases: (i) when there is rank invariance 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.[5] 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 rank invariance between treated potential outcomes and untreated potential outcomes for the treated group. This assumption can be written as

**Alternative Assumption 1** (Cross Sectional Rank Invariance)**.**

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

The Cross Sectional Rank Invariance Assumption implies that

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

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 re-

[5] In that paper and in the current paper, the Fréchet-Hoeffding bounds can rule out the common effects model (i.e. that the effect of the treatment is the same for all individuals); however, these bounds are much less useful for understanding some other aspects of individual-level treatment effect heterogeneity.

stricts the ability of treatment to have different effects across different individuals. In the context of job displacement, rank invariance seems unlikely to hold because it would prohibit individuals at the top of the pre-displacement earnings distribution from being unemployed or taking a part time job following job displacement.

With panel data, an alternative assumption that also leads to point identification is rank invariance in untreated potential outcomes over time:

**Alternative Assumption 2** (Rank Invariance Over Time)**.**

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

The Rank Invariance Over Time Assumption does not directly replace the unknown copula in Equation 2.1; however, it does lead to point identification of the joint distribution. To see this, note that under this assumption,

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

which implies that the joint distribution $F_{Y_{1t},Y_{0t}|D=1}$ is identified.

Rank invariance over time is a strong assumption. It says that, in the absence of participating in the treatment, individuals would keep the same rank in the distribution of outcomes over time. This seems unlikely to hold in most applications in economics. 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 rank invariance in untreated potential outcomes holds between periods $t-1$ and $t-2$ and this can provide evidence as to whether or not rank invariance is likely to hold between periods $t$ and $t-1$. In the application in the current paper, I find that this assumption does not hold. That being said, although rank invariance over time does not hold, there is strong positive dependence between earnings over time. The approach taken in the current paper exploits this strong positive dependence in order to deliver tighter bounds.

# 3 Identification

In the previous section, I have argued that assumptions that directly replace the unknown copula in Equation 2.1 are not likely to hold. This section considers an alternative approach that does not substitute for the copula in Equation 2.1 directly but limits the possibilities for the copula.

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

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

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

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, $\mathrm{F}_{Y_{0t}|D=1}$, is identified and the distribution of untreated potential outcomes for the treated at period $t-1$, $\mathrm{F}_{Y_{0t-1}|D=1}$, 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, $\mathrm{F}_{Y_{0t},Y_{0t-1}|D=1}$, is identified. This joint distribution is not of primary interest in the current paper. But knowledge of this joint distribution is important for deriving tighter 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 rank invariant) of untreated potential outcomes at period $t-2$, then untreated potential outcomes at period $t$ will continue to be independent (or rank invariant) of untreated potential outcomes at period $t-1$.[6] Or, for example, suppose the copula for

[6]This also implies that, if one conducted the pre-test of rank invariance over time mentioned in the previous section and

$(Y_{0t-1}, Y_{0t-2}|D = 1)$ is Gaussian with parameter $\rho$, the Copula Stability Assumption says that the copula for $(Y_{0t}, Y_{0t-1}|D = 1)$ is also Gaussian with parameter $\rho$ though the marginal distributions of outcomes can change in unrestricted ways. For example, the distribution of earnings can shift 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.[7]

**Assumption 3.** *(Outcomes are continuously distributed)*

*$Y_{0t}$, $Y_{0t-1}$ and $Y_{0t-2}$ are continuously distributed conditional on $D = 1$.*

Assumption 3 is helpful for utilizing the Copula Stability Assumption. Continuously distributed outcomes imply that $C_{Y_{0t-1}, Y_{0t-2}|D=1}$ is uniquely identified from the sampling process. In practice, this assumption allows for the quantile functions in the expressions below to be well-defined.[8] Importantly, Assumption 3 does not require that $Y_{1t}$ is continuously distributed. In the application on job displacement, this allows for some individuals to not be employed following job displacement because it can allow for a mass point in the distribution of treated potential outcomes.

Next, as a preliminary result, I show that, under the Copula Stability Assumption, the joint distribution of $(Y_{0t}, Y_{0t-1})|D = 1$ is identified. Recall that this is not the joint distribution of interest, but identifying this joint distribution is going to provide identifying power for distributional treatment effect parameters that depend on the joint distribution of $(Y_{1t}, Y_{0t})|D = 1$.

---

did not reject that assumption in the previous period, then the Copula Stability Assumption would impose rank invariance over time in the current period which would lead to point identification of all the parameters of interest.

[7]One could also make the Copula Stability Assumption conditional on some covariates $X$. This type of assumption might be more plausible in some applications. For example, earnings over time may be more strongly positively dependent for older workers than for younger workers.

[8]This condition could be weakened to: $\text{Range}(\text{F}_{Y_{0t}|D=1}) \subseteq \text{Range}(\text{F}_{Y_{0t-1}|D=1})$ and $\text{Range}(\text{F}_{Y_{0t-1}|D=1}) \subseteq \text{Range}(\text{F}_{Y_{0t-2}|D=1})$. In this case, all of the identification results in the paper would continue to go through. Assuming that untreated potential outcomes are continuously distributed seems more natural though and is a special case of this condition.

**Lemma 1.** *Under Assumptions 1 to 3 and the Copula Stability Assumption,*

$$F_{Y_{0t},Y_{0t-1}|D=1}(y_0, y') = F_{Y_{0t-1},Y_{0t-2}|D=1}\Big(F^{-1}_{Y_{0t-1}|D=1} \circ F_{Y_{0t}|D=1}(y_0), F^{-1}_{Y_{0t-2}|D=1} \circ F_{Y_{0t-1}|D=1}(y')\Big)$$

*and*

$$\begin{aligned} F_{Y_{0t}|Y_{0t-1},D=1}(y_0|y') &= F_{Y_{0t-1}|Y_{0t-2},D=1}\Big(F^{-1}_{Y_{0t-1}|D=1} \circ F_{Y_{0t}|D=1}(y_0)\Big|F^{-1}_{Y_{0t-2}|D=1} \circ F_{Y_{0t-1}|D=1}(y')\Big) \\ &= P\Big(Y_{0t-1} \leq F^{-1}_{Y_{0t-1}|D=1} \circ F_{Y_{0t}|D=1}(y_0)\Big|F^{-1}_{Y_{0t-1}|D=1} \circ F_{Y_{0t-2}|D=1}(Y_{0t-2}) = y', D = 1\Big) \end{aligned}$$

The first part of Lemma 1 shows that the joint distribution of $(Y_{0t}, Y_{0t-1})|D = 1$ is identified under the Copula Stability Assumption and gives an expression for it. The second part provides an expression for the conditional distribution $F_{Y_{0t}|Y_{0t-1},D=1}$ which turns out to be quite useful later as well. The intuition for Lemma 1 is that: the Copula Stability Assumption implies that one can learn about the joint distribution of $(Y_{0t}, Y_{0t-1})|D = 1$ from the joint distribution of $(Y_{0t-1}, Y_{0t-2})|D = 1$, but changes in the marginal distributions over time are unrestricted and therefore need to be adjusted. This is what the terms like $F^{-1}_{Y_{0t-1}|D=1} \circ F_{Y_{0t}|D=1}(y_0)$ do; they take the distribution of untreated potential outcomes at time period $t$ and adjust it back to the distribution of untreated potential outcomes in time period $t - 1$.

Next, I show how the Copula Stability Assumption can be used to derive tighter bounds on the joint distribution of potential outcomes. 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 2.** *Under Assumptions 1 to 3 and the Copula Stability Assumption, bounds on the joint distribution of treated and untreated potential outcomes for the treated group conditional on outcomes in the previous period are given by*

$$\mathrm{F}^L_{Y_{1t},Y_{0t}|Y_{0t-1},D=1}(y_1, y_0|y') \leq \mathrm{F}_{Y_{1t},Y_{0t}|Y_{0t-1},D=1}(y_1, y_0|y') \leq \mathrm{F}^U_{Y_{1t},Y_{0t}|Y_{0t-1},D=1}(y_1, y_0|y')$$

*where*

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

The next theorem is the main result for bounds on the joint distribution of potential outcomes for the treated group.

**Theorem 1.** *Under Assumptions 1 to 3 and the Copula Stability Assumption, bounds on the joint distribution of treated and untreated potential outcomes for the treated group are given by*

$$\mathrm{F}^{L}_{Y_{1t},Y_{0t}|D=1}(y_1,y_0) \leq \mathrm{F}_{Y_{1t},Y_{0t}|D=1}(y_1,y_0) \leq \mathrm{F}^{U}_{Y_{1t},Y_{0t}|D=1}(y_1,y_0)$$

*where*

$$\mathrm{F}^{L}_{Y_{1t},Y_{0t}|D=1}(y_1,y_0) = \mathrm{E}[\mathrm{F}^{L}_{Y_{1t},Y_{0t}|Y_{0t-1},D=1}(y_1,y_0|Y_{0t-1})|D=1]$$
$$\mathrm{F}^{U}_{Y_{1t},Y_{0t}|D=1}(y_1,y_0) = \mathrm{E}[\mathrm{F}^{U}_{Y_{1t},Y_{0t}|Y_{0t-1},D=1}(y_1,y_0|Y_{0t-1})|D=1]$$

The bounds in Theorem 1 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 rank invariance across treated and untreated groups (though it also includes a time dimension); Item (ii) is exactly the condition of rank invariance in untreated potential outcomes over time used as a point identifying assumption. 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 in the next example and provide a more formal proof in the proposition that follows.

**Example 1.** *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 the case in the current paper) are known (Joe (2015, Theorem 8.19)). Because the marginal distributions* $\mathrm{F}_{Y_{1t}|D=1}(Y_{1t})$, $\mathrm{F}_{Y_{0t}|D=1}(Y_{0t})$, *and* $\mathrm{F}_{Y_{0t-1}|D=1}(Y_{0t-1})$ *are uniformly distributed conditional on* $D = 1$, *their covariance matrix is given by*

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

*Consider the case where* $\rho_{13}$ *and* $\rho_{23}$ *are identified and* $\rho_{12}$ *is not known.* $\rho_{12}$ *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* $\rho_{23}$ *with* $|\rho_{23}| < 1$, *the width of the bounds on* $\rho_{12}$ *are decreasing as* $\rho_{13}$ *increases for* $\rho_{13} > 0$, *and width of the bounds are decreasing as* $\rho_{13}$ *decreases for* $\rho_{13} < 0$. *When either* $\rho_{13}$ *or* $\rho_{23}$ *is equal to one in absolute value,* $\rho_{12}$ *is point identified. This corresponds exactly to the case of rank invariance (or perfect negative dependence) mentioned above for point identification. The intuition of this result is that as the copula moves "closer" to rank invariance* *or perfect negative dependence, the bounds on the joint distribution of interest shrink.*

**Proposition 1.** *Fix the marginal distributions $F_{Y_{1t}|D=1}$, $F_{Y_{0t}|D=1}$, and $F_{Y_{0t-1}|D=1}$ and the conditional distribution $F_{Y_{1t}|Y_{0t-1},D=1}$. Now consider two possibilities for $F_{Y_{0t}|Y_{0t-1},D=1}$ given by $F_1$ and $F_2$. Assume that $F_1$, $F_2$, and $F_{Y_{1t}|Y_{0t-1},D=1}$ are stochastically increasing*[9] *and that $F_{Y_{1t}|Y_{0t-1},D=1} \prec^{SI} F_1 \prec^{SI} F_2$ where $F \prec^{SI} G$ indicates that $G$ is more stochastically increasing than $F$. Then, the bounds on the joint distribution given in Theorem 1 are at least as tight when $F_{Y_{0t}|Y_{0t-1},D=1} = F_2$ as when $F_{Y_{0t}|Y_{0t-1},D=1} = F_1$.*

Proposition 1 is a key result in the paper. It says that the bounds in the paper get tighter when there is stronger dependence between $Y_{0t}$ and $Y_{0t-1}$ (under the Copula Stability Assumption, this will be true when the dependence between $Y_{0t-1}$ and $Y_{0t-2}$ is stronger). An analogous result also holds for $Y_{1t}$ and $Y_{0t-1}$ – if the dependence is strong, the bounds will be tight. In the literature, assumptions of rank invariance have been made as approximations because in many applications there is strong positive dependence though less than rank invariance. Proposition 1 implies that the results in the current paper will be valid when rank invariance assumptions are violated, but the bounds will be "tight" in the case where these assumptions are not too far from the truth. This is likely to be the most relevant case in many applications. When the bounds are applied to job displacement later, I show that there is strong positive dependence (both between $Y_{1t}$ and $Y_{0t-1}$ and between $Y_{0t-1}$ and $Y_{0t-2}$ for the treated group) but less than rank invariance. This implies that the assumptions of rank invariance will be violated, but it also implies that the bounds can be tightened substantially over bounds that only use the information from the marginal distributions of $Y_{1t}$ and $Y_{0t}$, respectively.

Proposition 1 also implies that bounds obtained under the Copula Stability Assumption are robust to some violations of the Copula Stability Assumption. In particular, when there is stronger dependence (in terms of "more stochastically increasing") between $Y_{0t}$ and $Y_{0t-1}$ than there was for $Y_{0t-1}$ and $Y_{0t-2}$ for individuals in the treated group, then the bounds developed under the

[9]For two random variables $X$ and $W$, their conditional distribution $F_{X|W}$ is said to be stochastically increasing if $1 - F_{X|W}(x|w)$ is increasing in $w$ for all $x$. For two conditional distributions $F_{X|W}$ and $G_{X|W}$, having the same marginal distributions of $X$ and $W$, $F_{X|W}$ is said to be more stochastically increasing than $G_{X|W}$ if $F^{-1}_{X|W}(G_{X|W}(x|w)|w)$ is increasing in $w$ for all $x$. Stochastically increasing is a well-known dependence property and more stochastically increasing is a common dependence ordering (see Yanagimoto and Okamoto (1969) and Schriever (1987) as well as related discussion in Joe (1997) and Nelsen (2007)).

Copula Stability Assumption are conservative.

Just as knowledge of $F_{Y_{1t},Y_{0t-1}|D=1}$ and $F_{Y_{0t},Y_{0t-1}|D=1}$ leads to bounds on the joint distribution of interest $F_{Y_{1t},Y_{0t}|D=1}$, knowledge of these distributions can also be used to bound the DoTT, the QoTT, 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), and Williamson and Downs (1990)). I use these same bounds for the conditional joint distribution.

**Lemma 3.** *(Conditional Distribution of the Treatment Effect) Under Assumptions 1 to 3 and the Copula Stability Assumption, bounds on the distribution of the treatment effect for the treated group conditional on the outcome in the previous period are given by*

$$F^L_{Y_{1t}-Y_{0t}|Y_{0t-1},D=1}(\delta|y') \leq F_{Y_{1t}-Y_{0t}|Y_{0t-1},D=1}(\delta|y') \leq F^U_{Y_{1t}-Y_{0t}|Y_{0t-1},D=1}(\delta|y')$$

*where*

$$\begin{aligned}
F^L_{Y_{1t}-Y_{0t}|Y_{0t-1},D=1}(\delta|y') &= \sup_y \max\{F_{Y_{1t}|Y_{0t-1},D=1}(y|y') - F_{Y_{0t}|Y_{0t-1},D=1}(y-\delta|y'), 0\} \\
F^U_{Y_{1t}-Y_{0t}|Y_{0t-1},D=1}(\delta|y') &= 1 + \inf_y \min\{F_{Y_{1t}|Y_{0t-1},D=1}(y|y') - F_{Y_{0t}|Y_{0t-1},D=1}(y-\delta|y'), 0\}
\end{aligned}$$

The next result provides bounds for the DoTT.

**Theorem 2.** *(Distribution of the Treatment Effect) Under Assumptions 1 to 3 and the Copula Stability Assumption, bounds on $DoTT(\delta)$ are given by*

$$DoTT^L(\delta) \leq DoTT(\delta) \leq DoTT^U(\delta)$$

*where*

$$DoTT^L(\delta) = \mathrm{F}^L_{Y_{1t}-Y_{0t}|D=1}(\delta) = \mathrm{E}[\mathrm{F}^L_{Y_{1t}-Y_{0t}|Y_{0t-1},D=1}(\delta|Y_{0t-1})|D=1]$$
$$DoTT^U(\delta) = \mathrm{F}^U_{Y_{1t}-Y_{0t}|D=1}(\delta) = \mathrm{E}[\mathrm{F}^U_{Y_{1t}-Y_{0t}|Y_{0t-1},D=1}(\delta|Y_{0t-1})|D=1]$$

*where* $\mathrm{F}^L_{Y_{1t}-Y_{0t}|Y_{0t-1},D=1}(\delta|y')$ *and* $\mathrm{F}^U_{Y_{1t}-Y_{0t}|Y_{0t-1},D=1}(\delta|y')$ *are given in Lemma 3.*

**Remark 1.** *The results of Proposition 1 also imply that the bounds on the $DoTT$ are tighter under the conditions given in Proposition 1. To be specific, consider two possibilities for $F_{Y_{0t}|Y_{0t-1},D=1}$ given by $F_1$ and $F_2$ with $F_1 \prec^{SI} F_2$ as in Proposition 1 and assume all other conditions as in that proposition. Proposition 1 implies $C_1^L \prec^C C_2^L$ where $C_j^L$ denotes the lower bound on the copula of $Y_{1t}$ and $Y_{0t}$ when $F_{Y_{0t}|Y_{0t-1},D=1} = F_j$ for $j = 1, 2$. Bounds on the $DoTT$ get tighter when the lower bound of the copula is more concordant (Williamson and Downs (1990)). This implies that the bounds on the $DoTT$ will be tighter when there is more positive dependence between $Y_{0t}$ and $Y_{0t-1}$ (which occurs under the Copula Stability Assumption when there is more positive dependence between $Y_{0t-1}$ and $Y_{0t-2}$).*

**Remark 2.** *As mentioned above, the main identification results continue to go through when the assumptions hold conditional on covariates. In particular, in this case, one can derive bounds on the conditional $DoTT$, as in Lemma 3, that are conditional on both $Y_{0t-1}$ and $X$. Then, following the same arguments for the unconditional $DoTT$ as in Theorem 2, one can derive bounds on the unconditional $DoTT$ by averaging over both $Y_{0t-1}$ and $X$. These bounds will be tighter than the bounds that do not include covariates which follows using the same arguments as in Fan and Park (2009), Fan and Park (2010), Fan, Guerre, and Zhu (2017), and Firpo and Ridder (2019).*

Bounds on the QoTT can be obtained from the bounds on the DoTT. The upper bound on the QoTT comes from inverting the lower bound of the DoTT, and the lower bound on the QoTT comes from inverting the upper bound on the DoTT.

**Theorem 3.** *(Quantile of the Treatment Effect) Under Assumptions 1 to 3 and the Copula Sta-*

*bility Assumption, bounds on $QoTT(\tau)$ are given by*

$$QoTT^L(\tau) \leq QoTT(\tau) \leq QoTT^U(\tau)$$

*where*

$$QoTT^L(\tau) = \inf\{\delta : DoTT^U(\delta) \geq \tau\}$$
$$QoTT^U(\tau) = \inf\{\delta : DoTT^L(\delta) \geq \tau\}$$

*and $DoTT^L(\delta)$ and $DoTT^U(\delta)$ are given in Theorem 2.*

# 4 How Plausible is the Copula Stability Assumption?

Since the Copula Stability Assumption is the key identifying assumption used in the paper and is crucial for exploiting panel data to deliver tighter bounds on the distributional treatment effect parameters considered in this paper, it is worth considering how plausible this assumption is. In this section, I consider several types of evidence. First, I consider what additional restrictions need to be satisfied in typical panel data models (particularly models with individual heterogeneity) in order for the Copula Stability Assumption to hold. Second, I use data to test if the Copula Stability Assumption holds in an important application in economics. Finally, I propose a simple test for the Copula Stability Assumption in any application where there are more than two pre-treatment time periods that is similar to "pre-tests" that are commonly conducted in applied work with panel data.

## 4.1 Additional Conditions for the Copula Stability Assumption to Hold

In this section, I consider two leading models for untreated potential outcomes when there is time invariant unobserved heterogeneity and when panel data is available: (i) two-way fixed effects models and (ii) the Change in Changes model (Athey and Imbens (2006)).

### 4.1.1 Two-way Fixed Effects

Consider the following two-way fixed effects model for untreated potential outcomes. For $s \in \{t, t-1, t-2\}$

$$Y_{0is} = \theta_t + \eta_i + V_{is} \tag{4.1}$$

where $\theta_t$ is a time fixed effect, $\eta_i$ is time invariant unobserved heterogeneity that can be distributed differently for individuals in the treated group and untreated group, and $V_{is}$ are time varying unobservables. Equation (4.1) is a leading model in the treatment effects literature in the case with time invariant unobserved heterogeneity and panel data; in particular, it corresponds to the sort of model required for Difference in Differences designs to identify the ATT (see discussion in Blundell and Dias (2009)). The next result provides conditions under which the Copula Stability Assumption holds in the model in Equation (4.1).

**Proposition 2.** *In the two-way fixed effects model for untreated potential outcomes given above, and under the additional condition that $C_{\eta+V_t,\eta+V_{t-1}|D=1} = C_{\eta+V_{t-1},\eta+V_{t-2}|D=1}$, the Copula Stability Assumption holds.*

The additional condition in Proposition 2 is fairly weak. It allows for serial correlation in the time varying unobservables and for the distribution of time varying unobservables to change over time. It also allows for the distribution of time varying unobservables to depend on the value of the individual heterogeneity. There are also a number of special cases of this condition that are familiar as well. One example is when $F_{V_t,V_{t-1}|\eta,D=1} = F_{V_{t-1},V_{t-2}|\eta,D=1}$. This says that the joint distribution of time varying unobservables does not change over time conditional on individual unobserved heterogeneity. Another leading case is when $(V_t, V_{t-1}, V_{t-2}) \perp\!\!\!\perp \eta | D = 1$. In this case, the Copula Stability Assumption will hold as along as $C_{V_t,V_{t-1}|D=1} = C_{V_{t-1},V_{t-2}|D=1}$. This latter condition will hold, for example, in the case where the $V_s$ are mutually independent. These sorts of conditions are frequently invoked in the literature on treatment effects with panel data using identification arguments from the measurement error literature (e.g., Li and Vuong (1998), Evdokimov (2010), Bonhomme and Sauder (2011), Canay (2011), and Freyberger (2018)).

### 4.1.2 Change in Changes

In the application on job displacement, I use the Change in Changes (CIC) model (Athey and Imbens (2006)) to identify the counterfactual distribution of outcomes $F_{Y_{0t}|D=1}$ that individuals in the treated group would have experienced if they had not been displaced from their job. Therefore, it is useful to consider what extra conditions need to be placed on the CIC model in order for the Copula Stability Assumption to hold. The CIC model is based on the following setup:

$$Y_{0is} = h_s(U_{is}) \quad \text{for } s = t, t-1$$

with (i) $h_s(u)$ strictly increasing in $u$ and (ii) $U_s|D = d \sim F_{U|D=d}$ for $s = t, t-1$ and $d = 0, 1$. In addition to these conditions, I also assume $Y_{0it-2} = h_{t-2}(U_{it-2})$, $U_{is} = \eta_i + V_{is}$ for $s = t, t-1, t-2$, and $V_s|\eta, D = d \sim F_{V|\eta,D=d}$ for all $s = t, t-1, t-2$ (this last condition implies $U_s|D = d \sim F_{U|D=d}$ in each time period). These extra conditions simply extend the model to three periods and from the case with repeated cross sections to panel data. Importantly they allow the distribution of $\eta$ to differ across the treated and untreated group. I call the set of conditions above the Three Period Panel CIC model.

**Proposition 3.** *In the Three Period Panel CIC model given above, and under the additional assumption that $F_{V_t,V_{t-1}|\eta,D=1} = F_{V_{t-1},V_{t-2}|\eta,D=1}$, the Copula Stability Assumption holds.*

Proposition 3 gives the additional condition required for the Copula Stability Assumption to hold in the CIC model. The additional condition is weak and would be satisfied if the $V_s$ are iid, but it can also allows for dependence of $V_s$ over time as long as the dependence is constant.

**Remark 3.** *Both the two way fixed effects model and the Change in Changes model allow for individuals to select into treatment (i) based on having "good" treated potential outcomes and (ii) based on their unobserved heterogeneity, $\eta_i$, that shows up in the model for untreated potential outcomes.*[10] *The additional conditions for the Copula Stability Assumption to hold in each model*

[10] The first part holds because the models do not put any structure on how treated potential outcomes are generated, and the second part holds because the unobserved heterogeneity can be distributed differently for individuals in the treated group relative to individuals in the untreated group.

*allow for this type of selection into treatment as well.*

**Remark 4.** *Neither of the rank invariance assumptions mentioned in Section 2 are likely to hold in the two way fixed effects model nor in the Change in Changes model. Cross sectional rank invariance is related to the model for treated potential outcomes (which was not specified in either case above) as well as the model for untreated potential outcomes, but it would take an unusual model for treated potential outcomes to generate cross sectional rank invariance. Rank invariance over time is not compatible with the time varying shocks that show up in both models given in this section – these cause individuals to change their ranks in the distribution of outcomes over time.*

## 4.2 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 the correlation between the ranks of income over time (Kopczuk, Saez, and Song (2010)),[11] is the same over time.[12]

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.[13] 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 1).

Moreover, Figure 1 also confirms the intuition that there is strong positive dependence of yearly

[11]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.

[12]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 (Duncan et al. (1984), Hungerford (1993), Gottschalk (1997), and Carroll, Joulfaian, and Rider (2007)).

[13]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.

income over time though the dependence is less than rank invariance. This is precisely the case where the method developed in the current paper is likely to (i) provide more credible results than employing a rank invariance 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.

### 4.3 Pre-Testing the Copula Stability Assumption

In treatment effects applications with more periods than are required for identification, empirical researchers frequently "pre-test" their identifying assumptions (see, for example, Callaway and Sant'Anna (2019) and Roth (2018)). The idea is that, even though the identifying assumptions themselves are not directly testable, it is often reasonable to think that the identifying assumptions would have held in earlier periods as well. Then, one can compare results using observed untreated potential outcomes in pre-treatment periods to results coming from the identifying assumptions. In the current case, when there is an extra pre-treatment time period, one can pre-test the Copula Stability Assumption by testing if $C_{Y_{0t-1},Y_{0t-2}|D=1} = C_{Y_{0t-2},Y_{0t-3}|D=1}$; that is, one can test if the copula actually does not change over time in time periods before individuals in the treated group become treated. In principle, one could conduct this test by nonparametrically estimating each copula or, alternatively, building on results from the goodness-of-fit literature. A simple alternative though is to compute some dependence measure such as Spearman's Rho or Kendall's Tau and test if it remains constant across pre-treatment periods. This is a practical alternative in the sense that it is easy to implement though it could fail to detect certain violations of the Copula Stability Assumption in previous periods.

## 5 Estimation and Inference

This section proposes estimators for the DoTT and QoTT, provides the limiting distribution of these estimators, and shows the validity of the numerical bootstrap of Hong and Li (2018) to conduct inference. This section explicitly conditions on covariates $X$ in order to increase the clarity of the arguments though note that all of the unconditional results hold simply by taking the covariates to only include a constant. At a high level, the estimators of the DoTT and QoTT come from plugging in first step estimators into the expressions given in Theorems 2 and 3. In particular,

$$\widehat{DoTT}^L(\delta) = \frac{1}{n}\sum_{i=1}^{n}\frac{D_i}{p}\ \hat{F}^L_{Y_{1t}-Y_{0t}|Y_{0t-1},X,D=1}(\delta|Y_{it-1}, X_i)$$

$$\widehat{DoTT}^U(\delta) = \frac{1}{n}\sum_{i=1}^{n}\frac{D_i}{p}\ \hat{F}^U_{Y_{1t}-Y_{0t}|Y_{0t-1},X,D=1}(\delta|Y_{it-1}, X_i)$$

where $p$ is the fraction of individuals in the treated group.[14] The estimators of $F^L_{Y_{1t}-Y_{0t}|Y_{0t-1},X,D=1}$ and $F^U_{Y_{1t}-Y_{0t}|Y_{0t-1},X,D=1}$ further depend on preliminary estimators of $F_{Y_{1t}|Y_{0t-1},X,D=1}$ and $F_{Y_{0t}|Y_{0t-1},X,D=1}$ (see Lemma 3). $F_{Y_{1t}|Y_{0t-1},X,D=1}$ is identified by the sampling process and can be estimated directly; on the other hand, $F_{Y_{0t}|Y_{0t-1},X,D=1}$ requires further preliminary estimators of $F_{Y_{0t}|X,D=1}$, $F_{Y_{0t-1}|X,D=1}$, and $F_{Y_{0t-2}|X,D=1}$ (see Lemma 1). Recall that $F_{Y_{0t}|X,D=1}$, which is the counterfactual distribution of untreated potential outcomes for individuals in the treated group, requires an identifying assumption. I use a conditional version of the Change in Changes model (Melly and Santangelo (2015)) to identify this distribution.[15] Under this setup, the counterfactual distribution is given by

$$F_{Y_{0t}|X,D=1}(y|x) = F_{Y_{0t-1}|X,D=1}(F^{-1}_{Y_{0t-1}|X,D=0}(F_{Y_{0t}|X,D=1}(y|x)|x)|x)$$

[14] The term $D_i/p$ makes these terms equivalent to averaging over observations in the treated group only.

[15] As mentioned in Footnote 3, the assumptions used to identify the counterfactual distribution of untreated potential outcomes for individuals in the treated group may have identifying power for the joint distribution of treated and untreated potential outcomes. For the Change in Changes model, this is not the case. The reasons are that (i) this setup does not restrict how treated potential outcomes are generated at all, and (ii) it is consistent with any dependence structure between $Y_{0t}$ and $Y_{0t-1}$. Other models for untreated potential outcomes with panel data (e.g., Quantile Difference in Differences or the method proposed in Bonhomme and Sauder (2011)) do not imply any restrictions on the joint distribution of potential outcomes either. This is not always the case though. For example, if one assumes selection on observables (i.e., that treatment is as good as randomly assigned once one conditions on a lag of the outcome), that would identify the joint distributions of $(Y_{1t}, Y_{0-1})|D = 1$ and $(Y_{0t}, Y_{0t-1})|D = 1$ which would lead to the same bounds as the ones in the current paper without requiring the Copula Stability Assumption. That being said, in economics, models like the Change in Changes model are more common because they can be motivated using models that allow for an important role to be played by unobserved heterogeneity. For example, in the context of linear models, models such as Change in Changes are closely related to fixed effects models, while selection on observables is more similar to regressions that include lagged outcomes but not fixed effects.

Thus, to estimate the bounds on the DoTT and QoTT requires preliminary estimators of (i) $F_{Y_{1t}|Y_{0t-1},X,D=1}$ and $F_{Y_{0t}|Y_{0t-1},X,D=1}$ as well as (ii) $F_{Y_{0t-1}|X,D=1}$, $F_{Y_{0t-2}|X,D=1}$, $F_{Y_{0t}|X,D=0}$, and $F_{Y_{0t-1}|X,D=0}$. I propose using flexible parametric first step estimators based on distribution regression (for the first group) and quantile regression (for the second group).[16] Similar approaches have been used in Chernozhukov, Fernandez-Val, and Melly (2013), Melly and Santangelo (2015), and Wüthrich (2019). Using flexible parametric first step estimators for conditional quantiles and distributions is an attractive option in many applications. They are substantially more flexible than imposing that outcomes follow a particular distribution that is known up to a few parameters (e.g., normal), but they are also more feasible than nonparametric estimators that require extra regularity conditions and may be difficult to implement in many applications in economics. This is especially true in the frequently encountered case with a relatively large number of covariates and only a moderate number of observations.

Next, I develop some asymptotic theory for the estimators of the DoTT and QoTT. As mentioned above, I focus on the case where one estimates conditional distributions using distribution regression or quantile regression, but the arguments here are fairly general and could potentially be used with other first step estimators.

At a high level, the arguments in this section proceed by first showing that distribution regression estimators $F_{Y_{1t}|Y_{0t-1},X,D=1}$ and $F_{Y_{0t}|Y_{0t-1},X,D=1}$ converge uniformly to Gaussian processes. This step requires proving some new results on distribution regression estimators when one of the regressors is "generated" as well as when the index depends on an estimated transformation (see Appendix B.2). Second, building on work on inference on parameters that depend on the joint distribution of potential outcomes (Fan and Park (2010) and Fan and Wu (2010)) as well as recent results in the literature on Hadamard directionally differentiable functions (e.g, Fang and Santos (2019)), one can derive the limiting distribution of the upper and lower bound of the DoTT and QoTT. To conduct inference, in practice, the empirical bootstrap can be applied to simulate the limiting distribution of the first step estimators which can be combined with the numerical delta method (Hong and Li (2018)) to simulate the limiting distributions of estimators of the bounds.

It is also worth mentioning that I focus on constructing pointwise confidence intervals for the

[16] For these, I proceed by estimating all of their conditional quantiles using quantile regression and then inverting to obtain the conditional distribution.

bounds on the *DoTT* and *QoTT*. One main reason to consider pointwise confidence intervals is that the bounds themselves are pointwise (see the discussion in Firpo and Ridder (2019)). However, it does seem possible to extend these results to uniform confidence bands using, for example, the approach in Masten and Poirier (2019) and exploiting the monotonicity of the bounds. I also focus on inference for the bounds themselves though one could also consider inference on the *DoTT* or *QoTT* itself rather than the bounds (see, for example, Fan and Park (2012) and Chernozhukov, Lee, and Rosen (2013)).

For $(d,s) \in \{0,1\} \times \{t, t-1, t-2\}$, let $\hat{G}_{d,s}(y,x) = \sqrt{n}(\hat{F}_{Y_s|X,D=d}(y|x) - F_{Y_s|X,D=d}(y|x))$, and let $\hat{G}^0(y,x) = \sqrt{n}(\hat{F}_{Y_{0t}|X,D=1}(y|x) - F_{Y_{0t}|X,D=1}(y|x))$. Let $\mathcal{Y}_{ds}$, $\mathcal{X}_d$, and $\boldsymbol{\Delta}$ denote the supports of $Y$ for group $d$ in time period $s$, $X$ for group $d$, and $(Y_{1t} - Y_{0t})|D=1$, respectively. Also, let $\bar{\mathcal{Y}}_{0t}$ denote the support of $Y_{0t}|D=1$. Also, for some set $S$, let $l^\infty(S)$ denote the space of all uniformly bounded functions on $S$ that are equipped with the supremum norm and $\mathcal{C}(S)$ denote the space of all uniformly continuous functions on $S$. As a first step, I assume that a functional central limit theorem holds jointly for each of the first step estimators, and that they each converge uniformly at the parametric rate.

**Assumption 4** (Functional Central Limit Theorem for First-Step Estimators)**.**

$$\sqrt{n}(\hat{G}_{1,t}, \hat{G}_{1,t-1}, \hat{G}_{1,t-2}, \hat{G}_{0,t}, \hat{G}_{0,t-1}, \hat{G}^0) \rightsquigarrow (\mathbb{W}_{1,t}, \mathbb{W}_{1,t-1}, \mathbb{W}_{1,t-2}, \mathbb{W}_{0,t}, \mathbb{W}_{0,t-1}, \mathbb{W}^0)$$

*in the space* $\mathcal{S} = l^\infty(\mathcal{Y}_{1t}\mathcal{X}_1) \times l^\infty(\mathcal{Y}_{1t-1}\mathcal{X}_1) \times l^\infty(\mathcal{Y}_{1t-2}\mathcal{X}_1) \times l^\infty(\mathcal{Y}_{0t}, \mathcal{X}_0) \times l^\infty(\mathcal{Y}_{0t-1}\mathcal{X}_0) \times l^\infty(\boldsymbol{\Delta}\mathcal{X}_1)$ *where* $(\mathbb{W}_{1,t}, \mathbb{W}_{1,t-1}, \mathbb{W}_{1,t-2}, \mathbb{W}_{0,t}, \mathbb{W}_{0,t-1}, \mathbb{W}^0)$ *is a tight, mean zero Gaussian process.*

Assumption 4 says that first step estimators of the distribution of the outcomes in period $t$ and $t-1$ and for both the treated and untreated groups as well as the counterfactual distribution of untreated potential outcomes for the treated group converge uniformly to a Gaussian process at the parametric rate. In the application, I estimate the distributions of outcomes conditional on covariates for each time period and each group using quantile regression. Therefore, the part of Assumption 4 that involves observed outcomes holds under standard regularity conditions on quantile regression estimators (see Assumption SB.5 in Supplementary Appendix SB for these additional conditions). Besides first step quantile regression estimators, similar results will hold for first step estimators, under standard regularity conditions in the case where there are no covariates as well as in the case where $X$ includes only discrete covariates and each conditional distribution is estimated separately for each unique value of the covariates – this is a realistic case in many applications (see the related discussion in Chernozhukov, Fernandez-Val, Hahn, and Newey (2013) and Graham, Hahn, Poirier, and Powell (2018)). The reason to state Assumption 4 as an assumption rather than as a result is that it imposes a generic limiting process for the estimator of $F_{Y_{0t}|X,D=1}$. The limiting process will depend on the particular identifying assumptions that are invoked for this distribution. In Supplementary Appendix SB, I discuss the particular case where this distribution is identified using Change in Changes as in Athey and Imbens (2006) and Melly and Santangelo (2015), but, as noted above, other approaches could be used as well.

A key intermediate step is to establish the limiting distributions of estimators of $F_{Y_{1t}|Y_{0t-1},X,D=1}$ and $F_{Y_{0t}|Y_{0t-1},X,D=1}$ (see Theorem 2). I propose estimating each of these using distribution regression (Foresi and Peracchi (1995) and Chernozhukov, Fernandez-Val, and Melly (2013)). Handling $F_{Y_{1t}|Y_{0t-1},X,D=1}$ is relatively straightforward as it amounts to a standard distribution regression of $Y_{1t}$ on $Y_{0t-1}$ and $X$ which are both observed for individuals in the treated group, and the results on limiting processes for distribution regression estimators from Chernozhukov, Fernandez-Val, and Melly (2013) can be applied directly. In particular, it immediately follows from these results that

$$\sqrt{n}(\hat{F}_{Y_{1t}|Y_{0t-1},X,D=1} - F_{Y_{1t}|Y_{0t-1},X,D=1}) \rightsquigarrow \mathbb{G}_1$$

where the exact conditions, which are standard, are given in Supplementary Appendix SB, and the exact expression for $\mathbb{G}_1$ is given in Equation (SB.5) in Supplementary Appendix SB. However, $F_{Y_{0t}|Y_{0t-1},X,D=1}$ is more challenging as it depends on applying estimated transformations to $Y_{0t-1}$ and $Y_{0t-2}$ before carrying out the distribution regression. Next, I establish the limiting process for a distribution regression estimator for this case.

**Theorem 4.** *Under the Copula Stability Assumption, Assumptions 1 to 4 and SB.1 to SB.3,*

$$\sqrt{n}\begin{pmatrix} \hat{F}_{Y_{1t}|Y_{0t-1},X,D=1} - F_{Y_{1t}|Y_{0t-1},X,D=1} \\ \hat{F}_{Y_{0t}|Y_{0t-1},X,D=1} - F_{Y_{0t}|Y_{0t-1},X,D=1} \\ \hat{F}_{Y_{0t-1},X|D=1} - F_{Y_{0t-1},X|D=1} \end{pmatrix} \rightsquigarrow \begin{pmatrix} \mathbb{G}_1 \\ \mathbb{G}_2 \\ \mathbb{G}_3 \end{pmatrix}$$

*in the space* $l^\infty(\mathcal{Y}_{1t}\mathcal{Y}_{1t-1}\mathcal{X}_1) \times l^\infty(\bar{\mathcal{Y}}_{0t}\mathcal{Y}_{1t-1}\mathcal{X}_1) \times l^\infty(\mathcal{Y}_{1t-1}\mathcal{X}_1)$ *where* $(\mathbb{G}_1, \mathbb{G}_2, \mathbb{G}_3)$ *is a tight, mean zero Gaussian process and where* $\mathbb{G}_2$ *is defined in Equation* (B.9) *in Appendix B, and* $\mathbb{G}_3$ *is defined in Equation* (SB.6) *in Appendix SB.*

Theorem 4 provides the limiting process for the second step estimators. These are important inputs into estimating the DoTT and QoTT. In particular, the lower and upper bounds on the *DoTT* and *QoTT* can be viewed as functionals of the distributions of $F_{Y_{1t}|Y_{0t-1},X,D=1}$, $F_{Y_{0t}|Y_{0t-1},X,D=1}$ and $F_{Y_{0t-1},X|D=1}$; see Theorems 2 and 3. One can define the bounds of the DoTT and QoTT as maps from these distributions. Then, the key step for the main asymptotic results in this section is to show that this map is Hadamard directionally differentiable.

Before stating the main asymptotic results, I introduce a bit more notation. For any $\theta \in l^\infty(\mathcal{Y}_\delta \boldsymbol{\Delta} \mathcal{Y}_{1t-1}\mathcal{X}_1)$, define[17]

$$\Phi^L_{\mathcal{Y}_\delta}(\theta, \delta, y', x) := \underset{y \in \mathcal{Y}_\delta}{\operatorname{argmax}} \;\; \theta(y, \delta, y', x) \quad \text{and} \quad \Phi^U_{\mathcal{Y}_\delta}(\theta, \delta, y', x) := \underset{y \in \mathcal{Y}_\delta}{\operatorname{argmin}} \;\; \theta(y, \delta, y', x)$$

and also define

$$\theta_0(y, \delta, y', x) := F_{Y_{1t}|Y_{0t-1},X,D=1}(y|y', x) - F_{Y_{0t}|Y_{0t-1},X,D=1}(y - \delta|y', x)$$

The final results of this section provide the limiting distribution for the estimators of the lower and upper bounds on the *DoTT* and *QoTT* in the main text.

[17] Following Fan and Park (2010), I use the notation $\mathcal{Y}_\delta$ to denote the compact set of values of $y$ such that the lower and upper bounds on the *DoTT* are not trivially equal to each other and to either 0 or 1; this set depends on the value of $\delta$. See the discussion in Fan and Park (2010, Section 3).

**Proposition 4.** *Under the Copula Stability Assumption, Assumptions 1 to 4 and SB.1 to SB.3,*

$$\sqrt{n}(\widehat{DoTT}^L(\delta) - DoTT^L(\delta)) \rightsquigarrow \mathbb{V}^L \quad \text{and} \quad \sqrt{n}(\widehat{DoTT}^U(\delta) - DoTT^U(\delta)) \rightsquigarrow \mathbb{V}^U$$

*where* $\mathbb{V}^L := \mathbb{V}_0^L + \mathbb{V}_1^L$ *and* $\mathbb{V}^U := \mathbb{V}_0^U + \mathbb{V}_1^U$ *which are, in turn, given by*

$$\mathbb{V}_0^L = \int_{\mathcal{Y}_{t-1}} \int_{\mathcal{X}} F^L_{Y_{1t}-Y_{0t}|Y_{0t-1},X,D=1}(\delta|y', x)\, d\mathbb{G}_3$$

$$\mathbb{V}_0^U = \int_{\mathcal{Y}_{t-1}} \int_{\mathcal{X}} F^U_{Y_{1t}-Y_{0t}|Y_{0t-1},X,D=1}(\delta|y', x)\, d\mathbb{G}_3$$

$$\mathbb{V}_1^L = \int_{\mathcal{Y}_{1,t-1}} \int_{\mathcal{X}_1} \sup_{y \in \Phi^L_{\mathcal{Y}_t}(\theta_0,\delta,y',x)} \begin{cases} 0 & \text{if } \theta_0(y,\delta,y',x) < 0 \\ \max\{\mathbb{G}_1 - \mathbb{G}_2, 0\} & \text{if } \theta_0(y,\delta,y',x) = 0 \\ \mathbb{G}_1 - \mathbb{G}_2 & \text{if } \theta_0(y,\delta,y',x) > 0 \end{cases} dF_{Y_{0t-1},X|D=1}(y', x)$$

*and*

$$\mathbb{V}_1^U = \int_{\mathcal{Y}_{1,t-1}} \int_{\mathcal{X}_1} \inf_{y \in \Phi^U_{\mathcal{Y}_t}(\theta_0,\delta,y',x)} \begin{cases} \mathbb{G}_1 - \mathbb{G}_2 & \textit{if } \theta_0(y,\delta,y',x) < 0 \\ \min\{\mathbb{G}_1 - \mathbb{G}_2, 0\} & \textit{if } \theta_0(y,\delta,y',x) = 0 \\ 0 & \textit{if } \theta_0(y,\delta,y',x) > 0 \end{cases} dF_{Y_{0t-1},X|D=1}(y', x)$$

In Proposition 4, $\mathbb{V}_0^L$ and $\mathbb{V}_0^U$ give the components of the asymptotic variance if $F_{Y_{1t}|Y_{0t-1},X,D=1}$ and $F_{Y_{0t}|Y_{0t-1},X,D=1}$ were known and did not need to be estimated in the preliminary step. $\mathbb{V}_1^L$ and $\mathbb{V}_1^U$ are additional asymptotic variance terms that come from having to estimate each of these conditional distributions.

Finally, the main asymptotic results for the estimator of the $QoTT$ follow under a small extension to the results in Proposition 4, and I state these results next.

**Theorem 5.** *Under the Copula Stability Assumption, Assumptions 1 to 4 and SB.1 to SB.3 and* *for some* $\tau$ *satisfying* $0 < \epsilon < \tau < (1-\epsilon) < 1$ *for some* $\epsilon > 0$,

$$\sqrt{n}\left(\widehat{QoTT}^{L}(\tau) - QoTT^{L}(\tau)\right) \rightsquigarrow \mathbb{Z}^{L} \quad \text{and} \quad \sqrt{n}\left(\widehat{QoTT}^{U}(\tau) - QoTT^{U}(\tau)\right) \rightsquigarrow \mathbb{Z}^{U}$$

*where*

$$\mathbb{Z}^{L} = \frac{\mathbb{V}^{U}(QoTT^{L}(\tau))}{f_{DoTT^{U}}(QoTT^{L}(\tau))}$$

*and*

$$\mathbb{Z}^{U} = \frac{\mathbb{V}^{L}(QoTT^{U}(\tau))}{f_{DoTT^{L}}(QoTT^{U}(\tau))}$$

*where* $f_{DoTT^U}$ *and* $f_{DoTT^L}$ *denote the densities of the upper bound and lower bound on the* $DoTT$, *respectively.*

The above results derive the limiting process for estimators of the bounds of the DoTT and QoTT. As an intermediate result (see the proof of Proposition 4 in Appendix B), I show that the DoTT and QoTT are Hadamard directionally differentiable functionals of the the conditional distributions discussed in Theorem 4. However, they are not fully Hadamard differentiable. The results in Fang and Santos (2019) thus imply that the standard empirical bootstrap cannot be used to conduct inference. Instead, to conduct inference in practice, I use the numerical bootstrap method proposed in Hong and Li (2018). In particular, one can proceed by constructing bootstrap estimates of the first step estimators (here, using the standard empirical bootstrap), denote these $\hat{F}^* := (\hat{F}^*_{Y_{1t}|Y_{0t-1},X,D=1}, \hat{F}^*_{Y_{0t}|Y_{0t-1},X,D=1}, \hat{F}^*_{Y_{0t-1},X|D=1})$ and denote the first step estimators themselves by $\hat{F} := (\hat{F}_{Y_{1t}|Y_{0t-1},X,D=1}, \hat{F}_{Y_{0t}|Y_{0t-1},X,D=1}, \hat{F}_{Y_{0t-1},X|D=1})$. Here, generically, let $\phi$ denote one of the maps from distribution functions to a bounds parameter of interest, and let $\mathbb{V}$ generically represent the corresponding limiting distribution in either Proposition 4 (for the DoTT) or Theorem 5 (for the QoTT). Let $\epsilon_n$ denote a tuning parameter that satisfies the conditions that $\epsilon_n \to 0$ and

$\epsilon_n\sqrt{n} \to \infty$ as $n \to \infty$. The results in Hong and Li (2018) imply that

$$\frac{\phi\Big(\hat{F} + \epsilon_n\sqrt{n}(\hat{F}^* - \hat{F})\Big) - \phi(\hat{F})}{\epsilon_n} \rightsquigarrow \mathbb{V} \tag{5.1}$$

Essentially, the idea here is to combine a numerical derivative of the map $\phi$ with bootstrapping the first step estimators.[18] The term on the left hand side of Equation (5.1) can be simulated a large number of times, and the resulting distribution will approximate the limiting distribution of whichever bounds parameter one is interested in. In the application, I report confidence intervals for the bounds on DoTT and QoTT. For a lower $(1-\alpha)$ confidence interval on $DoTT^L(\delta)$ or $QoTT^L(\tau)$, one can compute $\phi^L(\hat{F}) - \hat{c}^L_{1-\alpha}/\sqrt{n}$ where $\hat{c}^L_{1-\alpha}$ is the $(1-\alpha)$-quantile of the simulated version of the term on the left hand side of Equation (5.1) and $\phi^L$ generically denotes the map from distribution functions to either $DoTT^L$ or $QoTT^L$. Similarly, an upper $(1-\alpha)$ confidence interval for $DoTT^U(\delta)$ or $QoTT^U(\tau)$ is given by $\phi^U(\hat{F}) - \hat{c}^U_{\alpha}/\sqrt{n}$ where $\hat{c}^U_{\alpha}$ is the $\alpha$-quantile of the simulated version of the term on the left hand side of Equation (5.1) and $\phi^U$ denotes the map from distribution functions to either $DoTT^U(\delta)$ or $QoTT^U(\tau)$. This is the approach that I take to conducting inference in the application.

# 6 Job Displacement during the Great Recession

This section studies the effect of job displacement during the Great Recession on yearly earnings of late prime-age workers. Using standard techniques, job displacement is estimated to decrease workers earnings by 34% on average relative to counterfactual earnings had they not been displaced.[19] The size of this effect is quite similar to 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 relatively older 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

[18]It is also helpful to understand the term on the left hand side of Equation (5.1) by noticing that it corresponds to the standard empirical bootstrap if $\epsilon_n = n^{-1/2}$ (though this case is ruled out by the rate conditions placed on $\epsilon_n$).

[19] In this part of the paper, the outcome is the logarithm of earnings. However, because many of the effects are relatively large, I convert estimated effects from "log points" into percentage changes using $\exp(\hat{\alpha}) - 1$ where $\hat{\alpha}$ is some estimated parameter of interest in log points.

developed in the paper. First, using the panel data methods developed in the paper provides substantially more identifying power for distributional treatment effects such as the QoTT 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 group of displaced workers in the period before the Great Recession. These bounds are tight enough to be inconsistent with the assumption of rank invariance 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 QoTT under the assumption of rank invariance. The bounds also imply that some workers have higher earnings after being displaced than they would have had if they not been displaced. Finally, as discussed above, one can use the bounds on the DoTT to study the fraction of workers who experience severe negative effects of job displacement. Here, I study the fraction of workers whose earnings are at least 50% lower following job displacement than they would have been if they had not been displaced.

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 paper 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. Also, 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. Kletzer and Fairlie (2003), using NLSY data, find that displaced workers have 11% lower earnings three years after displacement than they would have had if they had not been displaced. That paper uses the same dataset as in the current paper and finds considerably smaller effects of job displacement; however, it considers the period 1984-1993 where the workers are much younger (they would have ranged from 20-36 over those years) and the economy did not experience a deep recession which likely work together to explain the large differences. Stevens (1997), using PSID data, finds that workers initially lose 25% of their earnings following job displacement and have 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.

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 (Davis and Von Wachter (2011)). 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 late prime-age workers, ages 45 to 54, the unemployment rate doubled from 3.6% to 7.1% (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 (2017)). 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. Interestingly, Farber (2017) finds much heterogeneity in the effects of job displacement. First, he finds that there are significant differences in the effect of job displacement between workers who find full-time, part-time, or remained unemployed following job displacement. Second, comparing pre- and post-displacement earnings for displaced workers, he finds a substantial fraction (around 25-40% using different approaches) of workers who are employed following job displacement have higher earnings than they did before they were displaced.

## 6.1 Data

The data comes from the 1979 National Longitudinal Survey of Youth (NLSY). The main outcome variable is the log of yearly earnings in 2011. In 2011, NLSY respondents are between 47 and 54 years old. I limit the sample to individuals who worked at least 1000 hours in 2007 and classify workers as being displaced if they left a job in 2008 or 2009 and the reason given is (i) layoff, job eliminated or (ii) company, office, or workplace closed. This excludes other reasons for leaving a job such as being fired, quitting, moving, or the end of temporary employment.

A main issue in the job displacement literature is how to treat individuals with zero earnings. Most research drops individuals with zero earnings (for example, research that uses state-level administrative data cannot tell the difference between actually having zero earnings and moving to another state). Even most research on job displacement using the PSID or NLSY tends to drop individuals with zero earnings, then take the log of earnings as the outcome variable in some regression (Stevens (1997) and Kletzer and Fairlie (2003)). I follow this same approach and restrict the sample to individuals who have at least $1,000 in earnings in 2003, 2007, and 2011. This sample includes 2,775 individuals, of whom 122 are displaced in 2008 or 2009.

Summary statistics are provided in Table 1. Displaced workers and non-displaced workers have similar education levels. There are no statistically significant differences in the fraction that have less than high school education, graduated from high school, or graduated from college. On the other hand, there are larger differences in race; displaced workers are 10 percentage points more likely to be black (p-value: 0.02).

Table 1 also contains the path of average earnings for displaced and non-displaced workers from 2001-2013. In 2001, average yearly earnings for displaced workers was about $43,000 while average earnings for non-displaced workers was about $47,000 (the p-value on the difference is 0.40). From 2001-2007, which are all pre-displacement years, the gap remains roughly constant and does not display any obvious trend over time. However, by 2009, displaced workers have experienced a large decline in average earnings. The gap in 2009 is roughly $9,000 and almost all of this is explained by the decline in earnings of the group of displaced workers. By 2011, the gap in earnings has increased to almost $22,000; about a third of the increase in the gap is due to increased earnings for the group of non-displaced workers and the remaining part of the increase in the gap is due to decreased earnings of displaced workers.[20] The gap is still close to $20,000 in 2013.

## 6.2 Baseline Results

In this section, I estimate the average effect of job displacement on the earnings of late prime-age workers using the Change in Changes model. The results indicate that late prime-age workers lose 34% of their earnings on average due to job displacement. This effect is quite similar 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) and Von Wachter, Song, and Manchester (2009)) which are the largest in the literature. This estimate is broadly similar to the estimated effect of job displacement on all workers during the Great Recession reported in Farber (2017) and using the Displaced Workers Survey.

I also tried several alternative approaches including Difference in Differences, Difference in Differences where the trend can depend on covariates as in Heckman, Ichimura, and Todd (1997) and Abadie (2005), and the Change in Changes model with covariates. The results are very similar in all cases with estimates ranging from 30% lower earnings to 35% lower earnings using these alternative approaches.

## 6.3 The Distributional Effects of Job Displacement

This section uses the techniques developed earlier in the paper to understand the distributional effects of job displacement. I focus on estimating the QoTT, the fraction of workers who have higher earnings following displacement, and the fraction of workers who experience a severe negative effect of job displacement.

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.

[20] Recall that workers are displaced in either 2008 or 2009, so earnings in 2009 may mix pre-displacement earnings for workers who are displaced at some point in 2009 with their post-displacement earnings which may account for the smaller earnings gap in 2009 than in 2011 or 2013.

**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 Change in Changes method of Athey and Imbens (2006) though there are a variety of other methods that could be used for the first step estimation including the Panel DID method of Callaway and Li (2019), Quantile Difference in Differences (Athey and Imbens (2006)), or by assuming selection on observables holds after conditioning on a lag of earnings (Firpo (2007)). Figure SC.3 (in the Supplementary Appendix) plots the QTT under each of these alternative assumptions. The QTT results are not sensitive to changes in the underlying model for the counterfactual distribution of untreated potential outcomes. This result also implies that the results for parameters that depend on the joint distribution of potential outcomes are not sensitive to the choice of identifying assumption in the first stage.

Estimates of both marginal distributions are presented in Figure 2, and an estimate of the QTT is presented in Figure 3. The QTT is negative everywhere and statistically significant for all quantiles except the very largest (primarily due to standard errors increasing). The differences between the marginal distributions of observed outcomes for displaced workers and the distribution of outcomes that displaced workers would have experienced if they had not been displaced are large. The QTT is also increasing across quantiles. It is tempting to interpret Figure 3 as indicating that the effect of job displacement is largest for individuals at the lower part of the earnings distribution. However, as mentioned above, job displacement appears to cause many individuals to change their rank in the distribution of earnings which makes this interpretation of the QTT invalid. In fact, the QTT is also consistent with the effect of job displacement being very heterogeneous across individuals; that is, upon job displacement some individuals lose a large fraction of their earnings while others experience less severe effects.

**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 paper to estimate some parameters that depend on the joint distribution of treated and untreated potential outcomes. First, I consider the QoTT. Figure 4 plots bounds on the QoTT under no assumptions on the dependence between potential outcome distributions. There are several things to notice from the figure. First, immediately these bounds imply that there are heterogeneous effects – the upper bound on the 5th percentile of the effect of job displacement is estimated to be 56% lower earnings while the lower bound on the 95th percentile of the effect of job displacement is estimated to be 20% lower earnings. Finding that there is treatment effect heterogeneity is important here as it implies that the effect of job displacement can potentially be more severe for some displaced workers than the average effect would indicate. These bounds also imply that at least 5.9% of displaced workers (and up to 63.0% of displaced workers) lose more than half of their earnings relative to what they would have earned if they had not been displaced. However, these bounds are relatively less informative about other parameters of interest. For example, the median of the treatment effect is bounded to be between an earnings loss of 61% and an earnings gain of 65%. The bounds are consistent with up to 81% of displaced workers having higher earnings following displacement than they would have had if they had not been displaced, but they are also consistent with all displaced workers having lower earnings following displacement than they would have had if they had not been displaced. To better understand heterogeneous effects of job displacement, it would be helpful to have tighter bounds on these parameters. It is also worth mentioning that, although the confidence intervals appear small in the Figure 4, it is actually more the case that the width of the bounds stemming from the identification arguments is wide relative to the uncertainty due to estimation. For example, as mentioned above, the point estimate for the upper bound of the median of the treatment effect is 65% higher earnings, and the corresponding one-sided upper 95% confidence interval is 79% higher earnings.

Next, Figure 5 provides bounds under the Copula Stability Assumption. These bounds are indeed tighter. Earnings losses at the 5th percentile are between 73% and 96% which implies that, at least for some individuals, the effect of job displacement is much worse than the average effect. Similarly, the bounds imply that at least 9.0% of displaced workers earn less than half of what they would have earned if they had not been displaced. Interestingly, one can also conclude that at least 10.1% of individuals have higher earnings after being displaced than they would have had they not been displaced (the corresponding one-sided lower confidence interval is 4.0%). This type of conclusion was not available without exploiting the Copula Stability Assumption and is what makes the Copula Stability Assumption incompatible with the assumption of Monotone Treatment Response in the current case. This result provides another indication that the effect of job displacement is quite heterogeneous across workers.

Next, it is interesting to compare the results in the application to results that exploit having access to covariates. As discussed above, covariates can be used to tighten bounds on distributional treatment effect parameters, and they can also be used to tighten the bounds under the Copula Stability Assumption. Here, I use education levels, race, and gender as covariates. These results are available in Figure 6. Covariates do indeed tighten the bounds relative to the bounds that only use information on the marginal distributions of treated and untreated potential outcomes (this can be seen in the bottom-right panel of Figure 6); however, overall the effect is fairly moderate – the upper bound on the QoTT becomes somewhat tighter and the lower bound is basically the same. When covariates are used to further tighten the bounds that come from the Copula Stability Assumption, essentially the same pattern holds – the upper bound on the QoTT becomes somewhat tighter and the lower bound remains essentially the same. Finally, it is useful to compare the bounds (i) under the Copula Stability Assumption without covariates to the bounds using (ii) covariates but not the Copula Stability Assumption. Here, the bounds that use the Copula Stability Assumption are noticeably tighter. This indicates that, at least in the current application, exploiting panel data through the Copula Stability Assumption is particularly useful in deriving tighter bounds on the distributional treatment effects that depend on the joint distribution of potential outcomes.

Finally, Figure 7 plots the QoTT under several assumptions that would lead to point identification. First, it plots the QoTT under rank invariance 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 69% from being displaced. At the 95th percentile, they lose 12%. At the median, they lose 25%, and this effect is largely constant across most of the interior quantiles. Of course, the no-assumptions bounds cannot rule out rank invariance between $Y_{1t}$ and $Y_{0t}$, but, here, this kind of rank invariance is incompatible with the Copula Stability Assumption because the bounds imply more heterogeneity than occurs under rank invariance (see Figure 7).

The second part of Figure 7 plots the results under the assumption of rank invariance between $Y_{0t}$ and $Y_{0t-1}$. This assumption results in considerably more heterogeneity in the effect of job displacement than the assumption of rank invariance between $Y_{1t}$ and $Y_{0t}$. For example, at the 5th percentile, the estimated effect of job displacement is a loss of 89% of earnings. At the median, the estimated effect is 29% lower earnings per year. And at the 95th percentile, earnings are estimated to be 80% higher than they would have been without job displacement. Further, 30% of displaced workers are estimated to have higher earnings than they would have had they not been displaced, and 22% of displaced workers are estimated to have lost at least half of their earnings due to job displacement relative to what they would have earned if they had not been displaced. The estimates of the QoTT under the assumption of rank invariance over time fall completely within the bounds on the QoTT under the Copula Stability Assumption. Despite this, as discussed earlier, the assumption of rank invariance over time can be "pre-tested" (i.e., one can test if non-displaced potential earnings were rank invariant over time in pre-treatment periods). Spearman's Rho for $Y_{0t-1}$ and $Y_{0t-2}$ is estimated to be 0.79 for the group of displaced workers. This indicates that the assumption of rank invariance over time is rejected in the pre-treatment period and is therefore not likely to be valid in the current period either. The reason that the bounds in the current paper are close to the estimates of the QoTT under this point identifying assumption is that strong positive dependence, though not rank invariance, is observed between $Y_{0t-1}$ and $Y_{0t-2}$ for the group of displaced workers (Spearman's Rho = 0.79); the dependence between $Y_{1t}$ and $Y_{0t-1}$ is positive but smaller (Spearman's Rho = 0.54).

**Does the Copula Stability Assumption hold for job displacement?**

Figure 8 plots Spearman's Rho for $Y_{0s}$ and $Y_{0s-1}$ in periods before the Great Recession separately for the treated group and untreated group. Here, as above, periods are earnings separated by 4 years starting from 1983 and going to 2011. Spearman's Rho is broadly similar for both groups. For the group of displaced workers, Spearman's Rho is increasing from 1987 (about 0.5) to 1999 (about 0.75) and is essentially constant from 1999 - 2007. In 1987, individuals in the NLSY were between 23 and 30 years old. In 1999, they were between 35 and 42 years old. Figure 8 suggests that individuals moved around in the distribution of earnings more in their 20s, that there is less earnings mobility starting in their 30s and that the amount of mobility is the same through their 40s and early 50s. That Spearman's Rho is constant from 1999-2007 provides a strong piece of evidence in favor of the Copula Stability Assumption.

# 7 Conclusion

There is a long history in economics of exploiting access to panel data to identify parameters of interest when there is unobserved heterogeneity that affects outcomes. This paper has developed a new approach to deriving tighter bounds on distributional treatment effect parameters that depend on the joint distribution of potential outcomes in the presence of panel data. The results depend on three key ingredients: (i) access to at least three periods of panel data, (ii) identification of the marginal distribution of 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 late prime-age workers. Using standard techniques, I find that these workers lose on average 34% of their yearly earnings following job displacement. Using the techniques developed in the current paper, I find that this average effect masks substantial heterogeneity: some 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 if they had not been displaced. Having access to panel data and using the approach in the paper led to substantially tighter bounds on the partially identified parameters considered in the paper.

# References

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

[2] Jaap Abbring 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.

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

[4] Marianne P Bitler, Jonah B Gelbach, and Hilary W Hoynes. "What mean impacts miss: Distributional effects of welfare reform experiments". *American Economic Review* 96.4 (2006), pp. 988–1012.

[5] Marianne P Bitler, Jonah B Gelbach, and Hilary W Hoynes. "Can variation in subgroups' average treatment effects explain treatment effect heterogeneity? Evidence from a social experiment". *Review of Economics and Statistics* 99.4 (2017), pp. 683–697.

[6] Richard Blundell and Monica Costa Dias. "Alternative approaches to evaluation in empirical microeconomics". *Journal of Human Resources* 44.3 (2009), pp. 565–640.

[7] Richard Blundell, Amanda Gosling, Hidehiko Ichimura, and Costas Meghir. "Changes in the distribution of male and female wages accounting for employment composition using bounds". *Econometrica* 75.2 (2007), pp. 323–363.

[8] Stephane Bonhomme 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.

[9] Brantly Callaway and Tong Li. "Quantile treatment effects in difference in differences models with panel data". Working Paper. 2019.

[10] Brantly Callaway, Tong Li, and Tatsushi Oka. "Quantile treatment effects in difference in differences models under dependence restrictions and with only two time periods". *Journal of Econometrics* 206.2 (2018), pp. 395–413.

[11] Brantly Callaway and Pedro HC Sant'Anna. "Difference-in-differences with multiple time periods". Working Paper. 2019.

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

[13] Pedro Carneiro, 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.

[14] Robert Carroll, David Joulfaian, and Mark Rider. "Income mobility: The recent American experience". Working Paper, Andrew Young School of Policy Studies Research Paper Series. 2007.

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

[16] Victor Chernozhukov, Ivan Fernandez-Val, and Blaise Melly. "Inference on counterfactual distributions". *Econometrica* 81.6 (2013), pp. 2205–2268.

[17] Victor Chernozhukov, Sokbae Lee, and Adam M Rosen. “Intersection bounds: Estimation and inference”. *Econometrica* 81.2 (2013), pp. 667–737.

[18] Raj Chetty, 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.

[19] Kenneth A Couch and Dana W Placzek. “Earnings losses of displaced workers revisited”. *The American Economic Review* (2010), pp. 572–589.

[20] Steven J Davis and Till Von Wachter. “Recessions and the costs of job loss”. *Brookings Papers on Economic Activity* 43.2 (Fall) (2011), pp. 1–72.

[21] Habiba Djebbari and Jeffrey Smith. “Heterogeneous impacts in PROGRESA”. *Journal of Econometrics* 145.1 (2008), pp. 64–80.

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

[23] Yingying Dong and Shu Shen. “Testing for rank invariance or similarity in program evaluation”. *Review of Economics and Statistics* 100.1 (2018), pp. 78–85.

[24] Greg J Duncan, Richard Coe, Mary Corcoran, Martha S Hill, Saul Hoffman, and James N Morgan. *Years of poverty–years of plenty: The changing economic fortunes of American workers and families*. 1984.

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

[26] Yanqin Fan, Emmanuel Guerre, and Dongming Zhu. “Partial identification of functionals of the joint distribution of potential outcomes”. *Journal of Econometrics* 197.1 (2017), pp. 42–59.

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

[28] Yanqin Fan and Sang Soo Park. “Sharp bounds on the distribution of treatment effects and their statistical inference”. *Econometric Theory* 26.03 (2010), pp. 931–951.

[29] Yanqin Fan and Sang Soo Park. “Confidence intervals for the quantile of treatment effects in randomized experiments”. *Journal of Econometrics* 167.2 (2012), pp. 330–344.

[30] Yanqin Fan and Jisong Wu. “Partial identification of the distribution of treatment effects in switching regime models and its confidence sets”. *The Review of Economic Studies* 77.3 (2010), pp. 1002–1041.

[31] Zheng Fang and Andres Santos. “Inference on directionally differentiable functions”. *The Review of Economic Studies* 86.1 (2019), pp. 377–412.

[32] Henry S Farber. “The changing face of job loss in the United States, 1981-1995”. *Brookings Papers on Economic Activity. Microeconomics* (1997), pp. 55–142.

[33] Henry S Farber. “Employment, hours, and earnings consequences of job loss: US evidence from the displaced workers survey”. *Journal of Labor Economics* 35.S1 (2017), S235–S272.

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

[35] Sergio Firpo and Geert Ridder. “Partial identification of the treatment effect distribution and its functionals”. *Journal of Econometrics* (2019).

[36] Silverio Foresi and Franco Peracchi. “The conditional distribution of excess returns: An empirical analysis”. *Journal of the American Statistical Association* 90.430 (1995), pp. 451–466.

[37] Brigham R Frandsen and Lars J Lefgren. “Partial identification of the distribution of treatment effects with an application to the Knowledge Is Power Program (KIPP)”. Working Paper. 2017.

[38] Brigham R Frandsen and Lars J Lefgren. “Testing rank similarity”. *Review of Economics and Statistics* 1 (2018), pp. 86–91.

[39] Maurice Frank, 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.

[40] Maurice Fréchet. “Sur les tableaux de correlation dont les marges sont donnees”. *Annals University Lyon: Series A* 14 (1951), pp. 53–77.

[41] Joachim Freyberger. “Nonparametric panel data models with interactive fixed effects”. *Review of Economic Studies* 85.3 (2018), pp. 1824–1851.

[42] Michael Gechter. “Generalizing the results from social experiments: Theory and evidence from Mexico and India”. Working Paper. 2016.

[43] Peter Gottschalk. “Inequality, income growth, and mobility: The basic facts”. *The Journal of Economic Perspectives* (1997), pp. 21–40.

[44] Bryan S Graham, Jinyong Hahn, Alexandre Poirier, and James L Powell. “A quantile correlated random coefficients panel data model”. *Journal of Econometrics* 206.2 (2018), pp. 305–335.

[45] James J Heckman. “Micro data, heterogeneity, and the evaluation of public policy: Nobel lecture”. *Journal of Political Economy* 109.4 (2001), pp. 673–748.

[46] James Heckman, 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.

[47] James Heckman, 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.

[48] Wassily Hoeffding. “Massstabinvariante korrelationstheorie”. *Schriften des Mathematischen Instituts und des Institutes Fur Angewandte Mathematik der Universitat Berlin* 5 (1940), pp. 179–233.

[49] Han Hong and Jessie Li. “The numerical delta method”. *Journal of Econometrics* 206.2 (2018), pp. 379–394.

[50] Thomas L Hungerford. “US income mobility in the seventies and eighties”. *Review of Income and Wealth* 39.4 (1993), pp. 403–417.

[51] Guido Imbens and Jeffrey Wooldridge. “Recent developments in the econometrics of program evaluation”. *Journal of Economic Literature* 47.1 (2009), pp. 5–86.

[52] Louis S Jacobson, Robert J LaLonde, and Daniel G Sullivan. “Earnings losses of displaced workers”. *The American Economic Review* (1993), pp. 685–709.

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

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

[55] Ju Hyun Kim. “Identifying the distribution of treatment effects under support restrictions”. Working Paper. 2018.

[56] Lori G Kletzer and Robert W Fairlie. “The long-term costs of job displacement for young adult workers”. *ILR Review* 56.4 (2003), pp. 682–698.

[57] Patrick Kline and Andres Santos. “Sensitivity to missing data assumptions: Theory and an evaluation of the US wage structure”. *Quantitative Economics* 4.2 (2013), pp. 231–267.

[58] Patrick Kline and Melissa Tartari. “Bounding the labor supply responses to a randomized welfare experiment: A revealed preference approach”. *American Economic Review* 106.4 (2016), pp. 972–1014.

[59] Wojciech Kopczuk, 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.

[60] Erich Lehmann. *Nonparametrics: Statistical Methods Based on Ranks*. Holden-Day, San Francisco, 1974.

[61] Tong Li and Quang Vuong. “Nonparametric estimation of the measurement error model using multiple indicators”. *Journal of Multivariate Analysis* 65.2 (1998), pp. 139–165.

[62] GD Makarov. “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.

[63] Charles Manski. “Nonparametric bounds on treatment effects”. *The American Economic Review* (1990), pp. 319–323.

[64] Charles Manski. “Monotone treatment response”. *Econometrica* (1997), pp. 1311–1334.

[65] Charles Manski and John Pepper. “Monotone instrumental variables: with an application to the returns to schooling”. *Econometrica* 68.4 (2000), pp. 997–1010.

[66] Matthew A Masten and Alexandre Poirier. “Inference on breakdown frontiers”. Working Paper. 2019.

[67] Blaise Melly and Giulia Santangelo. “The changes-in-changes model with covariates”. Working Paper. 2015.

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

[69] Jonathan Roth. “Should we adjust for the test for pre-trends in difference-in-difference designs?” Working Paper. 2018.

[70] Ludger Rüschendorf. “Random variables with maximum sums”. *Advances in Applied Probability* (1982), pp. 623–632.

[71] Bertrand Ferdinand Schriever et al. “An ordering for positive dependence”. *The Annals of Statistics* 15.3 (1987), pp. 1208–1214.

[72] Alexander Shapiro. “On concepts of directional differentiability”. *Journal of optimization theory and applications* 66.3 (1990), pp. 477–487.

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

[74] Ann Huff Stevens. “Persistent effects of job displacement: The importance of multiple job losses”. *Journal of Labor Economics* (1997), pp. 165–188.

[75] U.S. Bureau of Labor Statistics. *All Employees: Total Nonfarm Payrolls [PAYEMS]*. 2015.

[76] U.S. Bureau of Labor Statistics. *Civilian Unemployment Rate [UNRATE]*. 2015.

[77] U.S. Bureau of Labor Statistics. *Unemployment Rate: 45 to 54 Years [LNU04000093]*. Retrieved from FRED, Federal Reserve Bank of St. Louis on August 9, 2017. 2015.

[78] Aad W van der Vaart and Jon A Wellner. *Weak Convergence and Empirical Processes*. Springer Science & Business Media, 1996.

[79] Till Von Wachter, 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).

[80] Robert Williamson and Tom Downs. “Probabilistic arithmetic. I. Numerical methods for calculating convolutions and dependency bounds”. *International Journal of Approximate Reasoning* 4 (1990), pp. 89–158.

[81] Kaspar Wüthrich. “A closed-form estimator for quantile treatment effects with endogeneity”. *Journal of Econometrics* 210.2 (2019), pp. 219–235.

[82] Takemi Yanagimoto and Masashi Okamoto. “Partial orderings of permutations and monotonicity of a rank correlation statistic”. *Annals of the Institute of Statistical Mathematics* 21.1 (1969), pp. 489–506.

# A Proofs

## A.1 Proof of Lemma 1

The first part holds under the Copula Stability Assumption as follows

$$
\begin{aligned}
F_{Y_{0t},Y_{0t-1}|D=1}(y_0,y') &= C_{Y_{0t},Y_{0t-1}|D=1}(F_{Y_{0t}|D=1}(y_0),F_{Y_{0t-1}|D=1}(y')) \\
&= C_{Y_{0t-1},Y_{0t-2}|D=1}\Big(F_{Y_{0t}|D=1}(y_0),F_{Y_{0t-1}|D=1}(y')\Big) \\
&= F_{Y_{0t-1},Y_{0t-2}|D=1}\Big(F^{-1}_{Y_{0t-1}|D=1}\circ F_{Y_{0t}|D=1}(y_0),F^{-1}_{Y_{0t-2}|D=1}\circ F_{Y_{0t-1}|D=1}(y')\Big)
\end{aligned}
$$

where the first equality holds from Sklar's Theorem, the second from the Copula Stability Assumption and the third holds from the definition of a copula.

For the second part, start with

$$
\begin{aligned}
&F_{Y_{0t}|Y_{0t-1},D=1}(y_0|y') \\
&\quad= \int_{\mathcal{Y}} \mathbb{1}\{\tilde{y}_0\le y_0\} f_{Y_{0t}|Y_{0t-1},D=1}(\tilde{y}_0 \mid y')\, \mathrm{d}\tilde{y}_0 \\
&\quad= \int_{\mathcal{Y}} \mathbb{1}\{\tilde{y}_0\le y_0\} \frac{f_{Y_{0t},Y_{0t-1},D=1}(\tilde{y}_0,y')}{f_{Y_{0t-1}|D=1}(y')}\, \mathrm{d}\tilde{y}_0 \\
&\quad= \int_{\mathcal{Y}} \mathbb{1}\{\tilde{y}_0\le y_0\} c_{Y_{0t},Y_{0t-1}|D=1}(\mathrm{F}_{Y_{0t}|D=1}(\tilde{y}_0),\mathrm{F}_{Y_{0t-1}|D=1}(y'))f_{Y_{0t}|D=1}(\tilde{y}_0)\, \mathrm{d}\tilde{y}_0 \\
&\quad= \int_{\mathcal{Y}} \mathbb{1}\{\tilde{y}_0\le y_0\} c_{Y_{0t-1},Y_{0t-2}|D=1}(\mathrm{F}_{Y_{0t}|D=1}(\tilde{y}_0),\mathrm{F}_{Y_{0t-1}|D=1}(y'))f_{Y_{0t}|D=1}(\tilde{y}_0)\, \mathrm{d}\tilde{y}_0 \\
&\quad= \int_{\mathcal{Y}} \mathbb{1}\{\tilde{y}_0\le y_0\} f_{Y_{0t-1},Y_{0t-2}|D=1}(\mathrm{F}^{-1}_{Y_{0t-1}|D=1}(\mathrm{F}_{Y_{0t}|D=1}(\tilde{y}_0)),\mathrm{F}^{-1}_{Y_{0t-2}|D=1}(\mathrm{F}_{Y_{0t-1}|D=1}(y'))) \\
&\qquad\quad \times \frac{f_{Y_{0t}|D=1}(\tilde{y}_0)}{f_{Y_{0t-1}|D=1}(\mathrm{F}^{-1}_{Y_{0t-1}|D=1}(\mathrm{F}_{Y_{0t}|D=1}(\tilde{y}_0)))\times f_{Y_{0t-2}|D=1}(\mathrm{F}^{-1}_{Y_{0t-2}|D=1}(\mathrm{F}_{Y_{0t-1}|D=1}(y')))}\, \mathrm{d}\tilde{y}_0 \\
&\quad= \int_{\mathcal{Y}} \mathbb{1}\{\tilde{y}_0\le y_0\} f_{Y_{0t-1}|Y_{0t-2},D=1}(\mathrm{F}^{-1}_{Y_{0t-1}|D=1}(\mathrm{F}_{Y_{0t}|D=1}(\tilde{y}_0))|\,\mathrm{F}^{-1}_{Y_{0t-2}|D=1}(\mathrm{F}_{Y_{0t-1}|D=1}(y'))) \\
&\qquad\quad \times \frac{f_{Y_{0t}|D=1}(\tilde{y}_0)}{f_{Y_{0t-1}|D=1}(\mathrm{F}^{-1}_{Y_{0t-1}|D=1}(\mathrm{F}_{Y_{0t}|D=1}(\tilde{y}_0)))}\, \mathrm{d}\tilde{y}_0
\end{aligned}
$$

where the first two equalities hold immediately, the third equality writes the joint density in terms of the copula and the marginal densities, the fourth equality uses the Copula Stability Assumption, the fifth equality converts the copula back into a joint density, and the sixth converts the joint density into a conditional density. Next, make the substitution $u=\mathrm{F}^{-1}_{Y_{0t-1}|D=1}(\mathrm{F}_{Y_{0t}|D=1}(\tilde{y}_0))$ which implies

$$
\tilde{y}_0 = \mathrm{F}^{-1}_{Y_{0t}|D=1}(\mathrm{F}_{Y_{0t-1}|D=1}(u))
$$

and

$$
\mathrm{d}\tilde{y}_0 = \frac{f_{Y_{0t-1}|D=1}(u)}{f_{Y_{0t}|D=1}(\mathrm{F}^{-1}_{Y_{0t}|D=1}(\mathrm{F}_{Y_{0t-1}|D=1}(u)))}\, du
$$

Plugging these back in implies

$$F_{Y_{0t}|Y_{0t-1},D=1}(y_0|y') = \int_{\mathcal{Y}} \mathbb{1}\{u \leq F^{-1}_{Y_{0t-1}|D=1}(F_{Y_{0t}|D=1}(y_0))\} f_{Y_{0t-1}|Y_{0t-2},D=1}(u|\, \mathrm{F}^{-1}_{Y_{0t-2}|D=1}(\mathrm{F}_{Y_{0t-1}|D=1}(y'))) \; du$$
$$= F_{Y_{0t-1}|Y_{0t-2},D=1}(F^{-1}_{Y_{0t-1}|D=1}(F_{Y_{0t}|D=1}(y_0))|\, \mathrm{F}^{-1}_{Y_{0t-2}|D=1}(\mathrm{F}_{Y_{0t-1}|D=1}(y')))$$

which completes the proof.

## A.2 Proof of Lemma 2

Lemma 2 follows by an application of the Fréchet-Hoeffding bounds to a conditional bivariate distribution.

## A.3 Proof of Lemma 3

Lemma 3 applies the sharp bounds on the difference between random variables with known marginal distributions but unknown copula of Williamson and Downs (1990) (See also: Makarov (1982), Rüschendorf (1982), Frank, Nelsen, and Schweizer (1987), and Fan and Park (2010)) to the difference conditional on the previous outcome.

## A.4 Proofs of Theorem 1 and Theorem 2

Theorem 1 and Theorem 2 follow from results in Fan and Park (2010, Section 5), Fan, Guerre, and Zhu (2017), and Firpo and Ridder (2019) which derive 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 paper, the marginal distributions are conditional on a lag of the outcome $Y_{0t-1}$.

## A.5 Proof of Theorem 3

Theorem 3 holds because inverting sharp bounds on a distribution implies sharp bounds on the quantiles (Williamson and Downs (1990) and Fan and Park (2010)).

## A.6 Proof of Proposition 1

To simplify notation throughout the proof, let $F_{1|0} = F_{Y_{1t}|Y_{0t-1},D=1}$ and let $F_{t-1} = F_{Y_{0t-1}|D=1}$.

**Upper Bound:**

Let $F_1^U(y_0, y_1)$ denote the upper bound on the joint distribution given in Theorem 1 when $F_{Y_{0t}|Y_{0t-1},D=1} = F_1$ and likewise let $F_2^U(y_0, y_1)$ denote the upper bound when $F_{Y_{0t}|Y_{0t-1},D=1} = F_2$ and under the conditions stated in the proposition. The goal is to show that $F_1^U(y_0, y_1) - F_2^U(y_0, y_1) \geq 0$ for all $(y_0, y_1) \in \mathcal{Y} \times \mathcal{Y}$ where, for simplicity, I suppose that $Y_{1t}$, $Y_{0t}$, and $Y_{0t-1}$ have common support $\mathcal{Y}$ conditional on $D = 1$. I also suppose (as elsewhere in the paper) that $\mathcal{Y}$ is compact, and that $Y_{1t}$ and $Y_{0t}$ are uniformly continuously distributed conditional on $Y_{0t-1}$ and $D = 1$.

First, it is straightforward to show that

$$
\begin{aligned}
F_1^U(y_0, y_1) &- F_2^U(y_0, y_1) \\
&= \int_{\mathcal{Y}} (F_1(y_0|y') - F_2(y_0|y'))\mathbb{1}\{F_1(y_0|y') \leq F_{1|0}(y_1|y'), F_2(y_0|y') \leq F_{1|0}(y_1|y')\} \, \mathrm{d}F_{t-1}(y') \\
&+ \int_{\mathcal{Y}} (F_1(y_0|y') - F_{1|0}(y_1|y'))\mathbb{1}\{F_1(y_0|y') \leq F_{1|0}(y_1|y'), F_2(y_0|y') > F_{1|0}(y_1|y')\} \, \mathrm{d}F_{t-1}(y') \\
&+ \int_{\mathcal{Y}} (F_{1|0}(y_1|y') - F_2(y_0|y'))\mathbb{1}\{F_1(y_0|y') > F_{1|0}(y_1|y'), F_2(y_0|y') \leq F_{1|0}(y_1|y')\} \, \mathrm{d}F_{t-1}(y')
\end{aligned} \tag{A.1}
$$

which holds from taking differences in the the conditional bounds on the joint distribution as in Lemma 1. Next, $F_{1|0} \prec^{SI} F_1 \prec^{SI} F_2$ implies that there exist $y_1^* = y_1^*(y_0, y_1)$ and $y_2^* = y_2^*(y_0, y_1)$ such that

$$
F_{1|0}(y_1|y_1^*) = F_1(y_0|y_1^*) \quad \text{and} \quad F_{1|0}(y_1|y_2^*) = F_2(y_0|y_2^*)
$$

and for $y' \in \mathcal{Y}$ such that

$$
y' \geq y_1^* \implies F_{1|0}(y_1|y') \geq F_1(y_0|y'), \qquad y' < y_1^* \implies F_{1|0}(y_1|y') < F_1(y_0|y')
$$

and

$$
y' \geq y_2^* \implies F_{1|0}(y_1|y') \geq F_2(y_0|y'), \qquad y' < y_2^* \implies F_{1|0}(y_1|y') < F_2(y_0|y')
$$

Note that the set of $y' \in \mathcal{Y}$ satisfying $y' < y_j^*$, for $j = 1, 2$ can be empty (particularly in the case where $y_j^* = y_{min}$ where $y_{min}$ is the smallest value in $\mathcal{Y}$). Plugging these into Equation A.1 implies that the difference in Equation A.1 is given by

$$
\int_{\mathcal{Y}} (F_1(y_0|y') - F_2(y_0|y'))\mathbb{1}\{y' \geq y_1^*, y' \geq y_2^*\} \, \mathrm{d}F_{t-1}(y') \tag{A.2}
$$

$$
+ \int_{\mathcal{Y}} (F_1(y_0|y') - F_{1|0}(y_1|y'))\mathbb{1}\{y_1^* \leq y' < y_2^*\} \, \mathrm{d}F_{t-1}(y') \tag{A.3}
$$

$$
+ \int_{\mathcal{Y}} (F_{1|0}(y_1|y') - F_2(y_0|y'))\mathbb{1}\{y_2^* \leq y' < y_1^*\} \, \mathrm{d}F_{t-1}(y') \tag{A.4}
$$

*Case 1:* $y_1^* \leq y_2^*$. In this case, Equation A.4 is equal to 0. Moreover, in Equation A.3, since $y'$ is restricted to be less than $y_2^*$, $F_{1|0}(y_1|y') < F_2(y_0|y')$ over this range; using this inequality and combining the terms from Equations (A.2) and (A.3) implies

$$
\begin{aligned}
F_1^U(y_0, y_1) - F_2^U(y_0, y_1) &\geq \int_{\mathcal{Y}} (F_1(y_0|y') - F_2(y_0|y'))\mathbb{1}\{y' \geq y_1^*\} \, \mathrm{d}F_{t-1}(y') \\
&= F_2(y_0, y_1^*) - F_1(y_0, y_1^*) \geq 0
\end{aligned}
$$

where the last inequality holds because $F_1 \prec^{SI} F_2 \implies F_1 \prec^{C} F_2$ (Joe (1997, Theorem 2.12)) where $F \prec^{C} G$ is the concordance ordering and indicates that $F(x_1, x_2) \leq G(x_1, x_2)$ for all $x_1$ and $x_2$ and where $F$ and $G$ are joint distributions with the same marginals.

*Case 2:* $y_1^* > y_2^*$. In this case, Equation A.3 is equal to 0. Also, Equation A.4 is greater than or equal to zero because $y'$ is restricted to be greater than $y_2^*$. Equation A.2 is then given by

$$\int_{\mathcal{Y}} (F_1(y_0|y') - F_2(y_0|y'))\mathbb{1}\{y' \geq y_1^*\} \, \mathrm{d}F_{t-1}(y') = F_2(y_0, y_1^*) - F_1(y_0, y_1^*) \geq 0$$

where the first part holds because $y_1^* > y_2^*$ and where the last inequality holds because $F_1 \prec^C F_2$. Therefore, in this case, all the terms in Equations (A.2) to (A.4) are greater than or equal to zero and therefore the result holds.

**Lower Bound:**

Let $F_1^L(y_0, y_1)$ denote the upper bound of the joint distribution given in Theorem 1 when $F_{Y_{0t}|Y_{0t-1},D=1} = F_1$ and likewise let $F_2^L(y_0, y_1)$ denote the upper bound when $F_{Y_{0t}|Y_{0t-1},D=1} = F_2$. For this part, the goal is to show that $F_2^L(y_0, y_1) - F_1^L(y_0, y_1) \geq 0$ for all $(y_0, y_1) \in \mathcal{Y} \times \mathcal{Y}$. Similar to the case for the upper bound, one can show that

$$\begin{aligned}
&F_2^L(y_0, y_1) - F_1^L(y_0, y_1) \\
&\quad = \int_{\mathcal{Y}} (F_2(y_0|y') - F_1(y_0|y'))\mathbb{1}\{F_{1|0}(y_1|y') + F_1(y_0|y') - 1 \geq 0, F_{1|0}(y_1|y') + F_2(y_0|y') - 1 \geq 0\} \, dF_{t-1}(y') \\
&\quad - \int_{\mathcal{Y}} (F_{1|0}(y_1|y') + F_1(y_0|y') - 1)\mathbb{1}\{F_{1|0}(y_1|y') + F_1(y_0|y') - 1 \geq 0, F_{1|0}(y_1|y') + F_2(y_0|y') - 1 < 0\} \, dF_{t-1}(y') \\
&\quad + \int_{\mathcal{Y}} (F_{1|0}(y_1|y') + F_2(y_0|y') - 1)\mathbb{1}\{F_{1|0}(y_1|y') + F_1(y_0|y') - 1 < 0, F_{1|0}(y_1|y') + F_2(y_0|y') - 1 \geq 0\} \, dF_{t-1}(y')
\end{aligned} \tag{A.5}$$

which holds by the same arguments as for the upper bound, but now using the conditional lower bounds on the joint distribution. Because $F_{1|0}$, $F_1$, and $F_2$ are all stochastically increasing, there exist $y_1^\dagger = y_1^\dagger(y_0, y_1)$ and $y_2^\dagger = y_2^\dagger(y_0, y_1)$ such that

$$F_{1|0}(y_1|y_1^\dagger) + F_1(y_0|y_1^\dagger) = 1 \quad \textit{and} \quad F_{1|0}(y_1|y_2^\dagger) + F_2(y_0|y_2^\dagger) = 1$$

and for $y' \in \mathcal{Y}$ such that

$$y' > y_1^\dagger \implies F_{1|0}(y_1|y') + F_1(y_0|y') < 1, \qquad y' \leq y_1^\dagger \implies F_{1|0}(y_1|y') + F_1(y_0|y') \geq 1$$

and

$$y' > y_2^\dagger \implies F_{1|0}(y_1|y') + F_2(y_0|y') < 1, \qquad y' \leq y_2^\dagger \implies F_{1|0}(y_1|y') + F_2(y_0|y') \geq 1$$

Similarly to the proof for the upper bound, there may not be any values of $y'$ in $\mathcal{Y}$ that satisfy the strict inequalities above. One can plug these into Equation A.5 to obtain

$$\begin{aligned}
&F_2^L(y_0, y_1) - F_1^L(y_0, y_1) \\
&\qquad = \int_{\mathcal{Y}} (F_2(y_0|y') - F_1(y_0|y'))\mathbb{1}\{y' \leq y_1^\dagger, y' \leq y_2^\dagger\} \, dF_{t-1}(y') && \text{(A.6)} \\
&\qquad - \int_{\mathcal{Y}} (F_{1|0}(y_1|y') + F_1(y_0|y') - 1)\mathbb{1}\{y_2^\dagger < y' \leq y_1^\dagger\} \, dF_{t-1}(y') && \text{(A.7)} \\
&\qquad + \int_{\mathcal{Y}} (F_{1|0}(y_1|y') + F_2(y_0|y') - 1)\mathbb{1}\{y_1^\dagger < y' \leq y_2^\dagger\} \, dF_{t-1}(y') && \text{(A.8)}
\end{aligned}$$

*Case 1: $y_2^\dagger \leq y_1^\dagger$*

In this case, Equation A.8 is equal to 0. For Equation A.7, $F_{1|0}(y_1|y') \leq 1 - F_2(y_0|y')$ when

$y' \geq y_2^\dagger$ which implies that

$$\begin{aligned}
F_2^L(y_0, y_1) &- F_1^L(y_0, y_1) \\
&\geq \int_{\mathcal{Y}} (F_2(y_0|y') - F_1(y_0|y'))\mathbb{1}\{y' \leq y_1^\dagger\} \; dF_{t-1}(y') \\
&= F_2(y_0, y_1^\dagger) - F_1(y_0, y_1^\dagger) \geq 0
\end{aligned}$$

which holds because $F_1 \prec^C F_2$ as implied by the assumptions in the proposition and which implies the result.

*Case 2:* $y_2^\dagger > y_1^\dagger$

In this case, Equation A.7 is equal to 0. Equation A.8 is greater than or equal to zero because $F_{1|0}(y_1|y') + F_2(y_0|y') > 1$ for $y' < y_2^\dagger$ (and equality holding when there is no $y' \in \mathcal{Y}$ that is strictly less than $y_2^\dagger$). And Equation A.6 is equal to $F_2(y_0, y_1^\dagger) - F_1(y_0, y_1^\dagger)$ which holds because $y_1^\dagger < y_2^\dagger$ in this case), and this term is greater than or equal to zero because $F_1 \prec^C F_2$. Thus, in this case, each term in Equations (A.6) to (A.8) is greater than or equal to zero which implies the result.

## A.7 Proof of Proposition 2

First, note that

$$\begin{aligned}
F_{Y_{0t}|D=1}(y) &= P(\theta_t + \eta + V_t \leq y | D = 1) \\
&= \int \mathbb{1}\{u \leq y - \theta_t\} f_{\eta+V_t|D=1}(u) \, du \\
&= \int\int \mathbb{1}\{u \leq y - \theta_t\} f_{\eta+V_t,\eta+V_{t-1}|D=1}(u, w) \, du \, dw \\
&= \int\int \mathbb{1}\{u \leq y - \theta_t\} c_{\eta+V_t,\eta+V_{t-1}|D=1}(F_{\eta+V_t|D=1}(u), F_{\eta+V_{t-1}|D=1}(w)) \\
&\qquad \times f_{\eta+V_t|D=1}(u) f_{\eta+V_{t-1}|D=1}(w) \, du \, dw \\
&= \int\int \mathbb{1}\{u \leq y - \theta_t\} c_{\eta+V_{t-1},\eta+V_{t-2}|D=1}(F_{\eta+V_t|D=1}(u), F_{\eta+V_{t-1}|D=1}(w)) \\
&\qquad \times f_{\eta+V_t|D=1}(u) f_{\eta+V_{t-1}|D=1}(w) \, du \, dw \\
&= \int\int \mathbb{1}\{F^{-1}_{\eta+V_t|D=1}(F_{\eta+V_{t-1}|D=1}(\tilde{u})) \leq y - \theta_t\} f_{\eta+V_{t-1},\eta+V_{t-2}|D=1}(\tilde{u}, \tilde{w}) \, d\tilde{u} \, d\tilde{w} \\
&= P(\eta + V_{t-1} \leq F^{-1}_{\eta+V_{t-1}|D=1}(F_{\eta+V_t|D=1}(y - \theta_t)) | D = 1) \\
&= P(Y_{0t-1} \leq F^{-1}_{\eta+V_{t-1}|D=1}(F_{\eta+V_t|D=1}(y - \theta_t)) + \theta_{t-1} | D = 1) \\
&= F_{Y_{0t-1}|D=1}\Big(F^{-1}_{\eta+V_{t-1}|D=1}(F_{\eta+V_t|D=1}(y - \theta_t)) + \theta_{t-1}\Big)
\end{aligned}$$

where the third equality holds just by integrating out the second argument of the joint density, the fourth equality writes the joint density in terms of the copula and the marginal densities, the fifth equality holds by the condition in the proposition, the sixth equality holds using similar arguments as the proof of Lemma 1, and the remaining equalities hold immediately. Similar arguments imply that

$$F_{Y_{0t-1}|D=1}(y') = F_{Y_{0t-2}|D=1}\Big(F^{-1}_{\eta+V_{t-2}|D=1}(F_{\eta+V_{t-1}|D=1}(y' - \theta_{t-1})) + \theta_{t-2}\Big)$$

Then, for any $(u,v) \in [0,1]^2$,

$$\begin{aligned}
C_{Y_{0t},Y_{0t-1}|D=1}(u,v) &= P(F_{Y_{0t}|D=1}(Y_{0t}) \leq u, F_{Y_{0t-1}|D=1}(Y_{0t-1}) \leq v|D=1) \\
&= P\Big(F_{Y_{0t-1}|D=1}(F^{-1}_{\eta+V_{t-1}|D=1} \circ F_{\eta+V_t|D=1}(Y_{0t}-\theta_t)+\theta_{t-1}) \leq u, \\
&\qquad F_{Y_{0t-2}|D=1}(F^{-1}_{\eta+V_{t-2}|D=1} \circ F_{\eta+V_{t-1}|D=1}(Y_{0t-1}-\theta_{t-1})+\theta_{t-2}) \leq v\Big) \\
&= P\Big(F_{Y_{0t-1}|D=1}(F^{-1}_{\eta+V_{t-1}|D=1} \circ F_{\eta+V_t|D=1}(\eta+V_t)+\theta_{t-1}) \leq u, \\
&\qquad F_{Y_{0t-2}|D=1}(F^{-1}_{\eta+V_{t-2}|D=1} \circ F_{\eta+V_{t-1}|D=1}(\eta+V_{t-1})+\theta_{t-2}) \leq v\Big) \\
&= P\Big(F_{Y_{0t-1}|D=1}(\eta+V_{t-1}+\theta_{t-1}) \leq u, F_{Y_{0t-2}|D=1}(\eta+V_{t-2}+\theta_{t-2}) \leq v\Big) \\
&= P\Big(F_{Y_{0t-1}|D=1}(Y_{0t-1}) \leq u, F_{Y_{0t-2}|D=1}(Y_{0t-2}) \leq v\Big) \\
&= C_{Y_{0t-1},Y_{0t-2}|D=1}(u,v)
\end{aligned}$$

where the second equality follows from the two results earlier in this section, the third equality follows by substituting for $Y_{0t}$ and $Y_{0t-1}$, the fourth holds under the additional condition in the proposition, the fifth holds from the model in the proposition, and the last by the definition of the copula.

## A.8 Proof of Proposition 3

Using the same arguments as in Athey and Imbens (2006), one can show that

$$F_{Y_{0t}|D=1}(y) = F_{Y_{0t-1}|D=1} \circ h_{t-1} \circ h_t^{-1}(y)$$

and

$$F_{Y_{0t-1}|D=1}(y') = F_{Y_{0t-2}|D=1} \circ h_{t-2} \circ h_{t-1}^{-1}(y')$$

These two imply,

$$\begin{aligned}
C_{Y_{0t},Y_{0t-1}|D=1}(u,v) &= P(F_{Y_{0t}|D=1}(Y_{0t}) \leq u, F_{Y_{0t-1}|D=1}(Y_{0t-1}) \leq v|D=1) \\
&= P(F_{Y_{0t-1}|D=1} \circ h_{t-1} \circ h_t^{-1}(Y_{0t}) \leq u, F_{Y_{0t-2}|D=1} \circ h_{t-2} \circ h_{t-1}^{-1}(Y_{0t-1}) \leq v|D=1) \\
&= P(F_{Y_{0t-1}|D=1} \circ h_{t-1}(\eta+V_t) \leq u, F_{Y_{0t-2}|D=1} \circ h_{t-2}(\eta+V_{t-1}) \leq v|D=1) \\
&= P(F_{Y_{0t-1}|D=1} \circ h_{t-1}(\eta+V_{t-1}) \leq u, F_{Y_{0t-2}|D=1} \circ h_{t-2}(\eta+V_{t-2}) \leq v|D=1) \\
&= P(F_{Y_{0t-1}|D=1}(Y_{0t-1}) \leq u, F_{Y_{0t-2}|D=1}(Y_{0t-2}) \leq v|D=1) \\
&= C_{Y_{0t-1},Y_{0t-2}|D=1}(u,v)
\end{aligned}$$

where the fourth equality holds because of the additional condition in the proposition which implies that $(V_t, V_{t-1}, \eta)$ follows the same distribution as $(V_{t-1}, V_{t-2}, \eta)$.

# B Asymptotic Results

## B.1 Verifying Assumption 4

First, I provide some additional discussion regarding Assumption 4. In the application in the paper, I estimate each of the distributions $F_{Y_s|X,D=d}$ using quantile regression and then inverting the estimated quantiles to obtain the distribution. That is, I impose that for all $u \in (0,1)$,

$$Q_{Y_s|X,D=d}(u|x) = x^\top \beta_{s,d}(u)$$

The results in this section hold under Assumptions SB.1, SB.2 and SB.5 which are standard regularity conditions for quantile regression estimators and which are given in Appendix SB. Define the following terms,

$$J_{s,d}(u) = E[f_{Y_s|X,D=d}(X^\top \beta_{s,d}(u)|X) X X^\top | D = d] \tag{B.1}$$

In particular, under Assumption SB.5, and for $(s,d) \in \{t, t-1, t-2\} \times \{0,1\}$,

$$\hat{G}_{d,s}(y,x) = -\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi_{d,s}^{(y,x)}(Y_{is}, X_i, D_i) + o_p(1)$$

which holds uniformly in $y$ and $x$ and where

$$\begin{aligned}\psi_{d,s}^{(y,x)}(Y_s, X, D) = \frac{\mathbb{1}\{D = d\}}{p_d} & f_{Y_s|X,D=d}(y|x) x^\top J_{s,d}(F_{Y_s|X,D=d}(y|x))^{-1} \\ & \times \Big( \mathbb{1}\{Y_s \leq X^\top \beta_{s,d}(F_{Y_s|X,D=d}(y|x))\} - F_{Y_s|X,D=d}(y|x) \Big) X\end{aligned}$$

and that $\psi_{d,s}^{(y,x)}$ is a Donsker class. This implies that

$$(\hat{G}_{1,t}, \hat{G}_{1,t-1}, \hat{G}_{1,t-2}, \hat{G}_{0,t}, \hat{G}_{0,t-1}) \rightsquigarrow (\mathbb{W}_{1,t}, \mathbb{W}_{1,t-1}, \mathbb{W}_{1,t-2}, \mathbb{W}_{0,t}, \mathbb{W}_{0,t-1}) \tag{B.2}$$

where $(\mathbb{W}_{1,t}, \mathbb{W}_{1,t-1}, \mathbb{W}_{1,t-2}, \mathbb{W}_{0,t}, \mathbb{W}_{0,t-1})$ is a tight, mean zero Gaussian process with covariance function

$$V(y,x,\tilde{y},\tilde{x}) = E[\psi^{(y,x)}(Y,X,D)\psi^{(\tilde{y},\tilde{x})}(Y,X,D)^\top]$$

for $y = (y_1, y_2, y_3, y_4)^\top$, $x = (x_1, x_2, x_3, x_4)^\top$, $\tilde{y} = (\tilde{y}_1, \tilde{y}_2, \tilde{y}_3, \tilde{y}_4)^\top$, $\tilde{x} = (\tilde{x}_1, \tilde{x}_2, \tilde{x}_3, \tilde{x}_4)^\top$, and where

$$\psi^{(y,x)}(Y,X,D) = \begin{pmatrix} \psi_{1t}^{(y_1,x_1)}(Y_t, X, D) \\ \psi_{1t-1}^{(y_2,x_2)}(Y_{t-1}, X, D) \\ \psi_{0t}^{(y_3,x_3)}(Y_t, X, D) \\ \psi_{0t-1}^{(y_4,x_4)}(Y_{t-1}, X, D) \end{pmatrix}$$

Finally in this section, I establish that Assumption 4 holds when each of $F_{Y_s|X,D=d}$ is estimated using quantile regression and when the counterfactual distribution of untreated potential outcomes for the treated group, $F_{Y_{0t}|X,D=1}$, is identified using the Change in Changes approach of Athey and Imbens (2006) and Melly and Santangelo (2015) and estimated using quantile regression.

**Proposition B.1.** *Under Assumptions 1 to 3, SB.1, SB.2, SB.4 and SB.5, Assumption 4 holds* *with*

$$\mathbb{W}^0 = \mathbb{W}_{1,t-1} \circ F^{-1}_{Y_{0t-1}|X,D=0} \circ F_{Y_{0t}|X,D=0} + f_{Y_{0t-1}|X,D=1}(F^{-1}_{Y_{0t-1}|X,D=0} \circ F_{Y_{0t}|X,D=0}) \frac{\mathbb{W}_{0,t} - \mathbb{W}_{0,t-1} \circ F^{-1}_{Y_{0t-1}|X,D=0} \circ F_{Y_{0t}|X,D=0}}{f_{Y_{0t-1}|X,D=0}(F^{-1}_{Y_{0t-1}|X,D=0} \circ F_{Y_{0t}|X,D=0})}$$

*Proof.* The result follows immediately from Equation (B.2) and Proposition SB.1 in Appendix SB. □

## B.2 Distribution Regression with "Generated" Outcomes and Regressors

This section establishes useful intermediate results for distribution regression estimators (Chernozhukov, Fernandez-Val, and Melly (2013)) for conditional distributions when the outcomes and covariates are "transformed" and the transformation needs to be estimated in a preliminary step. Recall that

$$F_{Y_{0t}|Y_{t-1},X,D=1}(y|y',x) = P(Y_{t-1} \leq \Gamma_{10}(y,X)|\Gamma_{20}(Y_{t-2},X) = y', X = x, D = 1) := \Lambda(w^\top \beta_0(y))$$

where $w = (y', x^\top)^\top$, $\Lambda$ is some link function (see discussion in Supplementary Appendix SB.2), and where the first equality holds by the identification result in Lemma 1 and the second by imposing a distribution regression model. Here,

$$\Gamma_{10}(y,x) := F^{-1}_{Y_{0t-1}|X,D=1}(F_{Y_{0t}|X,D=1}(y|x)|x) \quad \text{and} \quad \Gamma_{20}(\tilde{y},x) := F^{-1}_{Y_{0t-1}|X,D=1}(F_{Y_{0t-2}|X,D=1}(\tilde{y}|x)|x) \tag{B.3}$$

and

$$\hat{\Gamma}_1(y,x) := \hat{F}^{-1}_{Y_{0t-1}|X,D=1}(\hat{F}_{Y_{0t}|X,D=1}(y|x)|x) \quad \text{and} \quad \hat{\Gamma}_2(\tilde{y},x) := \hat{F}^{-1}_{Y_{0t-1}|X,D=1}(\hat{F}_{Y_{0t-2}|X,D=1}(\tilde{y}|x)|x) \tag{B.4}$$

As a first step, notice that, under the assumptions utilized in Proposition 4,

$$\sqrt{n}(\hat{\Gamma}_1 - \Gamma_{10}) \rightsquigarrow \mathbb{Z}_{01} := \frac{\mathbb{W}^0 - \mathbb{W}_{1,t-1} \circ F^{-1}_{Y_{0t-1}|X,D=1}(F_{Y_{0t}|X,D=1}(y|x))}{f_{Y_{0t-1}|X,D=1}(F^{-1}_{Y_{0t-1}|X,D=1}(F_{Y_{0t}|X,D=1}(y|x)))} \tag{B.5}$$

and

$$\sqrt{n}(\hat{\Gamma}_2 - \Gamma_{20}) \rightsquigarrow \mathbb{Z}_{02} := \frac{\mathbb{W}_{1,t-2} - \mathbb{W}_{1,t-1} \circ F^{-1}_{Y_{0t-1}|X,D=1}(F_{Y_{0t-2}|X,D=1}(y|x))}{f_{Y_{0t-1}|X,D=1}(F^{-1}_{Y_{0t-1}|X,D=1}(F_{Y_{0t-2}|X,D=1}(y|x)))} \tag{B.6}$$

where the results in Equations (B.5) and (B.6) hold by Assumption 4 and from the results in Lemma B.4.

Following Chernozhukov, Fernandez-Val, and Melly (2013), I can build on results from the literature on Z-estimators to establish the limiting process for the estimator of $F_{Y_{0t}|Y_{t-1},X,D=1}$. First, define $W_{\Gamma_2} := (\Gamma_2(Y_{t-2},X), X^\top)^\top$ which is the $k+1$ vector of regressors that are used in the distribution regression and which is indexed by the map $\Gamma_2$. Next, define

$$\Psi_{\Gamma_1,\Gamma_2}(\beta) := E\Big[\Big(\Lambda(W_{\Gamma_2}^\top \beta) - \mathbb{1}\{Y_{t-1} \leq \Gamma_1(y,X)\}\Big)H(W_{\Gamma_2}^\top \beta)W_{\Gamma_2}|D=1\Big] \tag{B.7}$$

which are indexed by $\Gamma_1$ and $\Gamma_2$, and where $H$ is given in Equation (SB.3). Also, define

$$\hat{\Psi}_{\Gamma_1,\Gamma_2}(\beta) := \frac{1}{n}\sum_{i=1}^n \frac{D_i}{p}\Big(\Lambda(W_{i,\Gamma_2}^\top \beta) - \mathbb{1}\{Y_{it-1} \leq \Gamma_1(y,X_i)\}\Big)H(W_{i,\Gamma_2}^\top \beta)W_{i,\Gamma_2} \tag{B.8}$$

Further, notice that $\Psi_{\Gamma_{10},\Gamma_{20}}(\beta_0) = 0$ which is the population version of the first order condition for estimating $\beta_0$, and the distribution regression regression estimator, $\hat{\beta}_0$, satisfies $\hat{\Psi}_{\hat{\Gamma}_1,\hat{\Gamma}_2}(\hat{\beta}_0) = 0$. The arguments of Chernozhukov, Fernandez-Val, and Melly (2013) imply that

$$\sup_{y\in\bar{\mathcal{Y}}_{0t}} |\sqrt{n}(\Psi_{\Gamma_{10},\Gamma_{20}}(\hat{\beta}_0) - \Psi_{\Gamma_{10},\Gamma_{20}}(\beta_0)) - \sqrt{n}\dot{\Psi}_{\Gamma_{10},\Gamma_{20},\beta_0}(\hat{\beta}_0 - \beta_0)| = o_p(1)$$

which further implies that

$$\begin{aligned}
\sqrt{n}(\hat{\beta}_0 - \beta_0) &= \dot{\Psi}^{-1}_{\Gamma_{10},\Gamma_{20},\beta_0}\sqrt{n}(\Psi_{\Gamma_{10},\Gamma_{20}}(\hat{\beta}_0) - \Psi_{\Gamma_{10},\Gamma_{20}}(\beta_0)) + o_p(1) \\
&= -\dot{\Psi}^{-1}_{\Gamma_{10},\Gamma_{20},\beta_0}\sqrt{n}(\hat{\Psi}_{\hat{\Gamma}_1,\hat{\Gamma}_2}(\hat{\beta}_0) - \Psi_{\Gamma_{10},\Gamma_{20}}(\hat{\beta}_0)) + o_p(1)
\end{aligned}$$

and which holds uniformly in $y$. I show in Lemma B.1 below that $\sqrt{n}(\hat{\Psi}_{\hat{\Gamma}_1,\hat{\Gamma}_2}(\hat{\beta}_0) - \Psi_{\Gamma_{10},\Gamma_{20}}(\hat{\beta}_0)) \rightsquigarrow \mathbb{Z}_0$ which implies that

$$\sqrt{n}(\hat{\beta}_0 - \beta_0) \rightsquigarrow -\dot{\Psi}^{-1}_{\Gamma_{10},\Gamma_{20},\beta_0}\mathbb{Z}_0$$

and where $\dot{\Psi}_{\Gamma_{10},\Gamma_{20},\beta_0}(y) = M_0(y)$ which is defined in Equation (SB.2) in Appendix SB. This result is similar to the one from Chernozhukov, Fernandez-Val, and Melly (2013) with the notable exception that the limiting process $\mathbb{Z}_0$ accounts for the first step estimations in the current case. Finally, the conditional distribution $F_{Y_{0t}|Y_{0t-1},X,D=1}(y|y',x) = \Lambda(w^\top \beta_0(y))$ (here, again, I set $w = (y', x^\top)^\top$) which can be viewed as a map from $l^\infty(\bar{\mathcal{Y}}_{0t})$ to $l^\infty(\bar{\mathcal{Y}}_{0t}\mathcal{Y}_{1t-1}\mathcal{X}_1)$. This map is Hadamard differentiable and thus,

$$\sqrt{n}(\hat{F}_{Y_{0t}|Y_{0t-1},X,D=1} - \hat{F}_{Y_{0t}|Y_{0t-1},X,D=1}) \rightsquigarrow \mathbb{G}_2 := \lambda(w^\top \beta_0(y))w^\top M_0(y)^{-1}\mathbb{Z}_0 \tag{B.9}$$

which is the desired result.

The final part of this section provides additional lemmas which are used for deriving the main results in this section.

**Lemma B.1.** *Under Assumptions 1 to 4 and SB.1 to SB.3,*

$$\sqrt{n}(\hat{\Psi}_{\hat{\Gamma}_1,\hat{\Gamma}_2}(\hat{\beta}_0) - \Psi_{\Gamma_{10},\Gamma_{20}}(\hat{\beta}_0)) \rightsquigarrow \mathbb{Z}_0 := \mathbb{Z}_{00} + \Psi'_{1,\Gamma_{10}}\mathbb{Z}_{01} + \Psi'_{2,\Gamma_{20}}\mathbb{Z}_{02}$$

*where expressions for* $\Psi'_{1,\Gamma_{10}}$ *and* $\Psi'_{2,\Gamma_{20}}$ *are given in Lemmas B.2 and B.3 below.*

*Proof.* To show the result, start by adding and subtracting some terms:

$$\begin{aligned}
&\sqrt{n}(\hat{\Psi}_{\hat{\Gamma}_1,\hat{\Gamma}_2}(\hat{\beta}_0) - \Psi_{\Gamma_{10},\Gamma_{20}}(\hat{\beta}_0)) \\
&\qquad = \sqrt{n}\left(\hat{\Psi}_{\Gamma_{10},\Gamma_{20}}(\beta_0) - \Psi_{\Gamma_{10},\Gamma_{20}}(\beta_0)\right)
\end{aligned} \tag{B.10}$$

$$+\sqrt{n}\left(\Psi_{\hat{\Gamma}_1,\Gamma_{20}}(\beta_0)-\Psi_{\Gamma_{10},\Gamma_{20}}(\beta_0)\right) \tag{B.11}$$
$$+\sqrt{n}\left(\Psi_{\Gamma_{10},\hat{\Gamma}_2}(\beta_0)-\Psi_{\Gamma_{10},\Gamma_{20}}(\beta_0)\right) \tag{B.12}$$
$$+\sqrt{n}\left(\hat{\Psi}_{\hat{\Gamma}_1,\hat{\Gamma}_2}(\hat{\beta}_0)-\hat{\Psi}_{\hat{\Gamma}_1,\hat{\Gamma}_2}(\beta_0)-\left(\Psi_{\Gamma_{10},\Gamma_{20}}(\hat{\beta}_0)-\Psi_{\Gamma_{10},\Gamma_{20}}(\beta_0)\right)\right) \tag{B.13}$$
$$+\sqrt{n}\left(\hat{\Psi}_{\hat{\Gamma}_1,\hat{\Gamma}_2}(\beta_0)-\hat{\Psi}_{\Gamma_{10},\hat{\Gamma}_2}(\beta_0)-\left(\Psi_{\hat{\Gamma}_1,\Gamma_{20}}(\beta_0)-\Psi_{\Gamma_{10},\Gamma_{20}}(\beta_0)\right)\right) \tag{B.14}$$
$$+\sqrt{n}\left(\hat{\Psi}_{\Gamma_{10},\hat{\Gamma}_2}(\beta_0)-\hat{\Psi}_{\Gamma_{10},\Gamma_{20}}(\beta_0)-\left(\Psi_{\Gamma_{10},\hat{\Gamma}_2}(\beta_0)-\Psi_{\Gamma_{10},\Gamma_{20}}(\beta_0)\right)\right) \tag{B.15}$$

The term in Equation (B.10) can be handled exactly the same way as in Chernozhukov, Fernandez-Val, and Melly (2013) (see also the related discussion in Appendix SB.2). It comes from distribution regression of transformed values of $Y_{t-1}$ on transformed values of $Y_{t-2}$ if the transformations did not need to be estimated. In particular, the term in Equation (B.10) weakly converges to $\mathbb{Z}_{00}$ which is a tight, mean zero Gaussian process with covariance function

$$V_{Z_{00}}(\tilde{y}_1,\tilde{y}_2)=E[\psi^1_{\Gamma_{10}(\tilde{y}_1,X);\beta_0}(Y_{0t-1},W_{\Gamma_{20}},D)\psi^1_{\Gamma_{10}(\tilde{y}_2,X);\beta_0}(Y_{0t-1},W_{\Gamma_{20}},D)^\top] \tag{B.16}$$

where $\psi^1$ is defined in Equation (SB.4) in Appendix SB.

The terms in Equations (B.13) to (B.15) converge uniformly to 0 using stochastic equicontinuity arguments. The terms in Equations (B.11) and (B.12) capture the estimation effect of the first step estimators. Thus, by Lemmas B.2 and B.3, it follows that

$$\begin{aligned}
&\sqrt{n}\left(\hat{\Psi}_{\hat{\Gamma}_1,\hat{\Gamma}_2}(\hat{\beta}_0)-\Psi_{\Gamma_{10},\Gamma_{20}}(\hat{\beta}_0)\right)\\
&\qquad=\sqrt{n}\left(\hat{\Psi}_{\Gamma_{10},\Gamma_{20}}(\beta_0)-\Psi_{\Gamma_{10},\Gamma_{20}}(\beta_0)\right)+\Psi'_{1,\Gamma_{10}}\sqrt{n}(\hat{\Gamma}_1-\Gamma_{10})+\Psi'_{2,\Gamma_{20}}\sqrt{n}(\hat{\Gamma}_2-\Gamma_{20})+o_p(1)\\
&\qquad\rightsquigarrow \mathbb{Z}_{00}+\Psi'_{1,\Gamma_{10}}\mathbb{Z}_{01}+\Psi'_{2,\Gamma_{20}}\mathbb{Z}_{02}=\mathbb{Z}_0
\end{aligned}$$

where the first equality holds uniformly in $y$ and holds by Lemmas B.2 and B.3 and where $\mathbb{Z}_{01}$ and $\mathbb{Z}_{02}$ are given in Equations (B.5) and (B.6). $\square$

**Lemma B.2.** *Let $\mathbb{D}=l^\infty(\mathcal{Y}_{1t-1}\mathcal{X}_1)$ and consider the map $\Psi_1:\mathbb{D}_0\subset\mathbb{D}\mapsto l^\infty(\bar{\mathcal{Y}}_{0t})$ given by*

$$\Psi_1(\Gamma_1):=\Psi_{\Gamma_1,\Gamma_{20}}(\beta_0)$$

*where $\mathbb{D}_0$ denotes the space of conditional distribution functions with uniformly bounded and continuous densities. Then, the map $\Psi_1$ is Hadamard differentiable at $\Gamma_{10}$ tangentially to $\mathbb{D}_0$ with derivative at $\Gamma_{10}$ in $\gamma_1\in\mathbb{D}_0$ given by*

$$\Psi'_{1,\Gamma_{10}}(\gamma_1)=E\left[f_{Y_{t-1}|W_{\Gamma_{20}},D=1}(\Gamma_{10}(y,X)|W_{\Gamma_{20}})H(W_{\Gamma_{20}}^\top\beta_0)W_{\Gamma_{20}}\gamma_1|D=1\right]$$

*Proof.* To simplify the notation in the proof, I omit the dependence of $\Gamma_{10}$ on $X$ throughout. Also, I use the shorthand notation $F_1(\cdot|\cdot):=F_{Y_{t-1}|W_{\Gamma_{20}},D=1}(\cdot|\cdot)$ and let $f_1$ denote the corresponding density function. Consider any sequence $t_k>0$ and $\Gamma_{1k}\in\mathbb{D}_0$ for $k=1,2,3,\dots$ with $t_k\downarrow 0$ and

$$\gamma_{1k}=\frac{\Gamma_{1k}-\Gamma_{10}}{t_k}\to\gamma_1\in\mathbb{D}_0 \text{ as } k\to\infty$$

As a first step, notice that

$$E[F_1(\Gamma_{1k}(y)|W_{\Gamma_{20}})-F_1(\Gamma_{10}(y)|W_{\Gamma_{20}})|D=1]=t_k\gamma_{1k}(y)\ E\left[\int_0^1 f_1(\Gamma_{10}(y)+rt_k\gamma_{1k}(y)|W_{\Gamma_{20}})\,dr|D=1\right] \tag{B.17}$$

where the expectation is with respect to $W_{\Gamma_{20}}$ and which holds by writing the conditional distribution as an integral and then a change of variables argument. Then,

$$
\begin{aligned}
&\frac{\Psi_1(\Gamma_{1k}) - \Psi_1(\Gamma_{10})}{t_k} - \Psi'_{1,\Gamma_{10}} \\
&= \frac{E\left[\left(\mathbb{1}\{Y_{t-1} \leq \Gamma_{1k}(y)\} - \mathbb{1}\{Y_{t-1} \leq \Gamma_{10}(y)\}\right) H(W_{\Gamma_{20}}^\top \beta_0) W_{\Gamma_{20}} | D=1\right]}{t_k} - \Psi'_{1,\Gamma_{10}} \\
&= \frac{E\left[\left(F_1(\Gamma_{1k}(y)|W_{\Gamma_{20}}) - F_1(\Gamma_{10}(y)|W_{\Gamma_{20}})\right) H(W_{\Gamma_{20}}^\top \beta_0) W_{\Gamma_{20}} | D=1\right]}{t_k} - \Psi'_{1,\Gamma_{10}} \\
&= E\left[\left(\int_0^1 f_1(\Gamma_{10}(y) + r t_k \gamma_{1k}(y)|W_{\Gamma_{20}})\, dr - f_1(\Gamma_{10}(y)|W_{\Gamma_{20}})\right) H(W_{\Gamma_{20}}^\top \beta_0) W_{\Gamma_{20}} \gamma_{1k}(y) | D=1\right] \\
&\qquad + E\left[f_1(\Gamma_{10}(y)|W_{\Gamma_{20}}) H(W_{\Gamma_{20}}^\top \beta_0) W_{\Gamma_{20}} (\gamma_{1k}(y) - \gamma_{10}(y)) | D=1\right] \qquad \text{(B.18)}
\end{aligned}
$$

where the third equality holds by the same argument as in Equation (B.17) and by adding and subtracting terms. The first term converges uniformly to 0 because $f_1$ is uniformly continuous, $\|\gamma_{1k} - \gamma_{10}\|_\infty \to 0$, and because $\gamma_{1k}$ is uniformly bounded. The second term converges uniformly to 0 because $f_1$ is uniformly bounded and $\|\gamma_{1k} - \gamma_{10}\|_\infty \to 0$. ☐

**Lemma B.3.** *Let $\mathbb{D} = l^\infty(\mathcal{Y}_{1t-1}\mathcal{X}_1)$ and consider the map $\Psi_2 : \mathbb{D}_0 \subset \mathbb{D} \mapsto l^\infty(\bar{\mathcal{Y}}_{0t})$ given by*

$$\Psi_2(\Gamma_2) := \Psi_{\Gamma_{10},\Gamma_2}(\beta_0)$$

*where $\mathbb{D}_0$ denotes the space of conditional distribution functions with uniformly bounded and continuous densities. Then, the map $\Psi_2$ is Hadamard differentiable at $\Gamma_{20}$ tangentially to $\mathbb{D}_0$ with derivative at $\Gamma_{20}$ in $\gamma_2 \in \mathbb{D}_0$ given by*

$$
\begin{aligned}
\Psi'_{2,\Gamma_{20}}(\gamma_2) &= E[\lambda(W_{\Gamma_{20}}^\top \beta_0) H(W_{\Gamma_{20}}^\top \beta_0) W_{\Gamma_{20}} \beta_0^{(1)} \gamma_2 | D=1] \qquad \text{(B.19)} \\
&+ E[(\Lambda(W_{\Gamma_{20}}^\top \beta_0) - \mathbb{1}\{Y_{t-1} \leq \Gamma_{10}(y,X)\}) h(W_{\Gamma_{20}}^\top \beta_0) W_{\Gamma_{20}} \beta_0^{(1)} \gamma_2 | D=1] \\
&+ E[(\Lambda(W_{\Gamma_{20}}^\top \beta_0)) - \mathbb{1}\{Y_{t-1} \leq \Gamma_{10}(y,X)\}) H(W_{\Gamma_{20}}^\top \beta_0) e_1 \gamma_2 | D=1]
\end{aligned}
$$

*where $\beta_0^{(1)}$ is the first element in the vector $\beta_0$ and $e_1$ is a $(k+1) \times 1$ vector with 1 as its first element and 0 for all the other elements.*

*Proof.* From Equation (B.20), write $\Psi'_{2,\Gamma_{20}}(\gamma_2) := A_1 + A_2 + A_3$. Consider any sequence $t_k > 0$ and $\Gamma_{2k} \in \mathbb{D}_0$ for $k = 1, 2, 3, \ldots$ with $t_k \downarrow 0$ and

$$\gamma_{2k} = \frac{\Gamma_{2k} - \Gamma_{20}}{t_k} \to \gamma_2 \in \mathbb{D}_0 \text{ as } k \to \infty$$

Now, notice that by adding and subtracting some terms, one can write

$$
\begin{aligned}
&\frac{\Psi_2(\Gamma_{2k}) - \Psi_2(\Gamma_{20})}{t_k} - \Psi'_{2,\Gamma_{20}}(\gamma_2) \\
&\qquad = E[(\Lambda(W_{\Gamma_{2k}}^\top \beta_0) - \Lambda(W_{\Gamma_{20}}^\top \beta_0)) H(W_{\Gamma_{20}}^\top \beta_0) W_{\Gamma_{20}} | D=1]/t_k - A_1 \qquad \text{(B.20)} \\
&\qquad + E[(\Lambda(W_{\Gamma_{20}}^\top \beta_0) - \mathbb{1}\{\Gamma_{10}(Y_{t-1}) \leq y\}) \qquad \text{(B.21)} \\
&\qquad\qquad \times \left(H(W_{\Gamma_{2k}}^\top \beta_0) - H(W_{\Gamma_{20}}^\top \beta_0)\right) W_{\Gamma_{20}} | D=1]/t_k - A_2
\end{aligned}
$$

$$
\begin{aligned}
&+ E[(\Lambda(\Gamma_{20}(Y_{t-2}^\top \beta_0)) - \mathbb{1}\{\Gamma_{10}(Y_{t-1}) \leq y\}) \\
&\qquad \times H(W_{\Gamma_{20}}^\top \beta_0)\,(W_{\Gamma_{2k}} - W_{\Gamma_{20}})\,|D = 1]/t_k - A_3
\end{aligned} \tag{B.22}
$$

which holds uniformly and up to some smaller order terms. For Equation (B.20), and by a Taylor expansion argument and for some $\bar{\Gamma}_2(y, x)$ between $\Gamma_{2k}(y, x)$ and $\Gamma_{20}(y, x)$, it is equal to

$$
\begin{aligned}
&= E[\lambda(W_{\bar{\Gamma}_2}^\top \beta_0) H(W_{\Gamma_{20}}^\top \beta_0) W_{\Gamma_{20}} \beta_0^{(1)} \gamma_{2k} | D = 1] - A_1 \\
&= E\Big[\Big(\lambda(W_{\bar{\Gamma}_2}^\top \beta_0) - \lambda(W_{\Gamma_{20}}^\top \beta_0)\Big) H(W_{\Gamma_{20}}^\top \beta_0) W_{\Gamma_{20}} \beta_0^{(1)} \gamma_{2k} | D = 1\Big] \\
&\qquad + E[\lambda(W_{\Gamma_{20}}^\top \beta_0) H(W_{\Gamma_{20}}^\top \beta_0) W_{\Gamma_{20}} (\gamma_{2k} - \gamma_2) | D = 1]
\end{aligned}
$$

Thus, the term in Equation (B.20) converges uniformly to 0 because $\bar{\Gamma}_2$ converges uniformly to $\Gamma_{20}$, $\gamma_{2k}$ is uniformly bounded, and $\gamma_{2k}$ converges uniformly to $\gamma_2$.

For Equation (B.21), and using similar arguments as for Equation (B.20), it is equal to

$$
\begin{aligned}
&= E\left[(\Lambda(W_{\Gamma_{20}}^\top \beta_0) - \mathbb{1}\{\Gamma_{10}(Y_{t-1}) \leq y\}) h(W_{\bar{\Gamma}_2}^\top \beta_0) W_{\Gamma_{20}} \beta_0^{(1)} \gamma_{2k} | D = 1\right] - A_2 \\
&= E\left[(\Lambda(W_{\Gamma_{20}}^\top \beta_0) - \mathbb{1}\{\Gamma_{10}(Y_{t-1}) \leq y\}) \left(h(W_{\bar{\Gamma}_2}^\top \beta_0) - h(W_{\Gamma_{20}}^\top \beta_0)\right) W_{\Gamma_{20}} \beta_0^{(1)} \gamma_{2k} | D = 1\right] \\
&\qquad + E\left[(\Lambda(W_{\Gamma_{20}}^\top \beta_0) - \mathbb{1}\{\Gamma_{10}(Y_{t-1}) \leq y\}) h(W_{\Gamma_{20}}^\top \beta_0) W_{\Gamma_{20}} \beta_0^{(1)} (\gamma_{2k} - \gamma_2) | D = 1\right]
\end{aligned}
$$

where $h$ is the derivative of $H$. The first term above converges to 0 because $\Gamma_{2k}$ converges uniformly to 0, $h$ is uniformly continuous, and the other terms are uniformly bounded. The second term converges to 0 for the same reasons as well as that $\gamma_{2k}$ converges uniformly to $\gamma_2$. Finally, for Equation (B.22), $W_{\Gamma_{2k}} = W_{\Gamma_{20}}$ each element in $W$ except for the first one. For the first element, notice that it is equal to

$$
\begin{aligned}
&= E\left[(\Lambda(\Gamma_{20}(Y_{t-2}^\top \beta_0)) - \mathbb{1}\{\Gamma_{10}(Y_{t-1}) \leq y\}) H(W_{\Gamma_{20}}^\top \beta_0) \left(\frac{\Gamma_{2k}(Y_{t-2}, X) - \Gamma_{20}(Y_{t-2}, X)}{t_k}\right)\right] - A_3 \\
&= E\left[(\Lambda(\Gamma_{20}(Y_{t-2}^\top \beta_0)) - \mathbb{1}\{\Gamma_{10}(Y_{t-1}) \leq y\}) H(W_{\Gamma_{20}}^\top \beta_0)\,(\gamma_{2k} - \gamma_2)\right] - A_3
\end{aligned}
$$

which converges uniformly to 0 since $\gamma_{2k}$ converges uniformly to $\gamma_2$ and each of the other terms are uniformly bounded. □

**Lemma B.4.** *Let $\mathbb{D} := l^\infty(\mathcal{Y}_{1t-1}\mathcal{X}_1) \times l^\infty(\mathcal{Y}_{1t-2}\mathcal{X}_1)$, define the map $\psi : \mathbb{D}_\psi \subset \mathbb{D} \mapsto l^\infty(\mathcal{Y}_{1t-2}\mathcal{X}_1)$, given by*

$$
\psi(F) = G^{-1} \circ H
$$

*for $F := (G, H) \in \mathbb{D}_\psi$ and with $\mathbb{D}_\psi := \mathbb{E}^2$ with $\mathbb{E}$ the set of all conditional distribution functions with a strictly positive and bounded conditional density. Then, the map $\psi$ is Hadamard differentiable at $F_0$ tangentially to $\mathbb{D}_\psi$ with derivative given by*

$$
\psi'_{F_0}(\gamma) = \frac{\gamma_2 - \gamma_1 \circ G_0^{-1} \circ H_0}{g_0 \circ G_0^{-1} \circ H_0}
$$

*where $\gamma := (\gamma_1, \gamma_2) \in \mathbb{D}_\psi$.*

*Proof.* This result follows from Lemma A.1 in Callaway, Li, and Oka (2018). □

## B.3 Additional Preliminary Results

This section presents some additional helpful preliminary results for establishing the limiting distribution of the estimators of the *DoTT* and *QoTT*. The key ingredients are establishing the Hadamard directional differentiability of the maps from conditional distribution functions to the *DoTT* and the *QoTT*. I establish piece-by-piece the main intermediate steps to proving this result in this section. In the next section, I provide proofs of the main asymptotic results.

**Lemma B.5.** *Consider the map* $\phi_3^U : \mathbb{D}_{\phi_3^U} \subset l^\infty(\mathcal{Y}_\delta \boldsymbol{\Delta} \mathcal{Y}_{1t-1} \mathcal{X}_1) \mapsto l^\infty(\mathcal{Y}_\delta \boldsymbol{\Delta} \mathcal{Y}_{1t-1} \mathcal{X}_1)$ *given by*

$$\phi_3^U(\theta)(y, \delta, y', x) = \min\{\theta(y, \delta, y', x), 0\}$$

*for* $\theta \in \mathbb{D}_{\phi_3^U}$. *Then, the map* $\phi_3^U$ *is Hadamard directionally differentiable at* $\theta_0 \in \mathbb{D}_{\phi_3^U}$ *tangentially to* $\mathbb{D}_{\phi_3^U}$ *in* $\varphi \in \mathbb{D}_{\phi_U^3}$ *with derivative given by*

$$\phi_{3,\theta_0}^{U'}(\varphi) = \begin{cases} \varphi(y, \delta, y', x) & \text{if } \theta_0(y, \delta, y', x) < 0 \\ \min\{\varphi(y, \delta, y', x), 0\} & \text{if } \theta_0(y, \delta, y', x) = 0 \\ 0 & \text{if } \theta_0(y, \delta, y', x) > 0 \end{cases}$$

*Proof.* The proof follows using the same argument as in Fang and Santos (2019, Example 2.1). □

**Lemma B.6.** *Consider the map* $\phi_3^L : \mathbb{D}_{\phi_3^L} \subset l^\infty(\mathcal{Y}_\delta \boldsymbol{\Delta} \mathcal{Y}_{1t-1} \mathcal{X}_1) \mapsto l^\infty(\mathcal{Y}_\delta \boldsymbol{\Delta} \mathcal{Y}_{1t-1} \mathcal{X}_1)$ *given by*

$$\phi_3^L(\theta)(y, \delta, y', x) = \max\{\theta(y, \delta, y', x), 0\}$$

*for* $\theta \in \mathbb{D}_{\phi_3^L}$. *Then, the map* $\phi_3^L$ *is Hadamard directionally differentiable at* $\theta_0 \in \mathbb{D}_{\phi_3^L}$ *tangentially to* $\mathbb{D}_{\phi_3^L}$ *in* $\varphi \in \mathbb{D}_{\phi_3^L}$ *with derivative given by*

$$\phi_{3,\theta_0}^{L'}(\varphi) = \begin{cases} 0 & \text{if } \theta_0(y, \delta, y', x) < 0 \\ \max\{\varphi(y, \delta, y', x), 0\} & \text{if } \theta_0(y, \delta, y', x) = 0 \\ \varphi(y, \delta, y', x) & \text{if } \theta_0(y, \delta, y', x) > 0 \end{cases}$$

*Proof.* The result follows immediately from Fang and Santos (2019, Example 2.1). □

For any $\theta \in l^\infty(\mathcal{Y}_\delta \boldsymbol{\Delta} \mathcal{Y}_{1t-1} \mathcal{X}_1)$, recall that

$$\Phi_{\mathcal{Y}_\delta}^L(\theta, \delta, y', x) = \underset{y \in \mathcal{Y}_\delta}{\operatorname{argmax}} \;\; \theta(y, \delta, y', x) \quad \text{and} \quad \Phi_{\mathcal{Y}_\delta}^U(\theta, \delta, y', x) = \underset{y \in \mathcal{Y}_\delta}{\operatorname{argmin}} \;\; \theta(y, \delta, y', x)$$

Then, the following results hold

**Lemma B.7.** *Consider the map* $\phi_2^L : \mathbb{D}_{\phi_2^L} \subset l^\infty(\mathcal{Y}_\delta \boldsymbol{\Delta} \mathcal{Y}_{1t-1}, \mathcal{X}_1) \mapsto l^\infty(\boldsymbol{\Delta} \mathcal{Y}_{1t-1}, \mathcal{X}_1)$ *given by*

$$\phi_2^L(\theta)(\delta, y', x) = \sup_{y \in \mathcal{Y}_\delta} \theta(y, \delta, y', x)$$

*for* $\theta \in \mathbb{D}_{\phi_2^L} := \mathcal{C}(\boldsymbol{\Delta} \mathcal{Y}_{1t-1} \mathcal{X}_1)$. *Then, the map* $\phi_2^L$ *is Hadamard directionally differentiable at* $\theta_0 \in \mathbb{D}_{\phi_2^L}$ *tangentially to* $\mathbb{D}_{\phi_2^L}$ *with derivative in* $\varphi \in \mathbb{D}_{\phi_2^L}$ *given by*

$$\phi_{2,\theta_0}^{L'}(\delta, y', x) = \sup_{y \in \Phi_{\mathcal{Y}_\delta}^L(\theta_0, \delta, y', x)} \varphi(y, \delta, y', x)$$

*Proof.* The proof follows immediately from Masten and Poirier (2019, Lemma 8) □

**Lemma B.8.** *Consider the map* $\phi_2^U : \mathbb{D}_{\phi_2^U} \subset l^\infty(\mathcal{Y}_\delta \boldsymbol{\Delta}\mathcal{Y}_{1t-1}\mathcal{X}_1) \mapsto l^\infty(\boldsymbol{\Delta}\mathcal{Y}_{1t-1}, \mathcal{X}_1)$ *given by*

$$\phi_2^U(\theta)(\delta, y', x) = \inf_{y \in \mathcal{Y}_\delta} \theta(y, \delta, y', x)$$

*for* $\theta \in \mathbb{D}_{\phi_2^U} := \mathcal{C}(\boldsymbol{\Delta}\mathcal{Y}_{1t-1}\mathcal{X}_1)$*. Then, the map* $\phi_2^U$ *is Hadamard directionally differentiable at* $\theta_0 \in \mathbb{D}_{\phi_2^U}$ *tangentially to* $\mathbb{D}_{\phi_2^U}$ *with derivative in* $\varphi \in \mathbb{D}_{\phi_2^U}$ *given by*

$$\phi_{2,\theta_0}^{U'}(\delta, y', x) = \inf_{y \in \Phi_{\mathcal{Y}_\delta}^U(\theta_0, \delta, y', x)} \varphi(y, \delta, y', x)$$

*Proof.* The result follows using essentially the same arguments as in Lemma B.7 which builds on Masten and Poirier (2019, Lemma 8). □

Finally, the next result restates Lemma D.1 of Chernozhukov, Fernandez-Val, and Melly (2013) with the notation adjusted to be the same as in the current paper.

**Lemma B.9.** *Consider the map* $\phi_1 : \mathbb{D}_{\phi_1} \subset l^\infty(\boldsymbol{\Delta}\mathcal{Y}_{1t-1}\mathcal{X}_1) \times l^\infty(\mathcal{Y}_{1t-1}\mathcal{X}_1) \mapsto l^\infty(\boldsymbol{\Delta})$ *given by*

$$\int_{\mathcal{Y}_{t-1}\mathcal{X}} \Lambda_1(\cdot|y_{t-1}, x)\, d\Lambda_2(y_{t-1}, x)$$

*for* $\Lambda = (\Lambda_1, \Lambda_2) \in \mathbb{D}_{\phi_1}$ *where* $\mathbb{D}_{\phi_1}$ *is the product of the space of measurable functions* $\Lambda_1 : \boldsymbol{\Delta}\mathcal{Y}_{1t-1}\mathcal{X}_1 \mapsto [0,1]$ *and of the bounded maps* $\Lambda_2 : \mathcal{F} \mapsto \mathbb{R}$ *given by* $f \mapsto \int f\, d\Lambda_2$ *where* $\Lambda_2$ *is a probability measure on* $\mathcal{Y}_{1t-1}\mathcal{X}_1$*. Then, the map* $\phi_1$ *is Hadamard differentiable at* $\Lambda_0 = (\Lambda_{10}, \Lambda_{20})$ *tangentially to* $\mathbb{D}_0$ *where* $\mathbb{D}_0$ *denotes the product of the space of uniformly continuous functions mapping* $\mathcal{Y}_{1t-1}\mathcal{X}_1$ *to* $[0,1]$ *times the space of uniformly continuous functions in* $\mathcal{F}$ *in* $\lambda = (\lambda_1, \lambda_2) \in \mathbb{D}_{\phi_1}$ *with derivative given by*

$$\phi_{1,\Lambda_0}'(\lambda) = \int_{\mathcal{Y}_{t-1}} \int_{\mathcal{X}} \lambda_1(\cdot|y_{t-1}, x)\, d\Lambda_{20}(y_{t-1}, x) + \int_{\mathcal{Y}_{t-1}} \int_{\mathcal{X}} \Lambda_{10}(\cdot|y_{t-1}, x)\, d\lambda_2(y_{t-1}, x)$$

*Proof.* The result follows immediately using the arguments of Chernozhukov, Fernandez-Val, and Melly (2013, Lemma D.1) □

## B.4 Proofs of Main Asymptotic Results

**Proof of Proposition 4**

Recall that Theorem 2 establishes identification of $DoTT^L$ and $DoTT^U$. Let $F_{10} := F_{Y_{1t}|Y_{0t-1},X,D=1}$, $F_{20} := F_{Y_{0t}|Y_{0t-1},X,D=1}$, and $F_{30} := F_{Y_{0t-1},X|D=1}$; also, let $\hat{F}_1 := \hat{F}_{Y_{1t}|Y_{0t-1},X,D=1}$, $\hat{F}_2 := \hat{F}_{Y_{0t}|Y_{0t-1},X,D=1}$, and $\hat{F}_3 := \hat{F}_{Y_{0t-1},X|D=1}$. Using the notation of this section and the definitions of $\phi_1$, $\phi_2^L$, $\phi_2^U$, $\phi_3^L$, and $\phi_3^U$ in the previous section, notice that

$$\begin{aligned} DoTT^L(\delta) &= \phi^L(F_{10}, F_{20}, F_{30}) \\ &:= \phi_1\Big(\phi_2^L \circ \phi_3^L(F_{10}, F_{20}), F_{30}\Big) \end{aligned} \quad \text{and} \quad \begin{aligned} DoTT^U(\delta) &= \phi^U(F_{10}, F_{20}, F_{30}) \\ &:= \phi_1\Big(\phi_2^U \circ \phi_3^U(F_{10}, F_{20}), F_{30}\Big) \end{aligned}$$

and that estimators of the lower and upper bounds of the distribution of the treatment effect are given by

$$\widehat{DoTT}^{L}(\delta) = \phi^{L}(\hat{F}_1, \hat{F}_2, \hat{F}_3) \quad \text{and} \quad \widehat{DoTT}^{U}(\delta) = \phi^{U}(\hat{F}_1, \hat{F}_2, \hat{F}_3)$$

Then, from the results in Lemmas B.5 to B.9 and by the chain rule for Hadamard directionally differentiable functions (Shapiro (1990) and Masten and Poirier (2019)), it holds that the maps $\phi^L$ and $\phi^U$ are Hadamard directionally differentiable with derivative at $F_0$ given in $\tilde{F} := (\tilde{F}_1, \tilde{F}_2, \tilde{F}_3)$ given by

$$\phi^{L'}_{F_0}(\tilde{F}) = \phi'_{1,(\phi_2^L \circ \phi_3^L(F_{10}, F_{20}), F_{30})}\left(\phi^{L'}_{2,\phi_3^L(F_{10},F_{20})} \circ \phi^{L'}_{3,(F_{10},F_{20})}(\tilde{F}_1, \tilde{F}_2), \tilde{F}_3\right)$$

and

$$\phi^{U'}_{F_0}(\tilde{F}) = \phi'_{1,(\phi_2^U \circ \phi_3^U(F_{10}, F_{20}), F_{30})}\left(\phi^{U'}_{2,\phi_3^U(F_{10},F_{20})} \circ \phi^{U'}_{3,(F_{10},F_{20})}(\tilde{F}_1, \tilde{F}_2), \tilde{F}_3\right)$$

Then, the delta method for Hadamard directionally differentiable functions (Fang and Santos (2019)) in combination with Theorem 4 and Lemmas B.5 to B.9 implies the result.

**Proof of Theorem 5**

The proof of Theorem 5 follows immediately from the result in Proposition 4, by the Hadamard differentiability of the quantile map (see van der Vaart and Wellner (1996, Lemma 3.9.23(ii))), and by recalling that the $QoTT^L$ comes from inverting $DoTT^U$ and $QoTT^U$ comes from inverting $DoTT^L$.

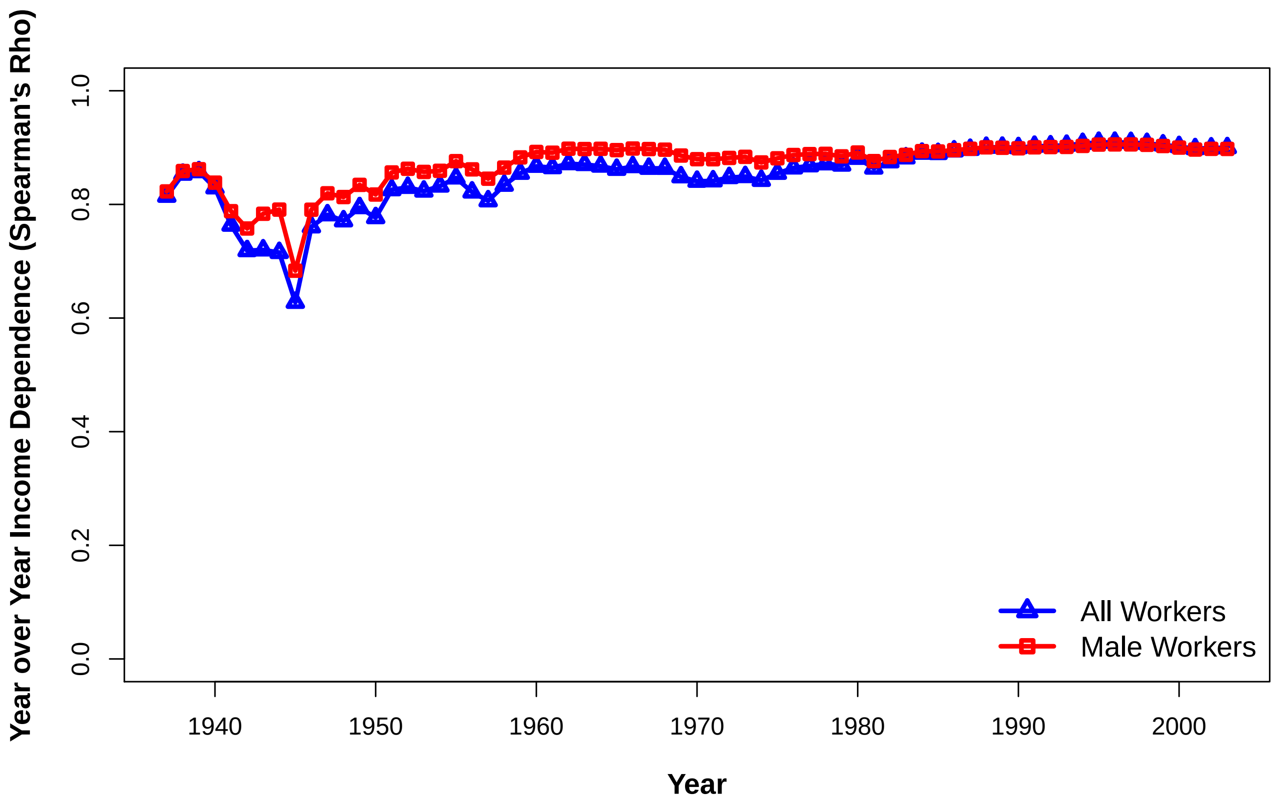
# C Tables and Figures

Table 1: Summary Statistics

| | Displaced | Non-Displaced | Diff | P-val on Diff |
|---|---|---|---|---|
| **Characteristics** | | | | |
| Less than HS | 0.10 | 0.08 | 0.02 | 0.43 |
| High School | 0.65 | 0.61 | 0.04 | 0.41 |
| College | 0.25 | 0.31 | -0.06 | 0.19 |
| Hispanic | 0.20 | 0.17 | 0.03 | 0.47 |
| Black | 0.35 | 0.25 | 0.10 | 0.02 |
| White | 0.45 | 0.57 | -0.12 | 0.01 |
| Male | 0.56 | 0.52 | 0.04 | 0.33 |
| Female | 0.44 | 0.48 | -0.04 | 0.33 |
| **Path of Earnings** | | | | |
| 2013 Earnings | 49.82 | 69.50 | -19.68 | 0.01 |
| 2011 Earnings | 45.51 | 67.06 | -21.55 | 0.00 |
| 2009 Earnings | 53.70 | 62.28 | -8.57 | 0.17 |
| 2007 Earnings | 59.17 | 60.93 | -1.76 | 0.78 |
| 2005 Earnings | 55.32 | 54.97 | 0.35 | 0.95 |
| 2003 Earnings | 49.71 | 49.88 | -0.17 | 0.97 |
| 2001 Earnings | 42.71 | 46.58 | -3.87 | 0.40 |

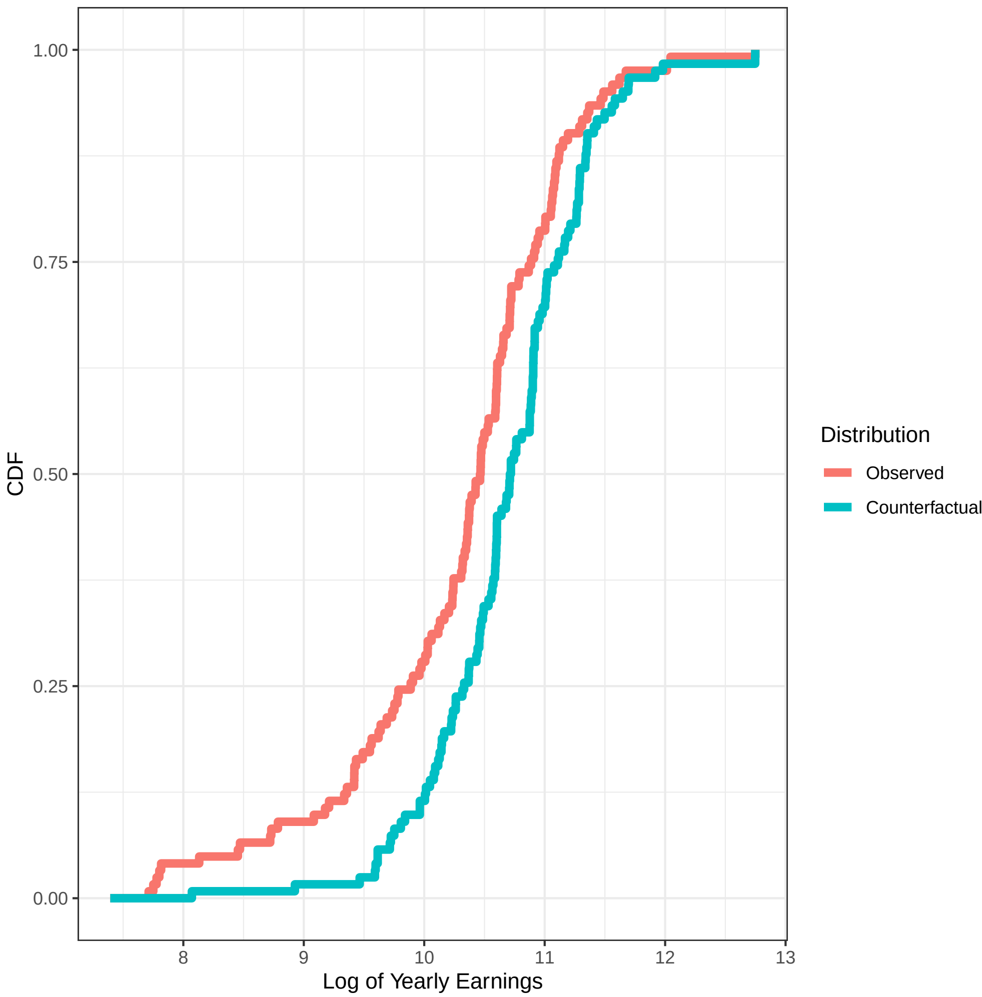
*Notes:* Summary statistics for individuals based on whether or not an individual was displaced from his job in 2008 or 2009. The top panel uses the sample used for the main results in the paper with a sample size of 2,775 of which 122 are displaced which amounts to 5.6% of the observations. The bottom panel uses a balanced panel subset of the data for which earnings are available in all years in the table which includes 2,077 observations. Earnings are in thousands of dollars.
*Sources:* 1979 National Longitudinal Study of Youths

Figure 1: Rank Correlation (Spearman's Rho) of Year over Year Annual Income Dependence for All Workers and Male Workers from 1937-2003



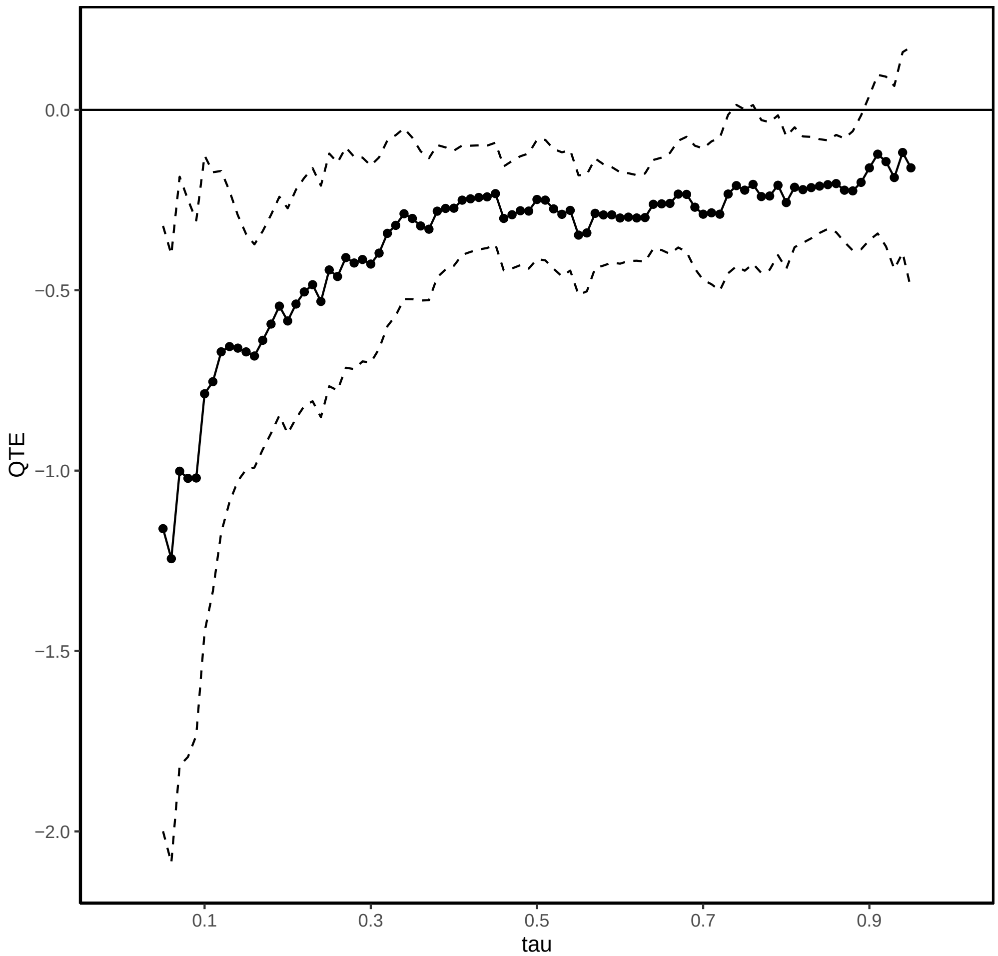

*Notes:* The data comes from Kopczuk, Saez, and Song (2010) and replicates part of Figure 4 in that paper.

Figure 2: Marginal Distributions of Displaced and Non-displaced Potential Outcomes for the Displaced Group



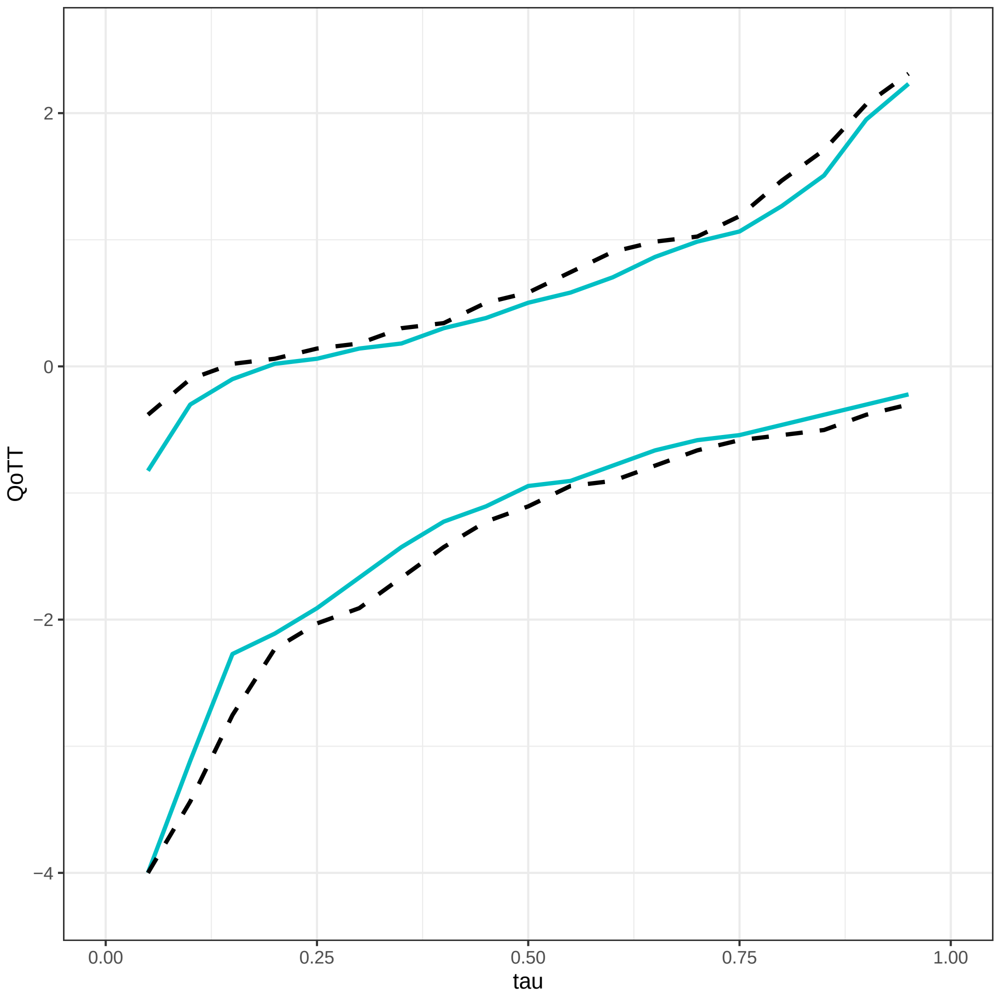

*Notes:* This figure provides estimates of the distribution of displaced potential earnings for the treated group and the counterfactual distribution of non-displaced potential earnings for the treated group. The latter is estimated using the Change in Changes model as described in the text.
*Sources:* 1979 National Longitudinal Survey of Youth

Figure 3: The Quantile Treatment Effect on the Treated



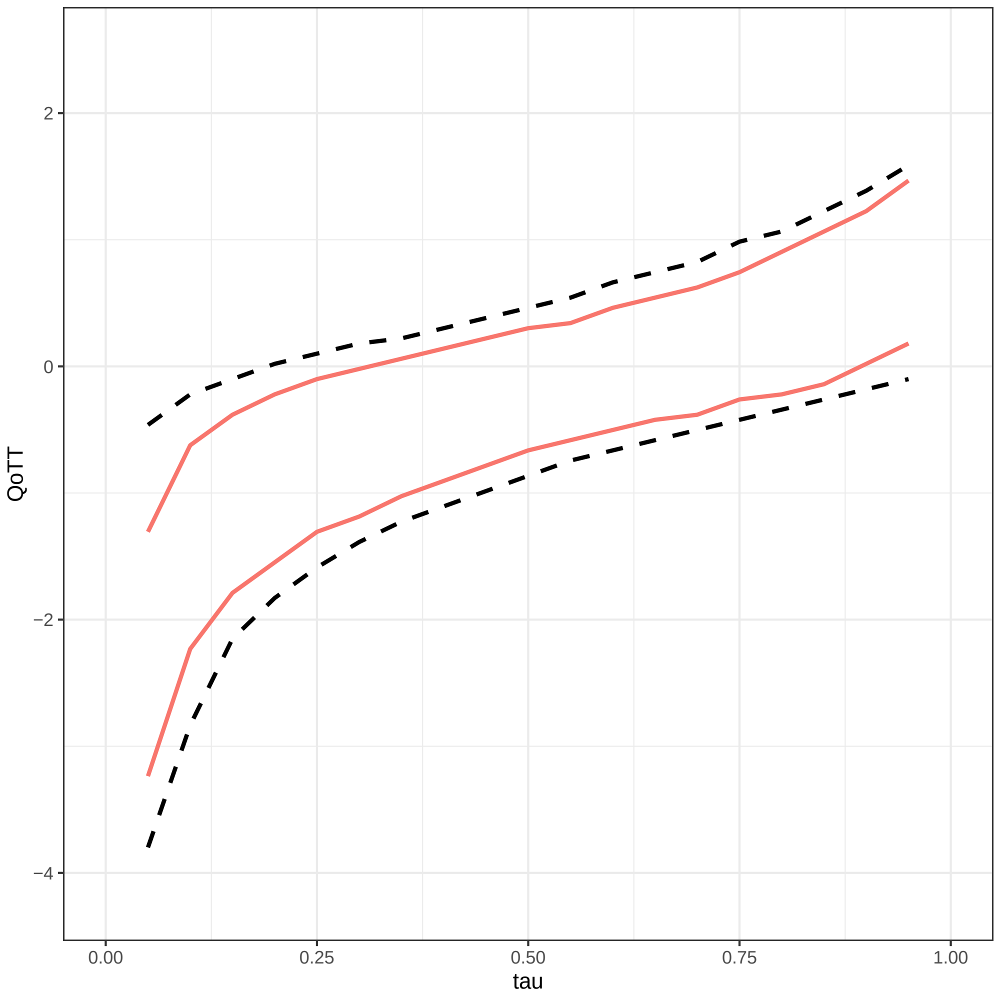

*Notes:* This figure provides estimates of the QTT of job displacement. The QTT is estimated using the Change in Changes model as described in the text. The scale of the y-axis is in log points. Most of the reported results in the text convert log points into percentage changes (see Footnote 19). The dotted lines provide pointwise 95% confidence intervals using the empirical bootstrap.
*Sources:* 1979 National Longitudinal Survey of Youth

Figure 4: Worst Case Bounds on the Quantile of the Treatment Effect



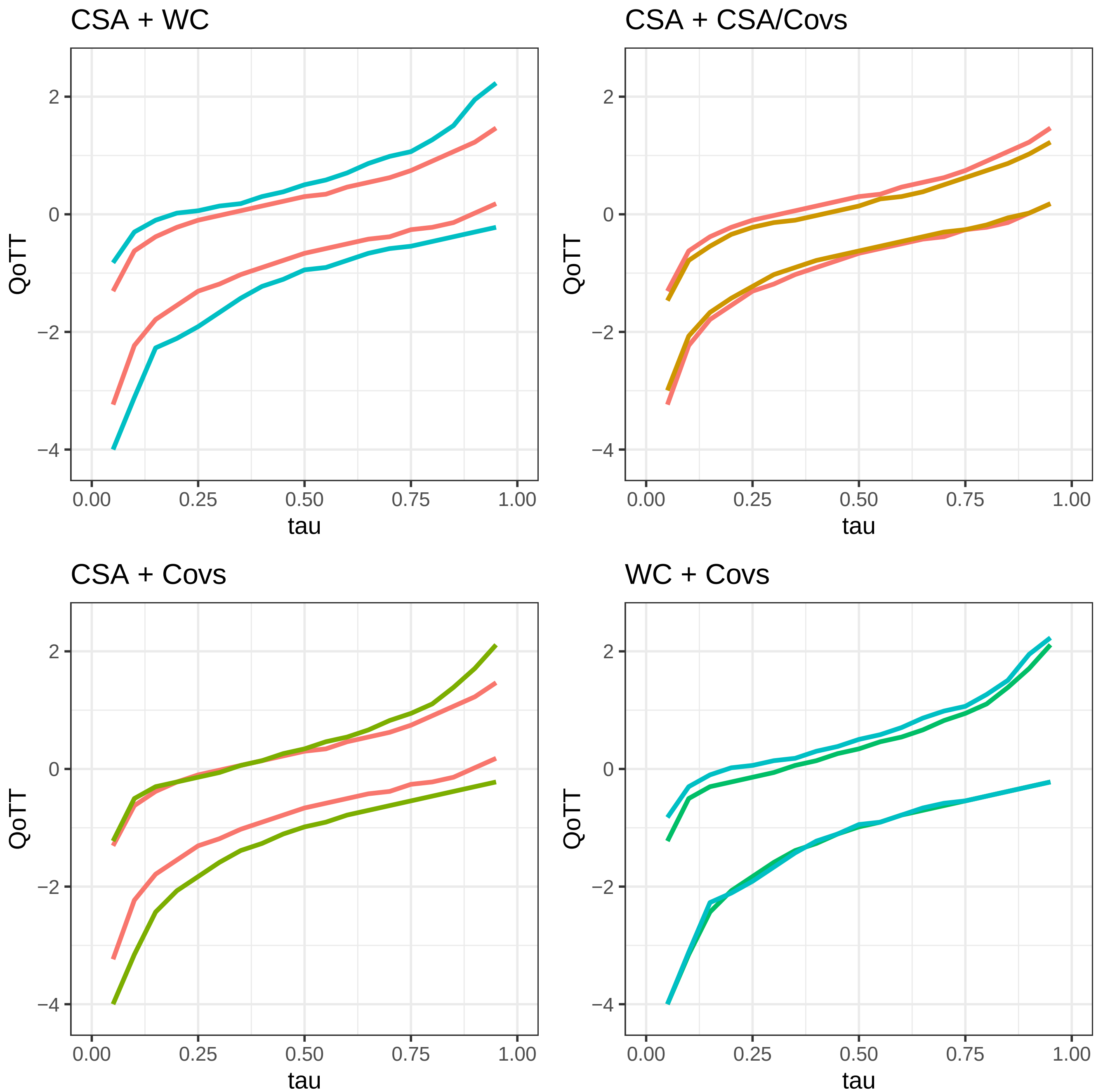

*Notes:* These are worst-case bounds that only use information from the marginal distributions of treated and untreated potential outcomes. The scale of the y-axis is in log points. Most of the reported results in the text convert log points into percentage changes (see Footnote 19). The dotted lines provide 95% confidence intervals for the estimated lower and upper bounds using the numerical bootstrap as discussed in the text.
*Sources:* 1979 National Longitudinal Survey of Youth

Figure 5: Bounds on the Quantile of the Treatment Effect under the Copula Stability Assumption



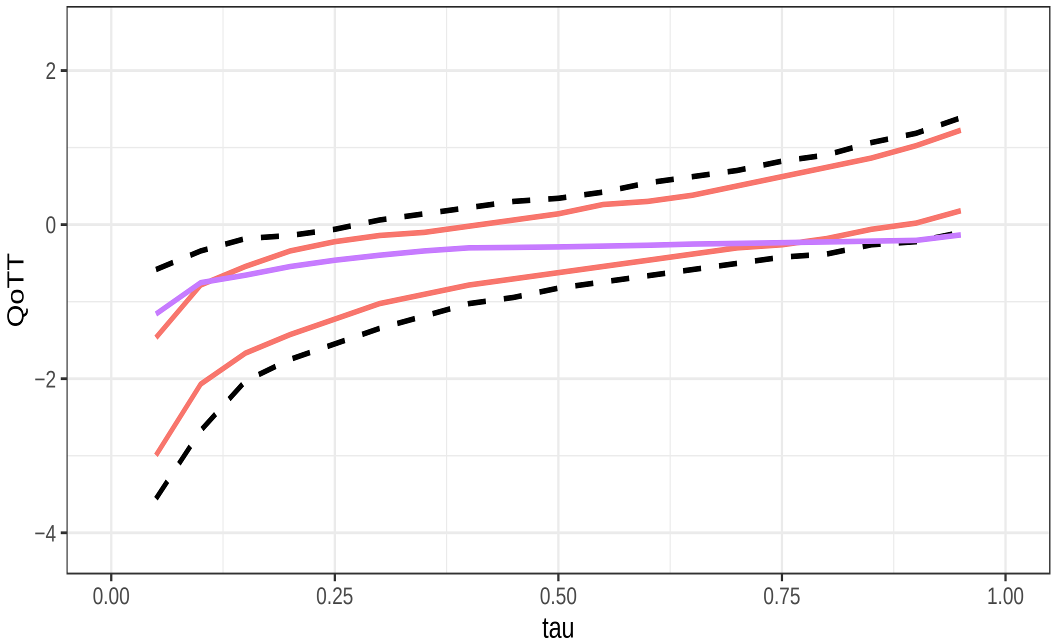

*Notes:* These are bounds that come from using the method developed in the current paper under the Copula Stability Assumption. The scale of the y-axis is in log points. Most of the reported results in the text convert log points into percentage changes (see Footnote 19). The dotted lines provide 95% confidence intervals for the estimated lower and upper bounds using the numerical bootstrap as discussed in the text.
*Sources:* 1979 National Longitudinal Survey of Youth

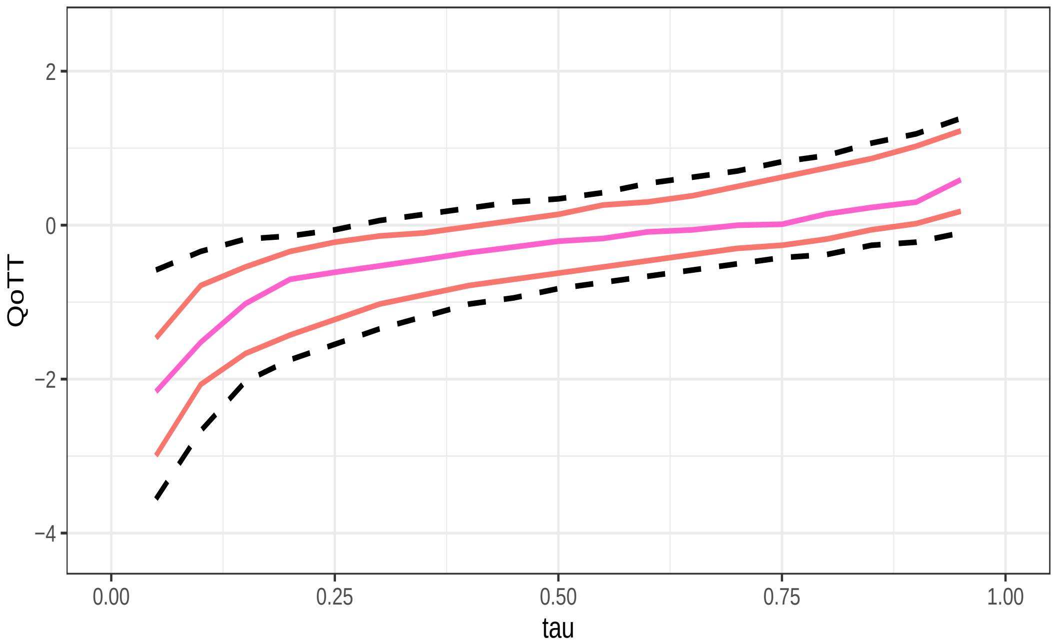
Figure 6: Comparison of Bounds under Different Conditions




*Notes:* The figure includes bounds under different conditions presented in the main text. The top left panel includes bounds under the Copula Stability Assumption along with the worst-case bounds (these the same as what is reported in Figures 4 and 5). The top right panel compares bounds under the Copula Stability Assumption in the unconditional case to bounds under Copula Stability Assumption that also use covariates to further tighten the bounds. The bottom left panel compares bounds under the Copula Stability Assumption to bounds that tighten the worst-case bounds only using covariates. The bottom right panel compares the worst-case bounds to the bounds that only use covariates to obtain tighter bounds. The scale of the y-axis is in log points. Most of the reported results in the text convert log points into percentage changes (see Footnote 19). Plots that include confidence intervals for the bounds in each case are available in the Supplementary Appendix.
*Sources:* 1979 National Longitudinal Survey of Youth

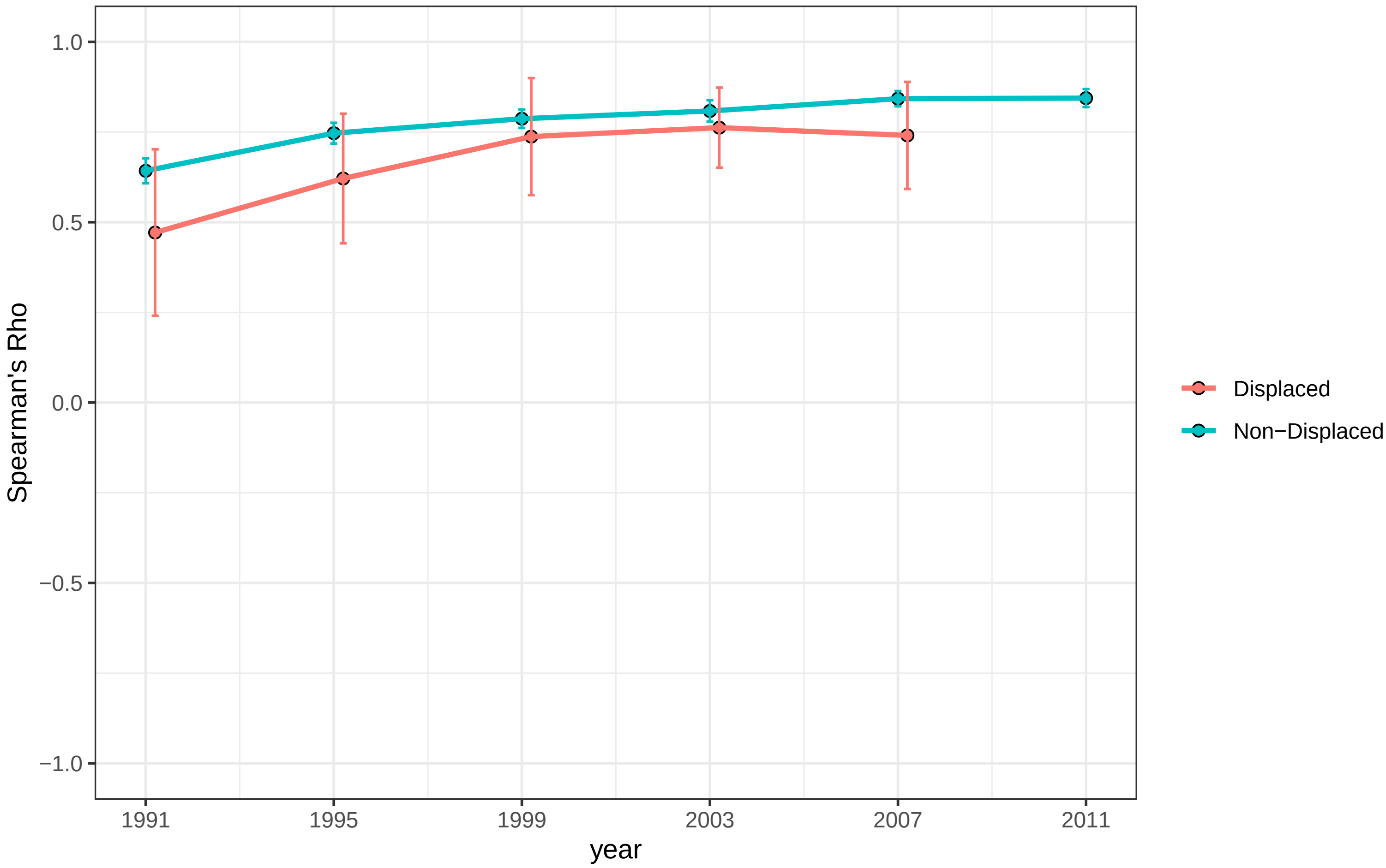
Figure 7: Plots of the QoTT under Rank Invariance Assumptions




*Notes:* This figure provides plots of the QoTT under the assumption of Cross Sectional Rank Invariance (top panel) and Rank Invariance Over Time (bottom panel). The red lines are the bounds on the QoTT under the Copula Stability Assumption and are the same as in Figure 5. The scale of the y-axis is in log points. Most of the reported results in the text convert log points into percentage changes (see Footnote 19).
*Sources:* 1979 National Longitudinal Survey of Youth

Figure 8: Spearman's Rho for every four year 1983-2007




*Notes:* This figure provides estimates of Spearman's Rho for the group of displaced workers and the group of non-displaced workers. Spearman's Rho is the correlation of the ranks of earnings in period $t$ and $t-1$ and depends only on the copula of earnings in period $t$ and period $t-1$. The sample includes a subset of the dataset used in the main analysis that includes 1,993 individuals that have positive earnings in each year from 1983-2011. Standard errors are computed using the block bootstrap with 1000 iterations.
*Sources:* 1979 National Longitudinal Survey of Youth