Supplementary Appendix: Bounds on Distributional Treatment Effect Parameters using Panel Data with an Application on Job Displacement

Brantly Callaway*

February 24, 2020

This Supplementary Appendix contains a number of additional results for "Bounds on Distributional Treatment Effect Parameters using Panel Data with an Application on Job Displacement." Appendix SA contains additional discussion of alternative rank invariance assumptions that lead to point identification of the distributional parameters contained in the main text as well as some additional empirical evidence on the Copula Stability Assumption. Appendix SB contains some Monte Carlo simulations to assess the finite sample properties of the estimators discussed in the paper. Appendix SC contains the proofs for the identification results in Lemma 1 and Propositions 2 and 3 in the main text. Appendix SD contains the proofs for the main asymptotic results in Section 4 of the main text. Appendix SE contains some additional low-level assumptions and results in the particular case where one uses distribution regression, quantile regression, and Change in Changes to estimate parameters of interest. Appendix SF develops a nonparametric pre-test of the Copula Stability Assumption. Appendix SG contains additional results for the application on job displacement.

SA More Details on Rank Invariance Assumptions and Evidence on the Copula Stability Assumption

This section contains (i) a more detailed discussion of two rank invariance assumptions that can be used to point identify parameters of interest, and (ii) additional empirical evidence on the Copula Stability Assumption in the particular case of the copula of yearly earnings in the United States since the middle of the 20th century.

SA.1 Alternative Approaches that Lead to Point Identification

This section provides additional details on two leading rank invariance assumptions that result in point identification of any parameters that depend on the joint distribution of treated and untreated

*Department of Economics, University of Mississippi, 319 Odom Hall, University, MS 38677. Email: bmcallaw@olemiss.edu

potential outcomes. The first assumption is cross-sectional rank invariance. 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 restricts 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, for example, individuals who would have been at the top of the earnings distribution if they had not been displaced 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 (in the main text); 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 without requiring the limiting case of rank invariance over time to hold exactly.

SA.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)),[1] is the same over time.[2]

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.[3] 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 SA.1).

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

SB Monte Carlo Simulations

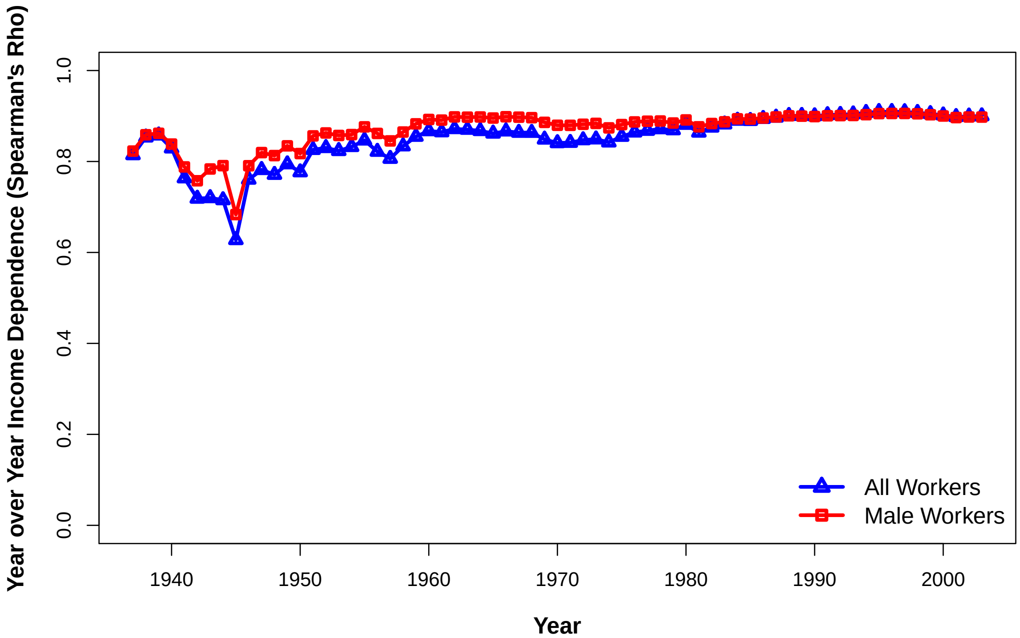
For the first set of results, I consider the finite sample performance of estimators of the upper bound of the QoTT. To keep things simple, I consider the case where the distribution of $(Y_{0t}, Y_{0t-1})|D = 1$

[1] 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.

[2] 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)).

[3] 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. See also the related discussion on pre-testing the Copula Stability Assumption in Section 4 of the main text.

Figure SA.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.

is known rather than needing to be estimated in a first step. In particular, I consider the case where $(Y_{1t}, Y_{0t-1})|D = 1 \sim N(0, V_1)$ and $(Y_{0t}, Y_{0t-1})|D = 1 \sim N(0, V_0)$ with

$$V_j = \begin{pmatrix} 1 & \rho_j \\ \rho_j & 1 \end{pmatrix}$$

for $j = 0, 1$.

With a (bivariate) Gaussian distribution, the results in Chen and Fan (2006) and Bouyé and Salmon (2009) imply that

$$P(Y_{jt} \leq y | Y_{0t-1} = y') = \Phi\left(\frac{(y - \rho_j y')}{\sqrt{1 - \rho_j^2}}\right) \tag{SB.1}$$

where Φ is the cdf of a standard normal random variable. Then, the bounds on the QoTT in Theorem 3 (in the main text) can be simulated quite accurately.

For each of the below simulations, I consider the case where $\rho_1 = 0$, and I vary $\rho_0 \in \{0, 0.5, 0.9\}$, and I report the bias, median absolute deviation, and root mean squared error for the upper bound on the 10th percentile, median, and 90th percentile of the treatment effect. I vary the number of observations between 100, 500, and 1000 and randomly assign half of them to being in the treated

Table SB.1: Monte Carlo Simulations

	N=100			**N=500**			**N=1000**		
	10%	50%	90%	10%	50%	90%	10%	50%	90%
$\rho_0 = 0$									
Actual	0.249	1.373	3.301						
Bias	−0.259	−0.194	−0.334	−0.096	−0.103	−0.189	−0.062	−0.081	−0.165
MAD	0.225	0.193	0.321	0.096	0.096	0.193	0.064	0.096	0.161
RMSE	0.300	0.258	0.407	0.118	0.131	0.218	0.081	0.101	0.185
$\rho_0 = 0.5$									
Actual	0.056	1.309	3.172						
Bias	−0.284	−0.215	−0.350	−0.115	−0.099	−0.183	−0.056	−0.081	−0.164
MAD	0.289	0.225	0.321	0.129	0.096	0.193	0.064	0.096	0.161
RMSE	0.335	0.275	0.422	0.143	0.126	0.212	0.100	0.097	0.179
$\rho_0 = 0.9$									
Actual	−0.779	0.859	2.691						
Bias	−0.168	−0.185	−0.331	−0.063	−0.084	−0.186	−0.068	−0.065	−0.168
MAD	0.193	0.193	0.321	0.064	0.096	0.193	0.064	0.064	0.161
RMSE	0.264	0.241	0.380	0.108	0.109	0.206	0.092	0.081	0.180

Notes: This table presents results from Monte Carlo experiments using the DGP given in Equation (SB.1). Each section in the table provides estimates across different values of ρ_0 which is the correlation between Y_{0t} and Y_{0t-1}. The rows labeled "Actual" give the actual upper bound on the QoTT for across different quantiles. These change with different values of ρ_0. "MAD" stands for median absolute deviation, and "RMSE" stands for root mean squared error. The columns provide simulated estimates of the bias, MAD, and RMSE at the 10th percentile, median, and 90th percentile of the treatment effect and using 1000 Monte Carlo simulations.

group.

The results of these Monte Carlo simulations are available in Table SB.1. The main takeaways are as follows. First, across all values of ρ_0, the performance of the estimators of the upper bound of the QoTT improves with larger sample sizes. Second, the estimators tend to be downward biased overall with the bias decreasing with the sample size. Similar results (which are not shown) hold for the lower bound on the QoTT – estimates of the lower bound on the QoTT tend to be upward biased. This is not surprising because the upper bound on the QoTT comes from inverting the lower bound on the DoTT – the lower bound on the DoTT is likely to tend to be downward biased because of sampling variation and taking the infimum in Lemma 3 (in the main text); see the related discussion in Manski and Pepper (2000) and Chernozhukov, Lee, and Rosen (2013). Third, the performance of the estimator of the QoTT is somewhat better overall for the median than for the 10th- or 90th-percentiles. Finally, there are not systematic differences in the performance of the estimator across

Table SB.2: Bounds on QoTT

ρ_1 / τ / ρ_0	**0.00**		**0.50**		**0.90**		**0.99**	
	0.5	0.9	0.5	0.9	0.5	0.9	0.5	0.9
0.00	2.683	3.544	2.583	3.203	1.702	1.922	0.701	0.741
0.50	2.583	3.223	2.322	3.063	1.582	1.882	0.681	0.741
0.90	1.722	1.902	1.582	1.882	1.161	1.542	0.561	0.681
0.99	0.701	0.741	0.681	0.761	0.561	0.681	0.380	0.501

Notes: The table provides bounds on $QoTT(\tau)$ for $\tau \in \{0.5, 0.9\}$ using the DGP given in this section. Each cell in the table provides $QoTT^U(\tau) - QoTT^L(\tau)$ for different values of the correlation parameters – this is the width of the identified set for different values of the parameters. Each row contains different values of ρ_0 – the correlation between Y_{0t} and Y_{0t-1}. Columns contain different values of ρ_1 – the correlation between Y_{1t} and Y_{0t-1} as well as different results across different values of τ. (Results for $\tau = 0.1$ are not reported because, due to the symmetry of the DGP, the results are the same as for $\tau = 0.9$).

different values of ρ_0 even though larger values of ρ_0 lead to substantially tighter bounds.

For the second set of results, I consider how tight the bounds are under different data generating processes. Here, I use the same DGP as given above. Under this DGP, $Y_{1t}|Y_{0t-1}, D = 1 \sim N(\rho_1 Y_{0t-1}, (1-\rho_1^2))$; similarly, $Y_{0t}|Y_{0t-1}, D = 1 \sim N(\rho_0 Y_{0t-1}, (1-\rho_0^2))$. Building on the results in Fan and Park (2010) and Fan and Wu (2010), it holds that bounds on the conditional DoTT are given by

$$F^L_{Y_{1t}-Y_{0t}|Y_{0t-1},D=1}(\delta|y') = \Phi\left(\frac{\sigma_1 s - \sigma_0 t}{\rho_0^2 - \rho_1^2}\right) + \Phi\left(\frac{\sigma_1 t - \sigma_0 s}{\rho_0^2 - \rho_1^2}\right) - 1$$

$$F^U_{Y_{1t}-Y_{0t}|Y_{0t-1},D=1}(\delta|y') = \Phi\left(\frac{\sigma_1 s + \sigma_0 t}{\rho_0^2 - \rho_1^2}\right) - \Phi\left(\frac{\sigma_1 t + \sigma_0 s}{\rho_0^2 - \rho_1^2}\right) + 1$$

where Φ is the cdf of a random variable that follows a standard normal distribution and where $\sigma_1^2 := (1-\rho_1^2)$, $\sigma_0^2 := (1-\rho_0^2)$, $s = \delta - (\rho_1 - \rho_0)y'$, and $t = \sqrt{s^2 + (\sigma_1^2 - \sigma_0^2)\log\left(\frac{\sigma_1^2}{\sigma_0^2}\right)}$; this holds in the case when $\rho_1 \neq \rho_0$. When $\rho_1 = \rho_0$, the bounds are given by

$$F^L_{Y_{1t}-Y_{0t}|Y_{0t-1},D=1}(\delta|y') = \mathbb{1}\{\delta \geq (\rho_1 - \rho_0)y'\} \left(2\Phi\left(\frac{\delta - (\rho_1 - \rho_0)y'}{2\sigma}\right) - 1\right)$$

$$F^U_{Y_{1t}-Y_{0t}|Y_{0t-1},D=1}(\delta|y') = \mathbb{1}\{\delta \geq (\rho_1 - \rho_0)y'\} + \mathbb{1}\{\delta \geq (\rho_1 - \rho_0)y'\}2\Phi\left(\frac{\delta - (\rho_1 - \rho_0)y'}{2\sigma}\right)$$

and the bounds on the $DoTT$ itself (and corresponding bounds on the QoTT) are given by averaging the conditional bounds over the distribution of $Y_{0t-1}|D = 1$.

Bounds on $QoTT(0.5)$ and $QoTT(0.9)$ are presented in Table SB.2. There are a few main patterns to notice. First, at least in this case, the bounds have roughly equal length for different values of τ. The bounds get tighter when there is stronger dependence between either Y_{0t} and Y_{0t-1} or Y_{1t} and Y_{0t-1}. Perhaps most interestingly, having very strong dependence between one pair of random variables tightens the bounds relatively more than having a moderate amount of dependence between both pairs of random variables. To see this, notice that the bounds are substantially tighter in the case when, for example, $\rho_0 = 0.9$ and $\rho_1 = 0$ than in the case where $\rho_0 = 0.5$ and $\rho_1 = 0.5$.

SC Proofs of Additional Identification Results

This section contains proofs for Lemma 1 and Propositions 2 and 3 that were not included in the main text.

SC.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

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

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

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

which completes the proof.

SC.2 Proof of Proposition 2

First, note that

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

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

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

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.

SC.3 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 $(V_t, V_{t-1}, \eta)|D=1$ follows the same distribution as $(V_{t-1}, V_{t-2}, \eta)|D=1$.

SD Asymptotic Results

SD.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 SE.1, SE.2 and SE.5 which are standard regularity conditions for quantile regression estimators and which are given in Appendix SE. Define the following terms,

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

In particular, under Assumption SE.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

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

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

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

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 $y=(y_1,y_2,y_3,y_4,y_5)^\top$, $x=(x_1,x_2,x_3,x_4,x_5)^\top$, $\tilde{y}=(\tilde{y}_1,\tilde{y}_2,\tilde{y}_3,\tilde{y}_4,\tilde{y}_5)^\top$, $\tilde{x}=(\tilde{x}_1,\tilde{x}_2,\tilde{x}_3,\tilde{x}_4,\tilde{x}_5)^\top$, and where

$$\psi^{(y,x)}(Y,X,D)=\begin{pmatrix}\psi_{1,t}^{(y_1,x_1)}(Y_t,X,D)\\ \psi_{1,t-1}^{(y_2,x_2)}(Y_{t-1},X,D)\\ \psi_{1,t-2}^{(y_3,x_3)}(Y_{t-2},X,D)\\ \psi_{0,t}^{(y_4,x_4)}(Y_t,X,D)\\ \psi_{0,t-1}^{(y_5,x_5)}(Y_{t-1},X,D)\end{pmatrix}$$

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 SD.1. *Under Assumptions 1 to 3, SE.1, SE.2, SE.4 and SE.5, Assumption 4 holds with*

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

Proof. The result follows immediately from Equation (SD.2) and Proposition SE.1 in Appendix SE. □

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

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

and

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

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

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

and

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

where the results in Equations (SD.5) and (SD.6) hold by Assumption 4 and from the results in Lemma SD.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 Γ_2. Next, define

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

which are indexed by Γ_1 and Γ_2, and where H is given in Equation (SE.3). Also, define

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

Further, notice that $\Psi_{\Gamma_{10},\Gamma_{20}}(\beta_0) = 0$ which is the population version of the first order condition for estimating β_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 SD.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 (SE.2) in Appendix SE. 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,

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

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 SD.1. *Under Assumptions 1 to 4 and SE.1 to SE.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 SD.2 and SD.3 below.

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

$$\begin{aligned}
&\sqrt{n}(\hat{\Psi}_{\hat{\Gamma}_1,\hat{\Gamma}_2}(\hat{\beta}_0) - \Psi_{\Gamma_{10},\Gamma_{20}}(\hat{\beta}_0)) \\
&\qquad = \sqrt{n}\left(\hat{\Psi}_{\Gamma_{10},\Gamma_{20}}(\beta_0) - \Psi_{\Gamma_{10},\Gamma_{20}}(\beta_0)\right) && \text{(SD.10)} \\
&\qquad + \sqrt{n}\left(\Psi_{\hat{\Gamma}_1,\Gamma_{20}}(\beta_0) - \Psi_{\Gamma_{10},\Gamma_{20}}(\beta_0)\right) && \text{(SD.11)} \\
&\qquad + \sqrt{n}\left(\Psi_{\Gamma_{10},\hat{\Gamma}_2}(\beta_0) - \Psi_{\Gamma_{10},\Gamma_{20}}(\beta_0)\right) && \text{(SD.12)} \\
&\qquad + \sqrt{n}\left(\hat{\Psi}_{\hat{\Gamma}_1,\hat{\Gamma}_2}(\hat{\beta}_0) - \hat{\Psi}_{\hat{\Gamma}_1,\hat{\Gamma}_2}(\beta_0) - \left(\Psi_{\Gamma_{10},\Gamma_{20}}(\hat{\beta}_0) - \Psi_{\Gamma_{10},\Gamma_{20}}(\beta_0)\right)\right) && \text{(SD.13)} \\
&\qquad + \sqrt{n}\left(\hat{\Psi}_{\hat{\Gamma}_1,\hat{\Gamma}_2}(\hat{\beta}_0) - \hat{\Psi}_{\Gamma_{10},\hat{\Gamma}_2}(\beta_0) - \left(\Psi_{\hat{\Gamma}_1,\Gamma_{20}}(\beta_0) - \Psi_{\Gamma_{10},\Gamma_{20}}(\beta_0)\right)\right) && \text{(SD.14)} \\
&\qquad + \sqrt{n}\left(\hat{\Psi}_{\Gamma_{10},\hat{\Gamma}_2}(\beta_0) - \hat{\Psi}_{\Gamma_{10},\Gamma_{20}}(\beta_0) - \left(\Psi_{\Gamma_{10},\hat{\Gamma}_2}(\beta_0) - \Psi_{\Gamma_{10},\Gamma_{20}}(\beta_0)\right)\right) && \text{(SD.15)}
\end{aligned}$$

The term in Equation (SD.10) can be handled exactly the same way as in Chernozhukov, Fernandez-Val, and Melly (2013) (see also the related discussion in Appendix SE.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 (SD.10) weakly converges to $\mathbb{Z}_{00}$ which is a tight, mean zero Gaussian process with covariance function

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

where ψ^1 is defined in Equation (SE.4) in Appendix SE.

The terms in Equations (SD.13) to (SD.15) converge uniformly to 0 using stochastic equicontinuity arguments. The terms in Equations (SD.11) and (SD.12) capture the estimation effect of the first step estimators. Thus, by Lemmas SD.2 and SD.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 SD.2 and SD.3 and where $\mathbb{Z}_{01}$ and $\mathbb{Z}_{02}$ are given in Equations (SD.5) and (SD.6). □

Lemma SD.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 Ψ_1 is Hadamard differentiable at Γ_{10} tangentially to $\mathbb{D}_0$ with derivative at Γ_{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 Γ_{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 $k = 1, 2, 3, \ldots$ 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

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

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,

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

where the third equality holds by the same argument as in Equation (SD.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 γ_{1k} is uniformly bounded. The second term converges uniformly to 0 because f_1 is uniformly bounded and $\|\gamma_{1k}-\gamma_{10}\|_\infty\to 0$. $\square$

Lemma SD.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 Ψ_2 is Hadamard differentiable at Γ_{20} tangentially to $\mathbb{D}_0$ with derivative at Γ_{20} in $\gamma_2\in\mathbb{D}_0$ given by

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

where $\beta_0^{(1)}$ is the first element in the vector β_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 (SD.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

$$
\begin{aligned}
&\frac{\Psi_2(\Gamma_{2k}) - \Psi_2(\Gamma_{20})}{t_k} - \Psi'_{2,\Gamma_{20}}(\gamma_2) \\
&\qquad = E[(\Lambda(W_{\Gamma_{2k}}^\top \beta_0) - \Lambda(W_{\Gamma_{20}}^\top \beta_0))H(W_{\Gamma_{20}}^\top \beta_0)W_{\Gamma_{20}}|D=1]/t_k - A_1 &&\text{(SD.20)}\\
&\qquad + E[(\Lambda(W_{\Gamma_{20}}^\top \beta_0) - \mathbb{1}\{\Gamma_{10}(Y_{t-1}) \le y\}) &&\text{(SD.21)}\\
&\qquad\qquad \times \left(H(W_{\Gamma_{2k}}^\top \beta_0) - H(W_{\Gamma_{20}}^\top \beta_0)\right) W_{\Gamma_{20}}|D=1]/t_k - A_2 \\
&\qquad + E[(\Lambda(\Gamma_{20}(Y_{t-2}^\top \beta_0)) - \mathbb{1}\{\Gamma_{10}(Y_{t-1}) \le y\}) &&\text{(SD.22)}\\
&\qquad\qquad \times H(W_{\Gamma_{20}}^\top \beta_0)\left(W_{\Gamma_{2k}} - W_{\Gamma_{20}}\right)|D=1]/t_k - A_3
\end{aligned}
$$

which holds uniformly and up to some smaller order terms. For Equation (SD.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 (SD.20) converges uniformly to 0 because $\bar{\Gamma}_2$ converges uniformly to Γ_{20}, γ_{2k} is uniformly bounded, and γ_{2k} converges uniformly to γ_2.

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

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

where h is the derivative of H. The first term above converges to 0 because Γ_{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 γ_{2k} converges uniformly to γ_2. Finally, for Equation (SD.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

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

which converges uniformly to 0 since γ_{2k} converges uniformly to γ_2 and each of the other terms are uniformly bounded. $\square$

Lemma SD.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* ψ *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). □

SD.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 SD.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* ϕ_3^U *is Hadamard directionally differentiable at* $\theta_0 \in \mathbb{D}_{\phi_3^U}$ *tangentially to* $\mathbb{D}_{\phi_3^U}$ *in* $\varphi \in \mathbb{D}_{\phi_3^U}$ *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 SD.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* ϕ_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*

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

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

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[4]

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

Then, the following results hold

Lemma SD.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 $\theta \in \mathbb{D}_{\phi_2^L} := \mathcal{C}(\mathcal{Y}_\delta \boldsymbol{\Delta}\mathcal{Y}_{1t-1}\mathcal{X}_1)$. *Then, the map* ϕ_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*

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

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

Lemma SD.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 $\theta \in \mathbb{D}_{\phi_2^U} := \mathcal{C}(\mathcal{Y}_\delta \boldsymbol{\Delta}\mathcal{Y}_{1t-1}\mathcal{X}_1)$. *Then, the map* ϕ_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*

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

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

[4]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 δ. See the discussion in Fan and Park (2010, Section 3).

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 SD.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* Λ_2 *is a probability measure on* $\mathcal{Y}_{1t-1}\mathcal{X}_1$*. Then, the map* ϕ_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) □

SD.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 ϕ_1, ϕ_2^L, ϕ_2^U, ϕ_3^L, and ϕ_3^U in the previous section, notice that

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

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 SD.5 to SD.9 and by the chain rule for Hadamard directionally differentiable functions (Shapiro (1990) and Masten and Poirier (2019)), it holds that the maps ϕ^L and ϕ^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

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

and

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

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

In addition, the limiting processes in Proposition 4 are given by $\mathbb{V}^L := \mathbb{V}_0^L + \mathbb{V}_1^L$ 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

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

where

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

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.

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 $DoTT^L$.

SE Supplementary Asymptotic Results

In the first part of this section, I provide some additional low-level assumptions on the preliminary estimators used in the main part of the text. In cases where a researcher used alternative first step estimators, these regularity conditions would need to be adjusted. The second set of results are on using Change in Changes (Athey and Imbens (2006) and Melly and Santangelo (2015)) in the first step to identify and estimate $F_{Y_{0t}|X,D=1}$. As discussed earlier, other approaches could be used to identify this counterfactual distribution, but the second part of this section supplies additional details for the method that I actually use in the main paper.

SE.1 Additional Assumptions for First Step Estimators

The first two assumptions are technical conditions used in deriving the main asymptotic results in the paper.

Assumption SE.1 (Compact Support).

For all $(s,d) \in \{t, t-1, t-2\} \times \{0,1\}$, $\mathcal{Y}_{ds}$ and $\mathcal{X}_d$, which denote the supports of Y_s and X conditional on $D=d$, are compact subsets of $\mathbb{R}$.

Assumption SE.2 (Continuously Distributed Outcomes).

(i) For all $(s,d) \in \{t, t-1, t-2\} \times \{0,1\}$, Y_s is continuously distributed conditional on X and $D=d$ with conditional density $f_{Y_d|X,D=d}(y|x)$ that is uniformly bounded away from 0 and ∞ and uniformly continuous in $(y,x) \in \mathcal{Y}_{ds}\mathcal{X}_d$.

(ii) Conditional on Y_{0t-1}, X, and $D=1$, Y_{1t} and Y_{0t} are continuously distributed with conditional densities $f_{Y_{1t}|Y_{0t-1},X,D=1}(y|y',x)$ and $f_{Y_{0t}|Y_{0t-1},X,D=1}(y|y',x)$ that are uniformly bounded away from 0 and ∞ and uniformly continuous in (y,y',x) on their supports.

The next assumption provides additional regularity conditions for the proposed distribution regression estimators of $F_{Y_{1t}|Y_{0t-1},X,D=1}$ and $F_{Y_{0t}|Y_{0t-1},X,D=1}$ which are also standard in the literature on distribution regression. First, for $W_{t-1} = (Y_{0t-1}, X^\top)^\top$ and $W_{\Gamma_{20}} = (\Gamma_{20}(Y_{0t-2},X), X^\top)^\top$, define

$$M_1(y) := E\left[\frac{\lambda(W_{t-1}^\top \beta_1(y))^2}{\Lambda(W_{t-1}^\top \beta_1(y))(1-\Lambda(W_{t-1}^\top \beta_1(y)))} W_{t-1}W_{t-1}^\top \Big| D=1\right] \tag{SE.1}$$

and

$$M_0(y) := E\left[\frac{\lambda(W_{\Gamma_{20}}^\top \beta_0(y))^2}{\Lambda(W_{\Gamma_{20}}^\top \beta_0(y))(1-\Lambda(W_{\Gamma_{20}} \top \beta_0(y)))} W_{\Gamma_{20}}W_{\Gamma_{20}}^\top \Big| D=1\right] \tag{SE.2}$$

Assumption SE.3 (Distribution Regression).

(i) $E\left[\|W_{t-1}\|^2|D=1\right] < \infty$ *and* $E\left[\|W_{\Gamma_{20}}\|^2|D=1\right] < \infty$.

(ii) The minimum eigenvalues of $M_1(y)$ *and* $M_0(y)$*, which are defined in Equations* (SE.1) *and* (SE.2)*, are uniformly bounded away from zero.*

Next, define $p_d := P(D=d)$ and $p_d(x) := P(D=d|X=x)$. The following assumptions are needed for the particular first step estimators that I use in the application.

Assumption SE.4 (Overlap).

$p_1 > 0$ *and, for all* $x \in \mathcal{X}_1$, $p_1(x) < 1$.

This assumption is standard in the treatment effects literature. The first part says that there are some treated individuals; the second part says that for any possible values of the covariates for the treated group, there is a positive probability that they there do not participate in the treatment. This guarantees that, for individuals in the treated group, one can find "matches" with the same characteristics. Because all the parameters that are considered in the paper are conditional on being in the treated group, I do not require that that the propensity score, $P(D=1|X)$, be bounded away from 0.[5]

The next assumption is an additional condition for the first step quantile regression estimators and is standard in the literature on quantile regression.

Assumption SE.5 (First Step Quantile Regression).

(i) For $d \in \{0,1\}$, $E\left[\|X\|^{2+\varepsilon}|D=d\right] < \infty$ *for some* $\varepsilon > 0$.

(ii) For $\{s,d\} \in \{0,1\} \times \{t, t-1, t-2\}$*, the minimum eigenvalues of* $J_{s,d}(u)$*, which is defined in Equation* (SD.1)*, are uniformly bounded away from zero.*

SE.2 Additional Details for Distribution Regression

Next, in this section, I discuss the results of Chernozhukov, Fernandez-Val, and Melly (2013) which apply directly to estimating $F_{Y_{1t}|Y_{0t-1},X,D=1}$ because this is just a distribution regression of an observed outcome on some observed covariates. Here, I suppose that $F_{Y_{1t}|Y_{0t-1},X,D=1}(y|y',x) = \Lambda(w^\top \beta_1(y))$ for some known link function Λ with derivative λ and where $w = (y', x^\top)^\top$. Define

$$H(z) = \frac{\lambda(z)}{\Lambda(z)(1-\Lambda(z))} \tag{SE.3}$$

and let h denote the derivative of H. Let $\Psi^1(\beta)$ and $\hat{\Psi}^1(\beta)$ be defined as in Equations (SD.7) and (SD.8) but for estimating $F_{Y_{1t}|Y_{0t-1},X,D=1}$. That is, they are the population and sample first order conditions for estimating β_1.

[5] This can potentially be important in applications like job displacement where job displacement may be much less common for individuals with particular characteristics.

Using the same arguments as in Chernozhukov, Fernandez-Val, and Melly (2013), one can show that

$$\sqrt{n}(\hat{\Psi}^1(\beta_1) - \Psi^1(\beta_1)) \rightsquigarrow \mathbb{Z}_1$$

where

$$\Psi^1(\beta) = E\Big[\psi^1_{y;\beta}(Y_t, W_{t-1}, D)\Big] \quad \text{and} \quad \hat{\Psi}^1(\beta) = \frac{1}{n}\sum_{i=1}^n \psi^1_{y;\beta}(Y_{it}, W_{it-1}, D_i)$$

with

$$\psi^d_{y;\beta}(Y, W, D) = \frac{\mathbb{1}\{D = d\}}{p_d}\big(\Lambda(W^\top\beta(y)) - \mathbb{1}\{Y \leq y\}\big)H(W^\top\beta(y))W \tag{SE.4}$$

where $\mathbb{Z}_1$ is a tight, mean zero Gaussian process with covariance function

$$V_{Z_1}(\tilde{y}_1, \tilde{y}_2) := E\big[\psi^1_{\tilde{y}_1;\beta_1}(Y_t, W_{t-1}, D)\psi^1_{\tilde{y}_2;\beta_1}(Y_t, W_{t-1}, D)^\top\big]$$

Continuing to follow the same arguments as in Chernozhukov, Fernandez-Val, and Melly (2013), it further holds that

$$\sqrt{n}(\hat{\beta}_1 - \beta_1) \rightsquigarrow -M_1(\cdot)^{-1}\mathbb{Z}_1$$

and

$$\sqrt{n}(\hat{F}_{Y_{1t}|Y_{0t-1},X,D=1} - F_{Y_{1t}|Y_{0t-1},X,D=1}) \rightsquigarrow \mathbb{G}_1 := \lambda(w^\top\beta_1(y))w^\top M_1(y)^{-1}\mathbb{Z}_1 \tag{SE.5}$$

Finally, I estimate $F_{Y_{0t-1},X|D=1}$ using the empirical cdf, i.e.,

$$\hat{F}_{Y_{0t-1},X|D=1}(y', x) = \frac{1}{n}\sum_{i=1}^n \psi^{ecdf}_{(y',x)}(Y_{it-1}, X_i, D_i)$$

where

$$\psi^{ecdf}_{(y',x)}(Y, X, D) := \frac{D}{p_1}\mathbb{1}\{Y \leq y', X \leq x\}$$

Noting that $F_{Y_{0t-1},X|D=1}(y', x) = E[\psi^{ecdf}_{(y',x)}(Y, X, D)]$, it follows immediately that

$$\sqrt{n}(\hat{F}_{Y_{0t-1},X|D=1} - F_{Y_{0t-1},X|D=1}) \rightsquigarrow \mathbb{G}_3 \tag{SE.6}$$

where $\mathbb{G}_3$ is a tight, mean zero Gaussian process with covariance function given by

$$V_3(\tilde{y}_1', \tilde{x}_1, \tilde{y}_2', \tilde{x}_2) \\ = E\Big[\big(\psi^{ecdf}_{(\tilde{y}_1', \tilde{x}_1}(Y_{t-1}, X, D) - F_{Y_{0t-1}, X|D=1}(\tilde{y}_1', \tilde{x}_1)\big)\big(\psi^{ecdf}_{(\tilde{y}_2', \tilde{x}_2)}(Y_{t-1}, X, D) - F_{Y_{0t-1}, X|D=1}(\tilde{y}_2', \tilde{x}_2)\big)^\top \Big]$$

SE.3 Additional Details for Change in Changes

Under Assumption 2 (in the main text), the distribution of untreated potential outcomes for individuals in the treated group is identified. In the application, I used Change in Changes (Athey and Imbens (2006) and Melly and Santangelo (2015)) to identify this distribution. In particular, in this setup,

$$\begin{aligned} F_{Y_{0t}|X,D=1}(y|x) &= \phi(F_{Y_{0t-1}|X,D=1}, F_{Y_{0t-1}|X,D=0}, F_{Y_{0t}|X,D=0})(y,x) \\ &:= F_{Y_{0t-1}|X,D=1}(F^{-1}_{Y_{0t-1}|X,D=0}(F_{Y_{0t}|X,D=0}(y|x)|x)|x) \end{aligned} \tag{SE.7}$$

where all the terms on the right hand side of Equation (SE.7) are identified. It is natural then to estimate the distribution of untreated potential outcomes for the treated group by

$$\begin{aligned} \hat{F}_{Y_{0t}|X,D=1}(y|x) &= \phi(\hat{F}_{Y_{0t-1}|X,D=1}, \hat{F}_{Y_{0t-1}|X,D=0}, \hat{F}_{Y_{0t}|X,D=0})(y,x) \\ &= \hat{F}_{Y_{0t-1}|X,D=1}(\hat{F}^{-1}_{Y_{0t-1}|X,D=0}(\hat{F}_{Y_{0t}|X,D=0}(y|x)|x)|x) \end{aligned}$$

The limiting process for each of the estimated distributions is given in Equation (SD.2), so all that remains to be show is that the function ϕ is Hadamard differentiable. Next, I provide the limiting process for $\sqrt{n}(\hat{F}_{Y_{0t}|X,D=1} - F_{Y_{0t}|X,D=1})$. This is closely related to Melly and Santangelo (2015) though my method of proof is somewhat different and my main result expands one of their intermediate results.

Proposition SE.1. *Let* $\mathbb{D} = l^\infty(\mathcal{Y}_{1t-1}\mathcal{X}_1) \times l^\infty(\mathcal{Y}_{0t-1}\mathcal{X}_1) \times l^\infty(\mathcal{Y}_{0t}\mathcal{X}_1)$ *and consider the map* $\phi : \mathbb{D}_\phi \subset \mathbb{D} \mapsto l^\infty(\bar{\mathcal{Y}}_{0t}\mathcal{X}_1)$ *given by*

$$\phi(F) := F_1 \circ F_2^{-1} \circ F_3$$

for $F := (F_1, F_2, F_3) \in \mathbb{D}_\phi$ *where* $\mathbb{D}_\phi := \mathbb{E}^3$ *where* $\mathbb{E}$ *denotes the set of all conditional distribution functions with conditional density function that is uniformly bounded from above and bounded away from zero. Then, the map* ϕ *is Hadamard differentiable at* $F_0 = (F_{10}, F_{20}, F_{30}) \in \mathbb{D}$ *with derivative given by*

$$\phi'_{F_0}(\lambda) = \lambda_1 \circ F_{20}^{-1} \circ F_{30} + f_{10}(F_{20}^{-1} \circ F_{30}) \frac{\lambda_3 - \lambda_2 \circ F_{20}^{-1} \circ F_{30}}{f_{20}(F_{20}^{-1} \circ F_{30})}$$

tangentially to $\mathbb{D}_\phi$ *in* $\lambda = (\lambda_1, \lambda_2, \lambda_3) \in \mathbb{D}_\phi$.

Proof. Let $\mathbb{D}_{\phi_2} = \mathbb{E} \times \mathbb{E}^{-1} \times \mathbb{E}$ where $\mathbb{E}^{-1}$ is the space of inverse functions in $\mathbb{E}$. Consider the maps $\phi_1 : \mathbb{D}_\phi \mapsto \mathbb{D}_{\phi_2}$ and $\phi_2 : \mathbb{D}_{\phi_2} \mapsto l^\infty(\bar{\mathcal{Y}}_{0t}\mathcal{X}_1)$ given by

$$\phi_1(F) = (F_1, F_2^{-1}, F_3) \quad \text{and} \quad \phi_2(\Gamma) = (\Gamma_1 \circ \Gamma_2 \circ \Gamma_3)$$

for $F = (F_1, F_2, F_3) \in \mathbb{D}_\phi$ and $\Gamma = (\Gamma_1, \Gamma_2, \Gamma_3) \in \mathbb{D}_{\phi_2}$. Notice that $\phi(F) = \phi_2(\phi_1(F))$. Next, the map ϕ_1 is Hadamard differentiable at F_0 with derivative given by

$$\phi'_{1,F_0}(\lambda) = \left(\lambda_1, -\frac{\lambda_2}{f_{20}} \circ F_{20}^{-1}, \lambda_3\right)$$

(see, for example, van der Vaart and Wellner (1996, Lemma 3.9.23(ii))). Next, the map ϕ_2 is Hadamard differentiable at Γ_0 in $\gamma \in \mathbb{D}_{\phi_2}$ with derivative given by

$$\phi'_{2,\Gamma_0}(\gamma) = \gamma_1 \circ \Gamma_{20} \circ \Gamma_{30} + \Gamma'_{1,\Gamma_{20}\circ\Gamma_{30}}\gamma_2 \circ \Gamma_{30} + \Gamma'_{1,\Gamma_{20}\circ\Gamma_{30}}\Gamma'_{2,\Gamma_{30}}\gamma_3$$

which follows using a similar argument as in van der Vaart and Wellner (1996, Lemma 3.9.27). Further, by the chain rule for Hadamard differentiable functions and for $\lambda \in \mathbb{D}_\phi$,

$$\begin{aligned}\phi'_{F_0}(\lambda) &= \phi'_{2,\phi_1(F_0)} \circ \phi'_{1,F_0}(\lambda) \\ &= \phi'_{2,(F_{10},F_{20}^{-1},F_{30})}\left(\lambda_1, -\frac{\lambda_2}{f_{20}} \circ F_{20}^{-1}, \lambda_3\right)\end{aligned}$$

which implies the result by plugging into the expression for $\phi'_{2,\Gamma_0}(\gamma)$ and because the derivative of F_{10} is f_{10} and the derivative of F_{20}^{-1} is $1/(f_{20} \circ F_{20}^{-1})$.

□

SF A Nonparametric Pre-Test of the Copula Stability Assumption

This section develops a nonparametric pre-test for the Copula Stability Assumption that is useful in cases where a researcher has access to additional pre-treatment time periods. In the main text, I also suggested a parametric test of the Copula Stability Assumption based on computing a dependence measure such as Spearman's Rho or Kendall's Tau in all pre-treatment times periods and testing whether the dependence measure is constant over time. One advantage of the parametric test relative to the nonparametric pre-test is that it is simple to implement. On the other hand, this sort of parametric test only tests an implication of the Copula Stability Assumption in pre-treatment periods. In other words, there are violations of the Copula Stability Assumption that would not be detected by the parametric test. The nonparametric test developed in this section can detect fixed nonparametric alternatives; i.e., violations of the Copula Stability Assumption that might not be detected by the parametric test mentioned above (for example, there exist copulas that have the same value of a dependence parameter but are different from each other). In applications, relative to the parametric test, the main disadvantages of this approach are that it is more computationally challenging to implement and that it does not deliver a figure like Figure 3 in the main text; this sort of figure can be useful for thinking about the Copula Stability Assumption in applied work.

Before proceeding, it is helpful to introduce a bit of extra notation for this section. Instead of supposing that we observe data from three time periods, we now consider the case where there are t total time periods and index these by $s = 1, \ldots, t$. Let n_1 denote the number of treated observations. Also, following the same setup as in the main part of the paper, consider the case where individuals in the treated group first become treated in time period t. Therefore, this section considers the following null hypothesis to test

$$H_0 : C_{Y_{0s_1}, Y_{0s_1-1}|D=1} = C_{Y_{0s_2}, Y_{0s_2-1}|D=1} \text{ for all } 2 \leq s_1 < s_2 < t \quad \text{vs.} \quad H_1 : \text{not } H_0$$

This null hypothesis is that the Copula Stability Assumption holds across all pre-treatment time periods. Rémillard and Scaillet (2009) develop a test for the equality of two copulas. The rest of the arguments in this section extend those arguments to the case of testing the equality of two or more copulas. To begin, estimators of copulas depend on first step estimators of distribution functions; i.e., it is natural to nonparametrically estimate a copula by

$$\hat{C}_{Y_{0s_1}, Y_{0s_1-1}|D=1}(u, v) = \frac{1}{n_1} \sum_{i=1}^{n_1} \mathbb{1}\{\hat{F}_{Y_{0s_1}|D=1}(Y_{0is_1}) \leq u, \hat{F}_{Y_{0s_1-1}|D=1}(Y_{0is_1-1}) \leq v\}$$

which depends on first step estimators of $F_{Y_{0s_1}|D=1}$ and $F_{Y_{0s_1-1}|D=1}$. The limiting process for this type of nonparametric copula estimator is established in Gaenssler and Stute (1987), Fermanian, Radulovic, and Wegkamp (2004), and van der Vaart and Wellner (2007) (among some others) under some regularity conditions (see, for example, Theorem 3 in Fermanian, Radulovic, and Wegkamp (2004) or Corollary 5.3 in van der Vaart and Wellner (2007)). Using essentially the same arguments, one can show that

$$\sqrt{n_1}(\hat{C}_2 - C_2, \hat{C}_3 - C_3, \ldots, \hat{C}_{t-1} - C_{t-1}) \rightsquigarrow (\mathbb{C}_2, \mathbb{C}_3, \ldots, \mathbb{C}_{t-1}) \tag{SF.1}$$

in $l^\infty([0,1])^{t-2}$ and where, for $s = 2, 3, \ldots, t-1$, $\mathbb{C}_s$ is a mean 0 Gaussian process (an exact expression and more detailed discussion is given in Fermanian, Radulovic, and Wegkamp (2004)). Define

$$\hat{J}_{s_1,s_2}(u, v) = \hat{C}_{Y_{0s_1}, Y_{0s_1-1|D=1}}(u, v) - \hat{C}_{Y_{0s_2}, Y_{0s_2-1}|D=1}(u, v) \tag{SF.2}$$

Under H_0,

$$\sqrt{n}_1 \hat{J}_{s_1,s_2} \rightsquigarrow \mathbb{C}_{s_1} - \mathbb{C}_{s_2}$$

and this holds jointly for $2 \leq s_1 < s_2 < t$. Then, I consider the following Cramer von Mises type of test statistic

$$\widehat{CvM} = \sum_{s_1=2}^{t-1} \sum_{s_2=2}^{t-1} \mathbb{1}\{s_1 < s_2\} \int_{[0,1]^2} \left(\sqrt{n}_1 \hat{J}_{s_1,s_2}(u,v)\right)^2 du\, dv$$

and where under H_0

$$\widehat{CvM} \xrightarrow{d} \sum_{s_1=2}^{t-1} \sum_{s_2=2}^{t-1} \mathbb{1}\{s_1 < s_2\} \int_{[0,1]^2} \left(\mathbb{C}_{s_1}(u,v) - \mathbb{C}_{s_2}(u,v)\right)^2 du\, dv$$

which holds by the continuous mapping theorem. On the other hand, under H_1, $\widehat{CvM}$ diverges. Under H_0, $\widehat{CvM}$ follows a nonstandard limiting distribution, but, following standard arguments, its limiting distribution can be simulated. In particular, Fermanian, Radulovic, and Wegkamp (2004) establish that the standard empirical bootstrap can be used to approximate the limiting process of the copula; in particular, let $\hat{C}_s^* = \hat{C}^*_{Y_{0s}, Y_{0s-1}|D=1}$ which is an estimated copula using a bootstrapped sample (i.e., one drawn from the original data with replacement). From the arguments in Fermanian, Radulovic, and Wegkamp (2004), it follows that

$$\sqrt{n}_1(\hat{C}_2^* - \hat{C}_2, \hat{C}_3^* - \hat{C}_3, \ldots, \hat{C}_{t-1}^* - \hat{C}_{t-1}) \rightsquigarrow_* (\mathbb{C}_2, \mathbb{C}_3 \ldots, \mathbb{C}_{t-1})$$

which is the same limiting process as in Equation (SF.1) and which further implies that

$$\sqrt{n}_1 \left(\hat{J}^*_{s_1,s_2} - \hat{J}_{s_1,s_2}\right) \rightsquigarrow_* \mathbb{C}_{s_1} - \mathbb{C}_{s_2}$$

which is the same process as in Equation (SF.2) and which holds jointly for $2 \leq s_1 < s_2 < t$. Finally, define the bootstrapped version of $\widehat{CvM}$ by

$$\widehat{CvM}^* = \sum_{s_1=2}^{t-1} \sum_{s_2=2}^{t-1} \mathbb{1}\{s_1 < s_2\} \int_{[0,1]^2} \left(\sqrt{n}_1(\hat{J}^*_{s_1,s_2}(u,v) - \hat{J}_{s_1,s_2}(u,v))\right)^2 du\, dv \tag{SF.3}$$

It immediately follows that

$$\widehat{CvM}^* \xrightarrow{d}_* \sum_{s_1=2}^{t-1} \sum_{s_2=2}^{t-1} \mathbb{1}\{s_1 < s_2\} \int_{[0,1]^2} \left(\mathbb{C}_{s_1}(u,v) - \mathbb{C}_{s_2}(u,v)\right)^2 du\, dv$$

which is the same limiting distribution as for $\widehat{CvM}$ under H_0. The suggests simulating the limiting distribution of $\widehat{CvM}$ under H_0 using the bootstrap; i.e., calculating the term in Equation (SF.3) B times where B is some large number. One can then test H_0 by comparing $\widehat{CvM}$ to the $(1-\alpha)$-quantile of $\{\widehat{CvM}^*_b\}_{b=1}^B$ where $b = 1, \ldots, B$ denotes a particular bootstrap iteration and α denotes a significance level.

SG More Details for the Application on Job Displacement

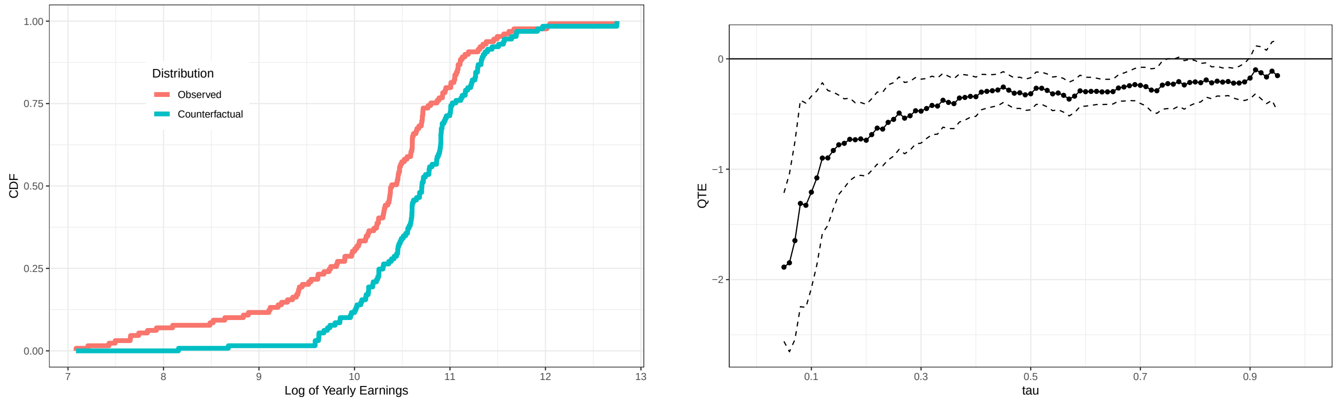
SG.1 Related Work on Job Displacement

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

Figure SG.1: Marginal Distributions of Displaced and Non-displaced Potential Outcomes for the Displaced Group



(a) Marginal Distributions of Potential Outcomes (b) Quantile Treatment Effect on the Treated

Notes: Panel (a) 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. Panel (b) provides estimates of the QTT of job displacement. 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 in the main text). The dotted lines provide pointwise 95% confidence intervals using the empirical bootstrap.

Sources: 1979 National Longitudinal Survey of Youth

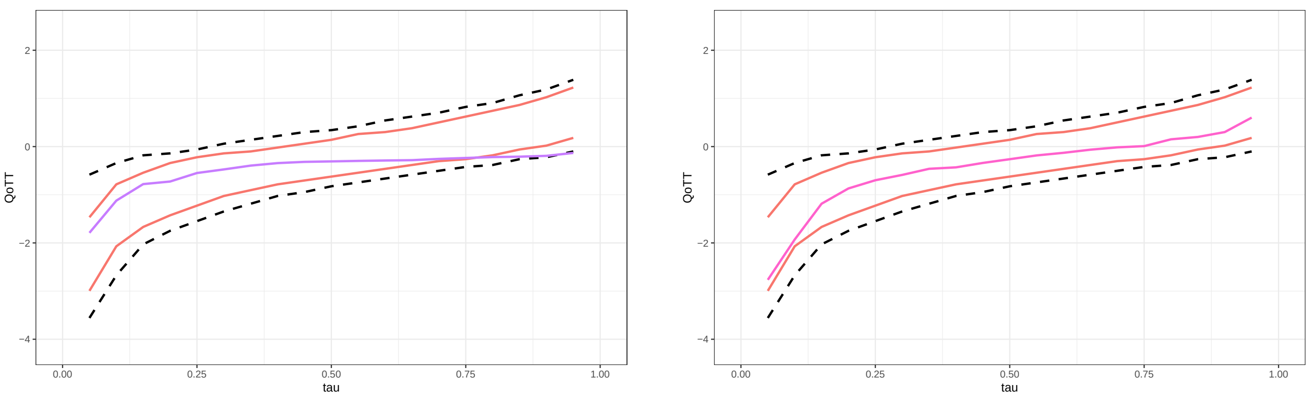
reemployment was very low with more workers being reemployed in part time jobs. Interestingly, Farber (2017) finds much heterogeneity in the effects of job displacement. First, he finds that there are significant differences in the effect of job displacement between workers who find full-time, part-time, 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.

SG.2 Additional Results on Job Displacement

This section contains several additional results for the application on job displacement. First, Figure SG.1a contains plots of the observed distribution of earnings for displaced workers as well as the counterfactual distribution of earnings that displaced workers would have experienced if they had not been displaced from their jobs. Figure SG.1b plots the QTT for displaced workers. Both of these figures are discussed in more detail in the main text.

Next, Figure SG.2 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 workers who would have been at the top of the earnings distribution if they had not been displaced 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

Figure SG.2: Plots of the QoTT under Rank Invariance Assumptions



(a) Cross-sectional rank invariance

(b) Rank invariance over time

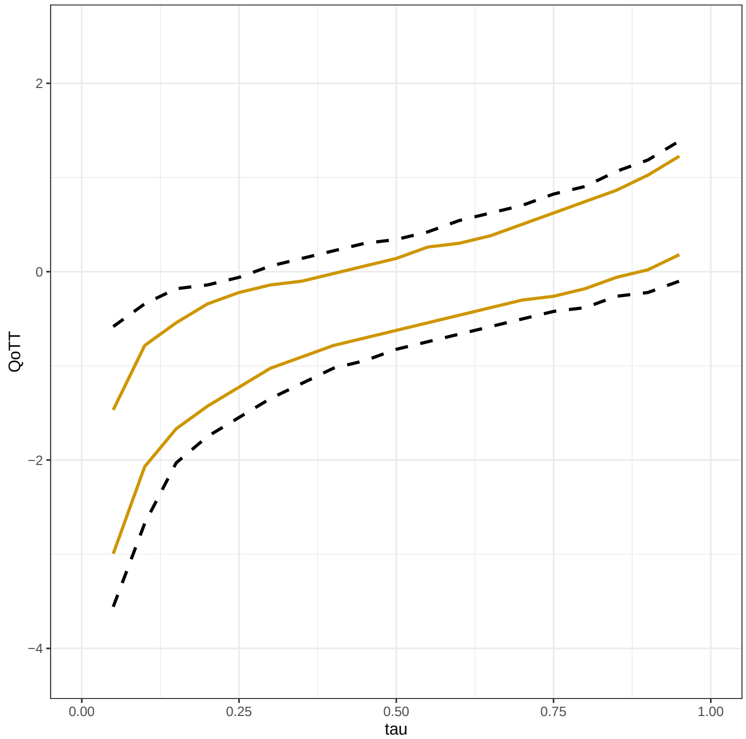
Notes: This figure provides plots of the QoTT under the assumption of cross-sectional rank invariance (left panel) and rank invariance over time (right panel). The red lines are the bounds on the QoTT under the Copula Stability Assumption and are the same as in Figure 1b. The scale of the y-axis is in log points. Most of the reported results in the text convert log points into percentage changes (see Footnote 19 in the main text).

Sources: 1979 National Longitudinal Survey of Youth

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 SG.2).

The second part of Figure SG.2 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 if 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. That being said, as discussed in the main text, one can pre-test the rank invariance over time assumption. To do this, I estimate Spearman's Rho for Y_{0t-1} and Y_{0t-2}. In order for rank invariance over time to hold in the past, it must be the case that Spearman's Rho is exactly equal to 1. Instead, I estimate that Spearman's Rho is equal to 0.79 for the group of displaced workers. Therefore, the assumption of rank invariance over time is rejected in the pre-treatment period. This also suggests that it is not likely to hold in the present periods

Figure SG.3: Bounds on the Quantile of the Treatment Effect under the Copula Stability Assumption with Covariates



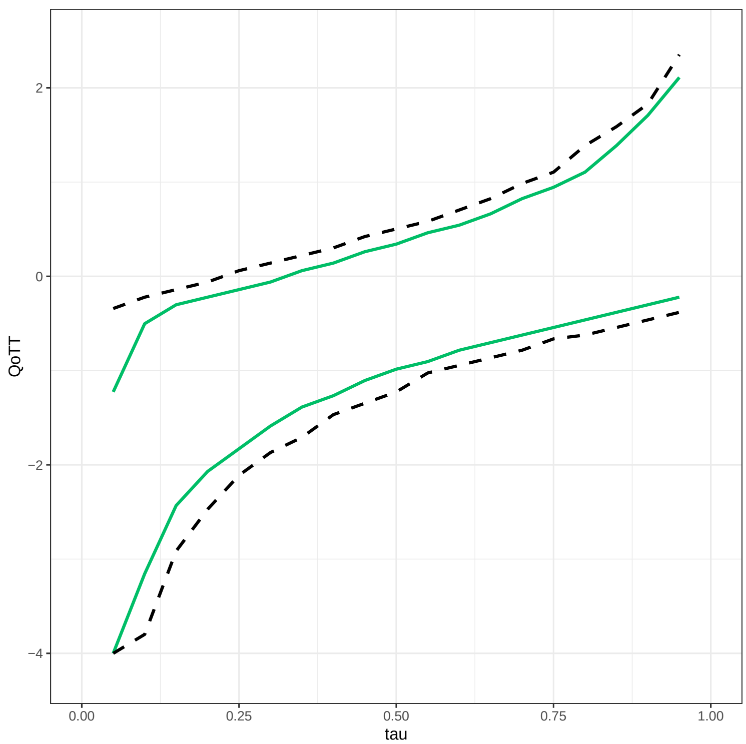
Notes: These are bounds that come from using the method developed in the current paper under the Copula Stability Assumption and through tightening bounds using covariates. The scale of the y-axis is in log points. Most of the reported results in the text convert log points into percentage changes (see Footnote 19 in the main text). 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

either. Relative to the results under cross-sectional rank invariance, the reason that the bounds in the current paper are closer to the estimates of the QoTT under rank invariance over time is that relatively stronger dependence is observed between Y_{0t-1} and Y_{0t-2} (Spearman's Rho $= 0.79$) than between Y_{1t} and Y_{0t-1} (Spearman's Rho $= 0.54$) for the group of displaced workers.

Next, I provide some additional results for the bounds on the QoTT when (i) the bounds are tightened using the Copula Stability Assumption and covariates and (ii) the bounds are tightened using covariates but not the Copula Stability Assumption. Figure 6 in the main text provides point estimates of each of these, but Figures SG.3 and SG.4 additionally provide 95% confidence intervals for the bounds in each of these cases.

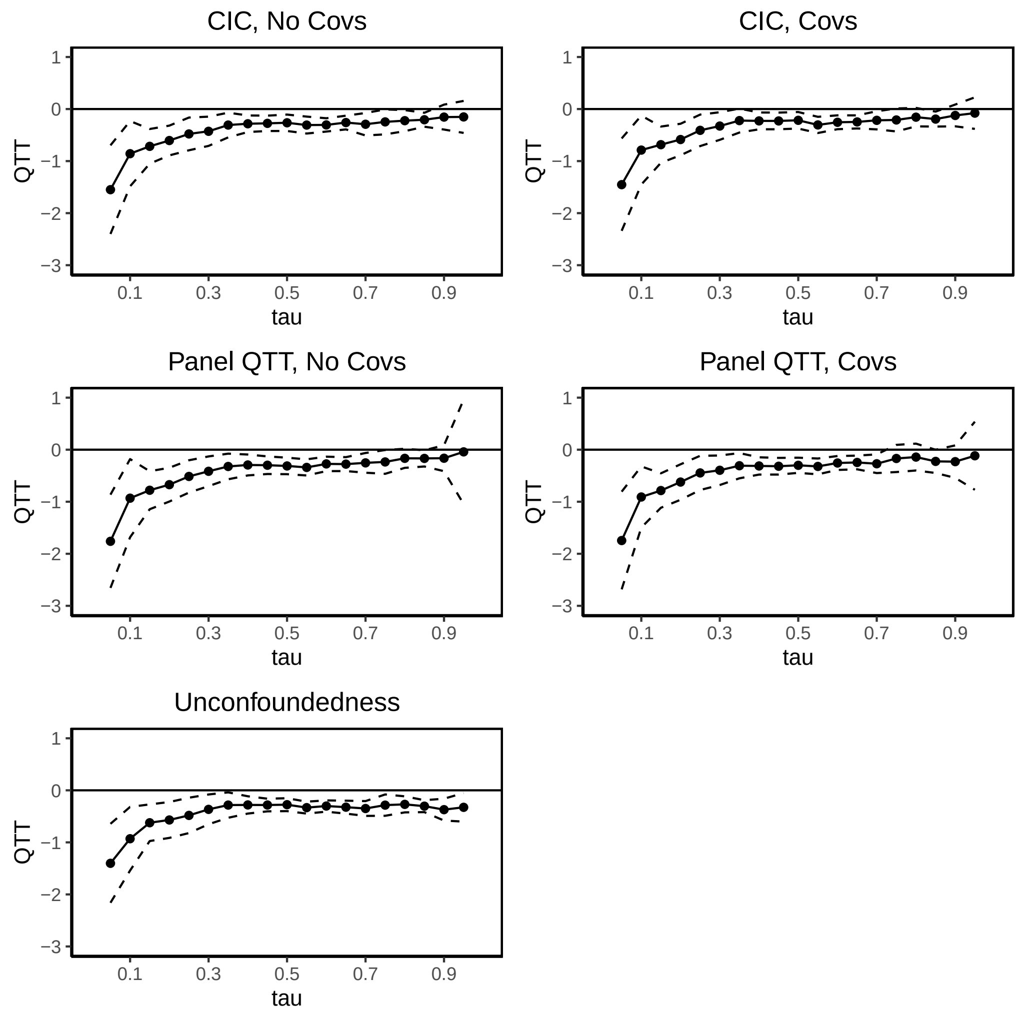
Finally, Figure SG.5 provides estimated QTTs under alternative first step estimators of $F_{Y_{0t}|D=1}$. These are largely similar to each other indicating that the bounds on the main parameters of interest in the paper are not sensitive to the choice of identification argument used for this counterfactual distribution.

Figure SG.4: Bounds on the Quantile of the Treatment Effect using Covariates



Notes: These are bounds that come from using available covariates to tighten the bounds but do not employ 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 in the main text). 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 SG.5: Plots of QTTs using Alternative First Step Assumptions



Notes: The figure plots QTTs using alternative assumptions to identify the counterfactual distribution of non-displaced potential earnings for the group of displaced workers. The panel "CIC, No Covs" provides estimates of the QTT using the Change in Changes model with no covariates; these are the same results as presented in Figure SG.1b. The panel "CIC, Covs" includes covariates in the Change in Changes models using the approach of Melly and Santangelo (2015) that uses first step quantile regression estimators. The panel "Panel QTT, No Covs" uses the Panel QTT method developed in Callaway and Li (2019) without covariates and the "Panel QTT, Covs" uses the same method after adjusting for covariates. The panel "Unconfoundedness" estimates the QTT under the assumption of unconfoundedness using the method developed in Firpo (2007). The last two estimates require a first stage estimation of the propensity score. It is estimated using a logit model and includes dummy variables for less than high school, high school, or college education; Hispanic, black, or white race; and gender. The unconfoundedness results also include the log of earnings in 2007 as an additional control.
Sources: 1979 National Longitudinal Survey of Youth

References

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

[2] Bouyé, Eric and Mark Salmon. "Dynamic copula quantile regressions and tail area dynamic dependence in Forex markets". *The European Journal of Finance* 15.7-8 (2009), pp. 721–750.

[3] Callaway, Brantly and Tong Li. "Quantile treatment effects in difference in differences models with panel data". *Quantitative Economics* 10.4 (2019), pp. 1579–1618.

[4] Callaway, Brantly, 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.

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

[6] Chen, Xiaohong and Yanqin Fan. "Estimation of copula-based semiparametric time series models". *Journal of Econometrics* 130.2 (2006), pp. 307–335.

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

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

[9] Chetty, Raj, Nathaniel Hendren, Patrick Kline, and Emmanuel Saez. "Where is the land of opportunity? The geography of intergenerational mobility in the United States". *The Quarterly Journal of Economics* 129.4 (2014), pp. 1553–1623.

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

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

[12] Duncan, Greg J, 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.

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

[14] Fan, Yanqin 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.

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

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

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

[18] Fermanian, Jean-David, Dragan Radulovic, and Marten Wegkamp. "Weak convergence of empirical copula processes". *Bernoulli* 10.5 (2004), pp. 847–860.

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

[20] Gaenssler, Peter and Winfried Stute. "Copula processes". *Seminar on Empirical Processes*. Springer. 1987, pp. 49–57.

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

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

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

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

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

[26] Kopczuk, Wojciech, Emmanuel Saez, and Jae Song. "Earnings inequality and mobility in the United States: Evidence from social security data since 1937". *The Quarterly Journal of Economics* 125.1 (2010), pp. 91–128.

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

[28] Masten, Matthew A and Alexandre Poirier. "Inference on breakdown frontiers". Working Paper. 2019.

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

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

[31] Rémillard, Bruno and Olivier Scaillet. "Testing for equality between two copulas". *Journal of Multivariate Analysis* 100.3 (2009), pp. 377–386.

[32] Shapiro, Alexander. “On concepts of directional differentiability”. *Journal of Optimization Theory and Applications* 66.3 (1990), pp. 477–487.

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

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

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

[36] 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.

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

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

[39] von Wachter, Till, Jae Song, and Joyce Manchester. “Long-term earnings losses due to mass layoffs during the 1982 recession: An analysis using US administrative data from 1974 to 2004”. *unpublished; New York: Columbia University* (2009).