

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

Brantly Callaway^{†‡}

October 24, 2015
Job Market Paper

Abstract

Older workers who were displaced during the Great Recession lost on average 40% of their earnings relative to their counterfactual earnings had they not been displaced. But the average effect masks substantial heterogeneity across workers. This paper develops new techniques to bound distributional treatment effect parameters that depend on the joint distribution of potential outcomes – an object not identified by standard identifying assumptions such as selection on observables or even when treatment is randomly assigned. I show that panel data and an additional assumption on the dependence between untreated potential outcomes for the treated group over time (i) provide more identifying power for distributional treatment effect parameters than existing bounds and (ii) provide a more plausible set of conditions than existing methods that obtain point identification.

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

JEL Codes: C31, C33, C14

*The latest version can be found at <http://www.brantlycallaway.com/research>

[†]Department of Economics, Vanderbilt University, VU Station B #351819, Nashville, TN 37235-1819. Phone: (678) 832-9884. Email: brantly.m.callaway@vanderbilt.edu

[‡]I would like to thank Alper Arslan, Bill Collins, Federico Gutierrez, Atsushi Inoue, and Pedro Sant'Anna for helpful comments. I especially thank my advisor Tong Li for his detailed comments and mentorship on this project.

1 Introduction

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

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

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

To give an example, suppose a researcher is interested in the fraction of workers who have higher earnings following displacement than they would have had if they not been displaced. Further, suppose hypothetically that workers are randomly assigned to being displaced or not

being displaced. In this case, the average effect of job displacement is identified – it is given by the difference between average earnings of those who are randomly assigned to be displaced and those who are randomly assigned to not be displaced. But the fraction of workers that benefit from displacement is not identified because, for workers randomly assigned to be displaced (non-displaced), where they *would* be in the distribution of non-displaced (displaced) earnings is not known.

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

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

Another approach is to assume that the dependence is known. The leading choice is perfect positive dependence.¹ This assumption says that individuals at a given rank in the distribution of earnings following displacement would have the same rank in the distribution of non-displaced potential earnings. This is a very strong assumption as it imposes severe restrictions on how heterogeneous the effect of treatment can be; for example, it prohibits any workers at the top

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

of the distribution of non-displaced earnings from retiring or taking a part time job following displacement. But the assumption is much stronger than that – it prohibits displacement from even swapping the rank of any workers relative to their rank had they not been displaced.

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

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

Instead of assuming perfect positive dependence of non-displaced potential earnings over time, I assume that the dependence (or copula) of non-displaced potential earnings over time is the same over time. I call this assumption the Copula Stability Assumption. Recent work on income mobility decomposes the income at two different points in time into the marginal distributions – which capture inequality – and the copula which captures income mobility (Chetty, Hendren, Kline, and Saez, 2014). Thus, in the context of job displacement, the Copula Stability Assumption requires that, in the absence of job displacement, earnings mobility would be constant over time for the group of displaced workers. Importantly, the Copula Stability Assumption does not restrict the distribution of earnings over time. For example, the distribution of earnings can shift to the right over time or the distribution of earnings can become

increasingly unequal over time.

I provide two pieces of evidence in favor of the Copula Stability Assumption. First, I show that the Copula Stability Assumption is likely to be satisfied in a very general model of the type typically estimated in panel data settings. Second, there is empirical evidence in favor of the Copula Stability Assumption. In the United States, despite large increases in inequality, there has been remarkably little change in yearly earnings mobility since the middle of the 20th century (Kopczuk, Saez, and Song, 2010, and also see Figure 1).

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

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

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

positive dependence which results in substantially tighter bounds.

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

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

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

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

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

statistical bounds approach. MTR would imply that earnings for displaced workers cannot be larger than earnings would have been had they not been displaced. This assumption is rejected by the bounds developed in the current paper.

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

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

2 Background

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

2.1 Treatment Effects Setup

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

unobserved. Let Y_t be the observed outcome in period t ; one can then write

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

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

2.1.1 Commonly Estimated Parameters

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

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

and the Average Treatment Effect on the Treated (ATT)

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

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

Quantile Treatment Effect

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

untreated potential outcomes

$$QTE(\tau) = F_{Y_{1t}}^{-1}(\tau) - F_{Y_{0t}}^{-1}(\tau)$$

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

In most panel data cases, only some fraction of individuals are ever treated. In this case, panel data is typically useful for identifying the related parameter, the Quantile Treatment Effect on the Treated (QTET), which is given by

$$QTET(\tau) = F_{Y_{1t}|D_t=1}^{-1}(\tau) - F_{Y_{0t}|D_t=1}^{-1}(\tau)$$

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

2.2 Distributional Parameters of Interest

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

There are many parameters of interest that depend on the joint distribution of treated and

untreated potential outcomes rather than just the marginal distributions. First, the fraction of treated individuals that would benefit from treatment depends on the joint distribution

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

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

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

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

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

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

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

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

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

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

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

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

Average Treatment Effect on the Treated Conditional on Previous Outcome

One important limitation of the QTET and the QoTET is that they can fail to provide information on which types of individuals experience the largest effects of treatment. For

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

example, with panel data one can address whether workers with high earnings or low earnings in the previous period experience larger effects of job displacement. This parameter, which I term the Average Treatment Effect on the Treated Conditional on Previous Outcomes (ATT-CPO)⁴ is given by

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

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

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

This example is less extreme than it appears. If any older workers from the middle or top of the non-displaced potential earnings distribution move to the lower part of the distribution of earnings following displacement – which could happen due to difficulty finding new

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

employment, moving to part time work, or retiring – then the QTET will be very difficult to interpret. However, the QoTET and the ATT-CPO can still be very helpful to understand the distributional effects of displacement.

2.3 The Identification Issue and Existing Solutions

This section explains in greater detail the fundamental reason why the joint distribution of potential outcomes is not point identified except under strong assumptions. First, I assume that both the marginal distribution of treated potential outcomes for the treated group $F_{Y_{1t}|D_t=1}(y_1)$ and the marginal distribution of untreated potential outcomes for the treated group $F_{Y_{0t}|D_t=1}(y_0)$ are identified. The first can be obtained directly from the data; the second is obtained under some identifying assumption which is assumed to be available. Sklar (1959) demonstrates that joint distributions can be written as the copula function of marginal distributions in the following way

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

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

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

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

which implies that

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

which means that for any individual in the treated group with observed outcome Y_{1t} , their

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

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

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

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

Lemma 1. *Under perfect positive dependence between untreated potential outcomes for the treated group over time,*

$$F_{Y_{1t}, Y_{0t}|D_t=1}(y_1, y_0) = F_{Y_{1t}, Y_{0t-1}|D_t=1}(y_1, F_{Y_{0t-1}|D_t=1}^{-1}(F_{Y_{0t}|D_t=1}(y_0)))$$

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

3 Identification

This section provides the main identification results in the paper. It provides bounds for the joint distribution of potential outcomes for the treated group, the distribution of the treatment effect for the treated group (DTET), and the quantile of the treatment effect for the treated

group (QoTET). It also provides point identification results for the Average Treatment Effect on the Treated Conditional on Previous Outcome (ATT-CPO).

The following are the main assumptions used in the paper.

Assumption 1. (*Data*)

- (a) *There are three periods of panel data $\{Y_t, Y_{t-1}, Y_{t-2}, D_t, X\}$*
- (b) *No one is treated before period t*

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

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

Assumption 2. (*Marginal Distributions*)

- (a) *The marginal distributions $F_{Y_{1t}|D_t=1}(y_1)$ and $F_{Y_{0t}|D_t=1}(y_0)$ are identified.*
- (b) *The conditional distributions $F_{Y_{1t}|X, D_t=1}(y_1)$ and $F_{Y_{0t}|X, D_t=1}(y_0)$ are identified.*

The first part of Assumption 2(a) says that the distribution of treated potential outcomes for the treated group is identified. This follows directly from Assumption 1 because treated potential outcomes are observed for the treated group. The second part says that the distribution of untreated potential outcomes for the treated group is identified. This distribution is counterfactual and its identification requires some identifying assumption. But this is precisely the distribution that most work on identifying the QTET with observations over time identifies (examples include Athey and Imbens, 2006; Bonhomme and Sauder, 2011; Chernozhukov, Fernández-Val, Hahn, and Newey, 2013; Callaway and Li, 2015). This counterfactual distribution is also available under the selection on observables assumption (Firpo, 2007) and could

be available under an instrumental variables assumption (Abadie, Angrist, and Imbens, 2002; Chernozhukov and Hansen, 2005; Carneiro and Lee, 2009; Frölich and Melly, 2013). Assumption 2(b) is a stronger assumption that says that the conditional versions of each marginal distribution are identified. These can be useful in the case where the Copula Stability Assumption (below) holds conditional on covariates. Also, when the conditional distributions are available, they can be useful for further tightening the bounds on the joint distribution (Fan, Guerre, and Zhu, 2015). If Assumption 2(b) holds, it implies that the marginal distributions in Assumption 2(a) will also be available; throughout most of the paper, I use Assumption 2(a) for simplicity.

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

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

$$C_{Y_{0t}, Y_{0t-1}|D_t=1}(u, v) = C_{Y_{0t-1}, Y_{0t-2}|D_t=1}(u, v)$$

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

$$C_{Y_{0t}, Y_{0t-1}|X, D_t=1}(u, v|x) = C_{Y_{0t-1}, Y_{0t-2}|X, D_t=1}(u, v|x)$$

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

To better understand the Copula Stability Assumption, it is helpful to consider some examples. As a first example, the Copula Stability Assumption says that if untreated potential outcomes at period $t - 1$ are independent (or perfectly positively dependent) of untreated potential at period $t - 2$, then untreated outcomes at period t will continue to be independent

(or perfectly positively dependent) of untreated outcomes at period $t - 1$. Or, for example, suppose the copula for $(Y_{0t-1}, Y_{0t-2} | D_t = 1)$ is Gaussian with parameter ρ , the Copula Stability Assumption says that the copula for $(Y_{0t}, Y_{0t-1} | D_t = 1)$ is also Gaussian with parameter ρ though the marginal distributions of outcomes can change in unrestricted ways. For example, the distribution of earnings could be increasing over time or could become more unequal over time. Likewise, if the copula is Archimedean, the Copula Stability Assumption says that the generator function does not change over time. For Archimedean copulas with a scalar parameter having a one-to-one mapping to dependence parameters such as Kendall's Tau or Spearman's Rho (examples include common Archimedean copulas such as the Clayton, Frank, and Gumbel copulas), the Copula Stability Assumption says that the dependence parameter is the same over time.

The Conditional Copula Stability Assumption may be more plausible in many applications. It says that the dependence is the same over time conditional on some covariates X . For example, earnings over time may be more strongly positively dependent for older workers than for younger workers. It should be noted, however, that the unconditional Copula Stability Assumption does not preclude covariates affecting outcomes; but it does place some restrictions on how covariates can affect the outcome of interest and especially how the effect of covariates changes over time – this issue is discussed more in Section 3.1.1 below.

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

Lemma 2.

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

where

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

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

Theorem 1.

$$F_{Y_{1t}, Y_{0t} | D_t=1}^L(y_1, y_0) \leq F_{Y_{1t}, Y_{0t} | D_t=1}(y_1, y_0) \leq F_{Y_{1t}, Y_{0t} | D_t=1}^U(y_1, y_0)$$

where

$$\begin{aligned}
F_{Y_{1t}, Y_{0t}|D_t=1}^L(y_1, y_0) &= \mathbb{E}[F_{Y_{1t}, Y_{0t}|Y_{0t-1}=y', D_t=1}^L(y_1, y_0|Y_{0t-1})] \\
F_{Y_{1t}, Y_{0t}|D_t=1}^U(y_1, y_0) &= \mathbb{E}[F_{Y_{1t}, Y_{0t}|Y_{0t-1}=y', D_t=1}^U(y_1, y_0|Y_{0t-1})]
\end{aligned}$$

and these bounds are sharp.

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

Example 2 Spearman’s Rho is the correlation of the ranks of two random variables; i.e. $\rho_S = \text{Corr}(F_1(X_1), F_2(X_2))$. Bounds on Spearman’s Rho can be derived when two out of three joint distributions and all marginal distributions (exactly our case) are known. Because the marginal distributions $F_{Y_{1t}|D_t=1}(Y_{1t})$, $F_{Y_{0t}|D_t=1}(Y_{0t})$, and $F_{Y_{0t-1}|D_t=1}(Y_{0t-1})$ are uniformly distributed, their covariance matrix is given by

$$\text{Cov } F_{Y_{1t}|D_t=1}(Y_{1t}), F_{Y_{0t}|D_t=1}(Y_{0t}), F_{Y_{0t-1}|D_t=1}(Y_{0t-1}) = \begin{pmatrix} 1 & \rho_{12} & \rho_{13} \\ \rho_{12} & 1 & \rho_{23} \\ \rho_{13} & \rho_{23} & 1 \end{pmatrix}$$

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

This results in the condition that

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

The width of the bounds is given by

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

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

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

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

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

Lemma 3. (*Conditional Distribution of the Treatment Effect*)

$$F_{Y_{1t}-Y_{0t}|Y_{0t-1}, D_t=1}^L(\Delta|y') \leq F_{Y_{1t}-Y_{0t}|Y_{0t-1}, D_t=1}(\Delta|y') \leq F_{Y_{1t}-Y_{0t}|Y_{0t-1}, D_t=1}^U(\Delta|y')$$

where

$$F_{Y_{1t}-Y_{0t}|Y_{0t-1},D_t=1}^L(\Delta|y') = \sup_y \max\{F_{Y_{1t}|Y_{0t-1},D_t=1}(y|y') - F_{Y_{0t}|Y_{0t-1},D_t=1}(y - \Delta|y'), 0\}$$

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

Theorem 2. (*Distribution of the Treatment Effect*)

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

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

where $F_{Y_{1t}-Y_{0t}|Y_{0t-1},D_t=1}^L(\Delta|y')$ and $F_{Y_{1t}-Y_{0t}|Y_{0t-1},D_t=1}^U(\Delta|y')$ are given in Lemma 3. These bounds are sharp.

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

Theorem 3. (*Quantile of the Treatment Effect*)

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

where

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

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

and $F_{Y_{1t}-Y_{0t}|D_t=1}^L(\Delta)$ and $F_{Y_{1t}-Y_{0t}|D_t=1}^U(\Delta)$ are given in Theorem 2. These bounds are sharp.

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

Assumption 3. (*Distribution of Untreated Potential Outcomes*)

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

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

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

Theorem 4.

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

3.1 More Evidence on the Copula Stability Assumption

Because the key identifying assumption in the paper, the Copula Stability Assumption, is new to the literature on evaluating the distributional impacts of program participation, this section considers whether or not it is likely to hold in applications. The first contribution of this section is to consider several models of varying generality and discusses whether or not they are consistent with the Copula Stability Assumption. The key requirement for a model to satisfy the Copula Stability Assumption is that the way unobservables affect outcomes cannot change over time. This requirement allows for untreated outcomes to be a nonseparable function of observed covariates, time-varying unobservables, and time invariant unobservables that can be correlated with observed covariates plus a time varying function of observed covariates. Second, this section provides empirical evidence in favor of the Copula Stability Assumption in the case where the outcome is yearly earnings in the United States.

3.1.1 Models that are Consistent with the Copula Stability Assumption

An important question is whether or not the Copula Stability Assumption is likely to hold in the types of models that economists most frequently use. This section shows that the Copula Stability Assumption holds in most of the cases most frequently considered in panel data or Difference in Differences settings such as Abadie (2005). Examples include (i) unobserved heterogeneity distributed differently across treated and control groups and (ii) different time

⁵This condition could be weakened to

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

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

trends in untreated potential outcomes for observations with different observable characteristics. The key restriction is that the effect of unobservables cannot change over time. When the effect of time invariant unobserved heterogeneity changes over time – for example, if the return to unobserved ability is increasing over time – then, the Copula Stability Assumption will not hold; however, in this case, panel data assumptions and Difference in Differences assumptions would also be violated which implies that this is not a unique restriction to the methods considered in the current paper. On the other hand, if the effect of time varying unobservables changes over time, then a Difference in Differences approach to identifying the average effect of participating in the treatment would still be valid, but the Copula Stability Assumption would not hold.

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

Model 1: $Y_{0t} = g(X_i, c_i, v_{it})$. This model is stationary in the sense that the same inputs produce the same outcomes in every time period though it is similar to models in recent work on identifying nonseparable models with panel data (for example Evdokimov, 2010; Chernozhukov, Fernández-Val, Hahn, and Newey, 2013). Both the unconditional Copula Stability Assumption and the Conditional Copula Stability Assumption hold in this model. Difference in Differences techniques would be useful for this model. It includes as a special case the model $Y_{0t} = X_i'\beta + c_i + v_{it}$.

Model 2: $Y_{0t} = g(X_i, c_i, v_{it}) + h_t(X_i)$. This model generalizes the previous model in that it allows for a trend in outcomes that can differ based on observable characteristics. The Conditional Copula Stability Assumption holds in this model, but the unconditional Copula Stability Assumption does not. This model includes as a special case $Y_{0t} = X_i'\beta_t + c_i + \theta_t + v_{it}$ where θ_t is an aggregate time fixed effect. However, this model also allows for the possibility of a much more general trend as a function of the observed covariates. The common aggregate time effect is a sufficient condition for the unconditional Copula Stability Assumption to hold even though there are covariates present in the model. In this type of model, one can include time varying observable characteristics – the key requirement is that they be additively separable from unobservables.

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

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

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

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

3.1.2 Empirical Evidence on the Copula Stability Assumption

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

⁶The 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.

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

A simple way to check if the copula is constant over time is to check if some dependence measure such as Spearman’s Rho or Kendall’s Tau is constant over time.⁸ Using administrative data from 1937-2003, Kopczuk, Saez, and Song (2010) find that the rank correlation (Spearman’s Rho) of yearly income is nearly constant in the U.S. Immediately following World War II, there was a slight decline in income mobility. Since then, there has been remarkable stability in income mobility (See Figure 1).

Moreover, Figure 1 also confirms the intuition that there is strong positive dependence of yearly income over time though the dependence is less than perfect positive dependence. This is precisely the case where the method developed in the current paper is likely to (i) provide more credible results than employing a perfect positive dependence over time assumption while (ii) yielding much tighter bounds on the joint distribution of potential outcomes than would be available using other methods that rely on purely statistical results to bound distributional treatment effects that depend on the joint distribution of potential outcomes.

4 Estimation

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

Estimating the QoTET

Estimation of the QoTET is based on the results of Lemma 3, Theorem 2, and Theorem 3. Broadly speaking, it is possible to use plug-in estimators for every term except $F_{Y_{0t}|Y_{0t-1}, D_t=1}(y_0|y')$. This term is identified under the Copula Stability Assumption, but it is not immediate how to estimate it. I consider how to estimate this term in Step 2 below.

Step 1: Estimate $F_{Y_{1t}|Y_{0t-1}, D_t=1}(y_1|y')$:

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

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

⁸It 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.

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

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

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

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

Step 2: Estimate $F_{Y_{0t}|Y_{0t-1}, D_t=1}(y_0|y')$:

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

Lemma 4.

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

With Lemma 4 in hand, it is fairly straightforward to estimate $F_{Y_{0t}|Y_{0t-1}, D_t=1}(y_0|y')$. Once again, I use local linear kernel estimators. Compared to Step 1, the only additional issue here is that I need first step estimators of the distribution and quantile functions in the result of Lemma 4. To estimate distribution functions, I use empirical cdfs:

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

To estimate quantile functions, I invert empirical cdfs:

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

With estimates of the distribution functions in hand, Lemma 4 can be estimated by the solution

$\hat{\alpha}_0(y_0, y')$ to

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

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

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

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

To obtain the bounds on the distribution of the treatment effect, one can plug in the above estimates into the results of Lemma 3 and Theorem 2. Recall, the lower bound on the DTE is identified and given by

$$F_{Y_{1t}-Y_{0t}|D_t=1}^L(\Delta) = E[F_{Y_{1t}-Y_{0t}|Y_{0t-1}, D_t=1}^L(\Delta|Y_{0t-1}) | D_t = 1] \quad (4.1)$$

Further, recall that

$$F_{Y_{1t}-Y_{0t}|Y_{0t-1}, D_t=1}^L(\Delta|y') = \sup_y \max\{F_{Y_{1t}|Y_{0t-1}, D_t=1}(y|y') - F_{Y_{0t}|Y_{0t-1}, D_t=1}(y - \Delta|y'), 0\}$$

which can be estimated by

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

Then, an estimate of Equation 4.1 is given by

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

Similarly, the upper bound on the DTE is given by

$$F_{Y_{1t}-Y_{0t}|D_t=1}^U(\Delta) = E[F_{Y_{1t}-Y_{0t}|Y_{0t-1}, D_t=1}^U(\Delta|Y_{0t-1})|D_t = 1] \quad (4.2)$$

where

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

which can be estimated by

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

and implies that an estimate of Equation 4.2 is given by

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

Step 4: Estimate the Bounds on the QoTET

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

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

and

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

Estimating the ATT-CPO

Recall that the ATT-CPO is given by

$$\begin{aligned} \text{ATT-CPO}(y') &= E[Y_{1t} - Y_{0t}|Y_{0t-1} = y', D_t = 1] \\ &= E[Y_{1t}|Y_{0t-1} = y', D_t = 1] \end{aligned} \quad (4.3)$$

$$- E[F_{Y_{0t}|D_t=1}^{-1}(F_{Y_{0t-1}|D_t=1}(Y_{0t-1})) | Y_{0t-2} = F_{Y_{0t-2}|D_t=1}^{-1}(F_{Y_{0t-1}|D_t=1}(y')), D_t = 1] \quad (4.4)$$

I estimate the ATT-CPO using local linear kernel estimators. An estimate of the term in Equation 4.3 comes from solving

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

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

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

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

Estimating the term in Equation 4.4 is more complicated because it depends on distribution functions and quantile functions. These need to be estimated in a first step. With estimates of the distribution functions in hand, Equation 4.3 can be estimated by the solution $\hat{a}_0(y')$ to

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

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

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

4.1 Inference

ATT-CPO

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

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

Let $g_1(y') = E[Y_{1t}|Y_{0t-1} = y', D_t = 1]$, $\epsilon_{it} = Y_{1it} - g_y(y')$, $\sigma_\epsilon^2(y') = E[\epsilon_{it}^2|Y_{0t-1} = y', D_t = 1]$, and h_1 be a bandwidth parameter. Also, let $z = F_{Y_{0t-2}|D_t=1}^{-1}(F_{Y_{0t-1}|D_t=1}(y'))$, $g(z) = E[F_{Y_{0t}|D_t=1}^{-1}(F_{Y_{0t-1}|D_t=1}(Y_{0t-1})) | Y_{0t-2} = z, D_t = 1]$, $u_{it} = F_{Y_{0t}|D_t=1}^{-1}(F_{Y_{0t-1}|D_t=1}(Y_{0it-1})) - g(F_{Y_{0t-2}|D_t=1}^{-1}(F_{Y_{0t-1}|D_t=1}(Y_{0it-2})))$, $\sigma_u^2(z) = E[u_{it}^2 | Y_{0t-2} = z, D_t = 1]$, and h be a bandwidth parameter. Finally, let $k(\cdot)$ be a kernel function, $\kappa = \int k(v)^2 dv$, and $\kappa_2 = \int k(v)v^2 dv$. I make the following assumptions

Assumption 4.

- (a) $g(z)$, $f_{Y_{0t-2}|D_t=1}(z)$, and $\sigma_u^2(z)$ are twice continuously differentiable
- (b) $g_1(y')$, $f_{Y_{0t-1}|D_t=1}(y')$, and $\sigma_\epsilon^2(y')$ are twice continuously differentiable
- (c) $k(\cdot)$ is a bounded second order kernel
- (d) As $n \rightarrow \infty$, $nh \rightarrow \infty$, $nh_1 \rightarrow \infty$, $nh^7 \rightarrow 0$, and $nh_1^7 \rightarrow 0$
- (e) Y_{0t} , Y_{0t-1} , and Y_{0t-2} have a common, compact support \mathcal{Y}
- (f) $f_{Y_{0t-1}|D_t=1}(\cdot)$ and $f_{Y_{0t-2}|D_t=1}(\cdot)$ are bounded away from 0 on \mathcal{Y} .

Under these assumptions, the following result holds,

Theorem 5. (*Asymptotic Normality of ATT-CPO*)

$$\sqrt{nh}\{(\hat{a}_1(y') - \hat{a}_0(y')) - ATT-CPO(y') - Bias(y')\} \xrightarrow{d} N(0, V) \quad (4.5)$$

where

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

and

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

5 Job Displacement of Older Workers during the Great Recession

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

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

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

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

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

Finally, analyzing the effect of job displacement on older workers is more challenging than

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

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

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

5.1 Data

The data comes from the Health and Retirement Study (HRS) (Health and Retirement Study, 2014). The HRS is a panel data set with interviews occurring every two years. The study follows participants ages 50 and older. Since its inception in 1992, the panel has added

new participants who meet the age requirement six times, most recently adding the Early Baby Boomers (birth years 1948-1953) in 2004 and the Mid Baby Boomers (birth years 1954-1959) in 2010. The primary data source for the current paper is the RAND HRS Version N longitudinal files which I also supplement with some of the base data. I use data from 2004, 2006, 2008, 2010, and 2012 which primarily covers the Early Baby Boomers who range in age from 50-56 in 2004 to 58-64 in 2012. In the context, of job displacement this dataset has been considered by Couch (1998), Chan and Stevens (1999), Chan and Stevens (2001), and Chan and Stevens (2004).

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

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

Summary statistics are presented in Table 1. The outcome of interest is yearly earnings. Prior to the recession, average earnings levels are higher for non-displaced workers than displaced workers (\$48,000 vs. \$43,200). For non-displaced workers, average earnings levels remain essentially flat – in 2012, non-displaced workers earn \$51,400 per year on average. For displaced workers, earnings fall dramatically following displacement. In 2010, average earnings for displaced workers are only \$30,100 (30% lower than 2006 earnings); by 2012, average earnings have increased somewhat to \$35,000 though this is still much lower than pre-displacement earnings.

There are only small differences in observable characteristics that may explain the differences in observed earnings for the displaced and non-displaced groups. 39% of non-displaced workers

have a college degree compared to 33% of displaced workers. 85% of non-displaced workers are white compared to 81% of displaced workers. And, for both groups, 44% are male.

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

5.2 Baseline Results

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

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

$$\log(Y_{it}) = c_i + \gamma_t + \alpha D_{it} + X'_{it}\beta + \epsilon_{it} \quad (5.1)$$

where c_i is an individual-specific fixed effect, γ_t is an aggregate time fixed effect, D_{it} is a binary variable indicating whether or not an individual is displaced, X_{it} is a vector of covariates, and ϵ_{it} is an error term. The coefficient of interest is α . I estimate the model using data on earnings from 2006 and 2012 using a Correlated Random Effects approach.¹⁰

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

¹⁰With two periods, estimates for variables that change over time are numerically identical to estimate from a Fixed Effects approach and also allow me to obtain estimates of the effects of variables that do not change over time though these

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

The last three models in Table 2 consider the younger subset of workers that are 64 or younger in 2012. The estimated effects are very similar for this group. When demographic characteristics (Model 5) and industry and occupation fixed effects are added (Model 6), job displacement is estimated to decrease earnings by 38% and 40%, respectively. Finally, when time varying occupations are included, the estimated effect is somewhat smaller at 28% and only borderline statistically significant (p-value=0.07).

A weakness of the previous specifications is that the aggregate time effect, γ_t , is common to all individuals. This means that the time trend is constrained by the model to be the same for individuals that may have very different observed and unobserved characteristics. This could potentially cause the effect of job displacement to be overestimated. For example, less educated workers may be more likely to be displaced than highly educated workers. Earnings are also likely to be increasing more over time for highly educated workers than for less educated workers. Since the trend in earnings for highly educated workers will be used to pin down γ_t in these specifications, the trend is likely to be overestimated for less educated workers; therefore, the size of the effect of job displacement would also be overestimated in this situation.

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

Instead of interacting observables and time fixed effects, I consider a nonparametric identifiy-

estimates should be interpreted more cautiously. The original work on correlated random effects models is Mundlak (1978), Chamberlain (1982), and Chamberlain (1984)

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

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

ing assumption consistent with the idea that individuals that differ in observable characteristics may have differing time trends in untreated outcomes

Assumption 5. (*Distributional Difference in Differences*)

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

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

$$\begin{aligned} Y_{1t} &= g_{1t}(X, c_i, \epsilon_{it}) \\ Y_{0t} &= g_{0t}(X) + h_0(X, c_i, \epsilon_{it}) \end{aligned}$$

where the model for treated potential outcomes is left completely unrestricted and the model for untreated potential outcomes depends on observable characteristics X , time invariant unobservables c_i that may be correlated with the covariates X , and time varying unobservables that are independent of (X, c_i) but may be serially correlated over time. The restriction on the model for untreated potential outcomes is that the path of untreated outcomes depends only on observable characteristics X but not on any unobservables. Assumption 5 identifies only the ATT and QTET but not the ATE or the QTE. Under this assumption,

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

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

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

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

¹⁴Couch and Placzek (2010) also find very similar results whether using a parametric model which is very similar to the

5.3 The Distributional Effects of Job Displacement

This section uses the techniques developed earlier in the paper to understand the distributional effects of job displacement for older workers. I focus on estimating the QoTET and the ATT-CPO. The QoTET is useful for understanding heterogeneity in the effect of job displacement. It is partially identified in the current application. The ATT-CPO is point identified and is useful for determining whether job displacement has had a relatively larger impact on workers who previously had high or low earnings.

There are three main findings using the methods developed in the current paper. First, the bounds developed in the current paper provide substantially more identifying power than bounds relying on only knowledge of the marginal distributions of potential outcomes. Those bounds are essentially uninformative in the current application. Second, the tighter bounds in the current paper rule out: (i) the assumption of perfect positive dependence across treated and untreated potential outcomes and (ii) the assumption of Monotone Treatment Response (Manski, 1997). Finally, estimates of the ATT-CPO (See Figure 5) imply that workers that had higher earnings in the previous period experience larger earnings losses due to job displacement than workers with lower earnings in the previous period. But, as a fraction of earnings, the effect of job displacement is very similar for workers across all initial earnings levels.

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

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

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

current parametric model or a very similar semiparametric Difference in Differences technique.

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

The key identifying assumption of Callaway and Li (2015) is Assumption 5. That paper also imposes a Copula Stability Assumption that is similar to the one in the current paper though not exactly the same. However, that Copula Stability Assumption is also satisfied in the same types of models that satisfy the Copula Stability Assumption in the current paper which suggests that this extra condition is not likely to add much empirical content. The counterfactual distribution of potential outcomes is point identified in this setup. Estimates of both marginal distributions are presented in Figure 2, and an estimate of the QTET is presented in Figure 3.

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

Next, I use the techniques presented earlier in the paper to estimate some parameters that depend on the joint distribution of treated and untreated potential outcomes. First, I consider the QoTET. Figure 4 plots (i) bounds on the QoTET under no assumptions on the dependence between potential outcome distributions, (ii) the bounds developed in the current paper, (iii) point estimates of the QoTET under the assumption of perfect positive dependence between treated and untreated potential outcomes, and (iv) point estimates of the QoTET under the assumption of Rank Invariance between untreated potential outcomes over time. There are several things to notice from the figure. First, when no information besides the identified marginal distributions is used, the bounds on the QoTET are very wide. For example, the median of the treatment effect is bounded to be between an earnings losses of 79% and an earnings gain of 118%. The effect of displacement among those most affected by displacement, for example the 5th percentile of the QoTET is estimated to be between a 62% and 99% loss of earnings. The effect of displacement on those least affected by displacement, for example the 95th percentile of the QoTET is bounded between a 21% loss of earnings and a 1163% increase in earnings. From these bounds, one is not able to determine much. These bounds indicate that at least 19% of displaced workers have lower earnings than they would have had they not been displaced.

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

Figure 4 also plots the QoTET under several assumptions that would lead to point identification. First, it plots the QoTET under perfect positive dependence between Y_{1t} and Y_{0t} . I have argued that this is an especially strong assumption in this case. For example, it essentially restricts any previously high earnings individuals from moving into much lower paying positions following displacement. This identifying assumption implies the least amount of heterogeneity in the effect of being displaced. At the 5th percentile, individuals lose 82% from being displaced. At the 95th percentile, they lose 16%. At the median, they lose 32%, and this effect is largely constant across most of the interior quantiles. Of course, the no-assumptions bounds cannot rule out perfect positive dependence between Y_{1t} and Y_{0t} , but under the Copula Stability Assumption, perfect positive dependence is ruled out because the bounds imply more heterogeneity than occurs under perfect positive dependence.

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

6 Conclusion

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

Using these methods, I have studied the distributional effects of job displacement during the Great Recession for older workers. Using standard techniques, I find that older workers

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

References

- [1] Abadie, Alberto. “Semiparametric Difference-in-Differences Estimators”. *The Review of Economic Studies* 72.1 (2005), pp. 1–19.
- [2] Abadie, Alberto, Joshua Angrist, and Guido Imbens. “Instrumental variables estimates of the effect of subsidized training on the quantiles of trainee earnings”. *Econometrica* 70.1 (2002), pp. 91–117.
- [3] Abbring, Jaap and James Heckman. “Econometric Evaluation of Social Programs, Part III: Distributional Treatment Effects, Dynamic Treatment Effects, Dynamic Discrete Choice, and General Equilibrium Policy Evaluation”. *Handbook of Econometrics* 6 (2007), pp. 5145–5303.
- [4] Athey, Susan and Guido Imbens. “Identification and Inference in Nonlinear Difference-in-Differences Models”. *Econometrica* 74.2 (2006), pp. 431–497.
- [5] Bitler, Marianne P, Jonah B Gelbach, and Hilary W Hoynes. “Welfare Reform and Children’s Living Arrangements”. *Journal of Human Resources* 41.1 (2006), pp. 1–27.
- [6] Bonhomme, Stéphane and Ulrich Sauder. “Recovering Distributions in Difference-in-Differences models: A Comparison of Selective and Comprehensive Schooling”. *Review of Economics and Statistics* 93.2 (2011), pp. 479–494.
- [7] Callaway, Brantly and Tong Li. “Quantile Treatment Effects in Difference in Differences Models with Panel Data” (2015).
- [8] Carneiro, Pedro, Karsten Hansen, and James Heckman. “Removing the Veil of Ignorance in assessing the distributional impacts of social policies”. *Swedish Economic Policy Review* 8 (2001), pp. 273–301.
- [9] Carneiro, Pedro, Karsten Hansen, and James Heckman. “Estimating distributions of treatment effects with an application to the returns to schooling and measurement of the effects of uncertainty on college choice”. *International Economic Review* 44.2 (2003), pp. 361–422.
- [10] Carneiro, Pedro and Sokbae Lee. “Estimating distributions of potential outcomes using local instrumental variables with an application to changes in college enrollment and wage inequality”. *Journal of Econometrics* 149.2 (2009), pp. 191–208.
- [11] 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.

- [12] Chamberlain, Gary. “Multivariate Regression Models for Panel Data”. *Journal of Econometrics* 18.1 (1982), pp. 5–46.
- [13] Chamberlain, Gary. “Panel Data”. *Handbook of Econometrics* 2 (1984), pp. 1248–1318.
- [14] Chan, Sewin and Ann Huff Stevens. “Employment and retirement following a late-career job loss”. *American Economic Review* (1999), pp. 211–216.
- [15] Chan, Sewin and Ann Huff Stevens. “The Effects of Job Loss on Older Workers: Employment, Earnings and Wealth”. *Ensuring Health and Income Security for an Aging Workforce* (2001), pp. 189–211.
- [16] Chan, Sewin and Ann Huff Stevens. “How Does Job Loss Affect the Timing of Retirement?” *The BE Journal of Economic Analysis & Policy* 3.1 (2004), pp. 1–26.
- [17] Chernozhukov, Victor, Iván Fernández-Val, Jinyong Hahn, and Whitney Newey. “Average and quantile effects in nonseparable panel models”. *Econometrica* 81.2 (2013), pp. 535–580.
- [18] Chernozhukov, Victor and Christian Hansen. “An IV model of quantile treatment effects”. *Econometrica* 73.1 (2005), pp. 245–261.
- [19] 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.
- [20] Couch, Kenneth A. “Late life job displacement”. *The Gerontologist* 38.1 (1998), pp. 7–17.
- [21] Couch, Kenneth A and Dana W Placzek. “Earnings losses of displaced workers revisited”. *The American Economic Review* (2010), pp. 572–589.
- [22] 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.
- [23] Djebbari, Habiba and Jeffrey Smith. “Heterogeneous Impacts in PROGRESA”. *Journal of Econometrics* 145.1 (2008), pp. 64–80.
- [24] Doksum, Kjell. “Empirical probability plots and statistical inference for nonlinear models in the two-sample case”. *The annals of statistics* (1974), pp. 267–277.
- [25] Dong, Yingying and Shu Shen. “Testing for Rank Invariance or Similarity in Program Evaluation: The Effect of Training on Earnings Revisited”. Working Paper. 2015.
- [26] Evdokimov, Kirill. “Identification and estimation of a nonparametric panel data model with unobserved heterogeneity”. Working Paper, Princeton University. 2010.

- [27] Fan, Jianqing and Irene Gijbels. *Local Polynomial Modelling and its Applications: Monographs on Statistics and Applied Probability 66*. Vol. 66. CRC Press, 1996.
- [28] Fan, Yanqin, Emmanuel Guerre, and Dongming Zhu. “Partial Identification and Confidence Sets for Functionals of the Joint Distribution of Potential Outcomes” (2015).
- [29] Fan, Yanqin and Sang Soo Park. “Partial identification of the distribution of treatment effects and its confidence sets”. *Advances in Econometrics* 25 (2009), pp. 3–70.
- [30] 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.
- [31] Fan, Yanqin and Sang Soo Park. “Confidence intervals for the quantile of treatment effects in randomized experiments”. *Journal of Econometrics* 167.2 (2012), pp. 330–344.
- [32] 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.
- [33] Farber, Henry S. “Job Loss in the Great Recession and its Aftermath: U.S. Evidence from the Displaced Workers Survey”. Working Paper. 2015.
- [34] Firpo, Sergio. “Efficient Semiparametric Estimation of Quantile Treatment Effects”. *Econometrica* 75.1 (2007), pp. 259–276.
- [35] Firpo, Sergio and Geert Ridder. “Bounds on functionals of the distribution of treatment effects” (2008).
- [36] Frandsen, Brigham R and Lars J Lefgren. “A Test for Rank Similarity and Partial Identification of the Distribution of Treatment Effects”. Working Paper. 2015.
- [37] Frank, Maurice, Roger Nelsen, and Berthold Schweizer. “Best-possible bounds for the distribution of a sum-a problem of Kolmogorov”. *Probability Theory and Related Fields* 74.2 (1987), pp. 199–211.
- [38] Fréchet, Maurice. “Sur Les Tableaux de Correlation Dont Les Marges Sont Donnees”. *Annals University Lyon: Series A* 14 (1951), pp. 53–77.
- [39] Frölich, Markus and Blaise Melly. “Unconditional quantile treatment effects under endogeneity”. *Journal of Business & Economic Statistics* 31.3 (2013), pp. 346–357.
- [40] Gibbons, Robert and Lawrence F Katz. “Layoffs and Lemons”. *Journal of Labor Economics* 9 (1991), pp. 351–380.
- [41] Gottschalk, Peter. “Inequality, Income Growth, and Mobility: The Basic Facts”. *The Journal of Economic Perspectives* (1997), pp. 21–40.

- [42] Hall, Peter, Rodney CL Wolff, and Qiwei Yao. “Methods for estimating a conditional distribution function”. *Journal of the American Statistical Association* 94.445 (1999), pp. 154–163.
- [43] Hansen, Bruce E. “Nonparametric estimation of smooth conditional distributions”. Working Paper: Department of Economics, University of Wisconsin. 2004.
- [44] Health and Retirement Study. *RAND HRS Data file public use dataset*. Produced and distributed by the University of Michigan with funding from the National Institute on Aging (grant number NIA U01AG009740). Ann Arbor, MI, 2014.
- [45] Heckman, James and V Joseph Hotz. “Choosing among alternative nonexperimental methods for estimating the impact of social programs: The case of manpower training”. *Journal of the American statistical Association* 84.408 (1989), pp. 862–874.
- [46] Heckman, James, Hidehiko Ichimura, and Petra Todd. “Matching as an econometric evaluation estimator: Evidence from evaluating a job training programme”. *The review of economic studies* 64.4 (1997), pp. 605–654.
- [47] Heckman, James and Jeffrey Smith. “Evaluating the Welfare State James J. Heckman and Jeffrey Smith”. *Econometrics and Economic Theory in the 20th Century: The Ragnar Frisch Centennial Symposium*. 31. Cambridge University Press. 1998, pp. 241–318.
- [48] Heckman, James and Jeffrey Smith. “The Pre-programme Earnings Dip and the Determinants of Participation in a Social Programme. Implications for Simple Programme Evaluation Strategies”. *The Economic Journal* 109.457 (1999), pp. 313–348.
- [49] Heckman, James, Jeffrey Smith, and Nancy Clements. “Making the Most out of Programme Evaluations and Social Experiments: Accounting for Heterogeneity in Programme Impacts”. *The Review of Economic Studies* 64.4 (1997), pp. 487–535.
- [50] Hoeffding, Wassily. “Massstabinvariante korrelationstheorie”. *Schriften des Mathematischen Instituts und des Institutes Fur Angewandte Mathematik der Universitat Berlin* 5 (1940), pp. 179–233.
- [51] Hungerford, Thomas L. “US Income Mobility in the Seventies and Eighties”. *Review of Income and Wealth* 39.4 (1993), pp. 403–417.
- [52] Jacobson, Louis S, Robert J LaLonde, and Daniel G Sullivan. “Earnings losses of displaced workers”. *The American Economic Review* (1993), pp. 685–709.
- [53] Joe, Harry. *Multivariate models and multivariate dependence concepts*. CRC Press, 1997.
- [54] Joe, Harry. *Dependence modeling with copulas. No. 134 in monographs on statistics and applied probability*. CRC Press, Boca Raton, FL, 2015.

- [55] Kim, Ju Hyun. “Identifying the Distribution of Treatment Effects under Support Restrictions”. Working Paper. 2014.
- [56] 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.
- [57] Lehmann, Erich. *Nonparametrics: Statistical Methods Based on Ranks*. 1974.
- [58] Li, Qi and Jeffrey Scott Racine. *Nonparametric Econometrics: Theory and Practice*. Princeton University Press, 2007.
- [59] Makarov, GD. “Estimates for the Distribution Function of a Sum of Two Random Variables When the Marginal Distributions are Fixed”. *Theory of Probability & its Applications* 26.4 (1982), pp. 803–806.
- [60] Manski, Charles. “Monotone Treatment Response”. *Econometrica* (1997), pp. 1311–1334.
- [61] Melly, Blaise and Giulia Santangelo. “The changes-in-changes model with covariates”. Working Paper. 2015.
- [62] Mundlak, Yair. “On the Pooling of Time Series and Cross Section Data”. *Econometrica: Journal of the Econometric Society* (1978), pp. 69–85.
- [63] Nelsen, Roger. *An introduction to copulas*. Springer, 2007.
- [64] Rüschendorf, Ludger. “Random variables with maximum sums”. *Advances in Applied Probability* (1982), pp. 623–632.
- [65] Sen, Amartya. *On economic inequality*. Clarendon Press, 1997.
- [66] Sklar, Abe. *Fonctions de répartition à n dimensions et leurs marges*. Publications de L Institut de Statistique de L Universite de Paris, 1959.
- [67] Stevens, Ann Huff. “Persistent Effects of Job Displacement: The Importance of Multiple Job Losses”. *Journal of Labor Economics* (1997), pp. 165–188.
- [68] U.S. Bureau of Labor Statistics. *All Employees: Total Nonfarm Payrolls [PAYEMS]*. Retrieved from FRED, Federal Reserve Bank of St. Louis on October 6, 2015. 2015.
- [69] U.S. Bureau of Labor Statistics. *Civilian Unemployment Rate [UNRATE]*. Retrieved from FRED, Federal Reserve Bank of St. Louis on October 6, 2015. 2015.
- [70] U.S. Bureau of Labor Statistics. *Unemployment Rate: 55 years and over [LNS14024230]*. Retrieved from FRED, Federal Reserve Bank of St. Louis on October 6, 2015. 2015.

- [71] 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).
- [72] Williamson, Robert and Tom Downs. “Probabilistic Arithmetic. I. Numerical Methods for Calculating Convolutions and Dependency Bounds”. *International Journal of Approximate Reasoning* 4 (1990), pp. 89–158.

A Proofs

A.1 Proof of Lemma 1

Proof.

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

where the third equality uses the assumption of perfect positive dependence. \square

A.2 Proofs of Lemma 2, Lemma 3, Theorem 1, Theorem 2, and Theorem 3

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

Lemma 3 applies the sharp bounds on the difference between random variables with known marginal distributions but unknown copula of Williamson and Downs (1990) to the difference conditional on the previous outcome.

Theorem 1 and Theorem 2 follow from results in Fan and Park (2010, Section 5) and Fan, Guerre, and Zhu (2015) which derive sharp bounds on the unconditional distribution of the treatment effect when conditional marginal distributions are known. In those cases, the marginal distributions are conditional on observed covariates X ; in the current paper, the marginal distributions are conditional on a lag of the outcome Y_{0t-1} .

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

A.3 Proof of Theorem 4

Proof. To show the result in Theorem 4, it must be shown that $E[Y_{0t}|Y_{0t-1} = y', D_t = 1] = E[F_{Y_{0t}|D_t=1}^{-1}(F_{Y_{0t-1}|D_t=1}(Y_{0t-1})) \mid Y_{0t-2} = F_{Y_{0t-2}|D_t=1}^{-1}(F_{Y_{0t-1}|D_t=1}(y'))]$.

$$E[Y_{0t}|Y_{0t-1} = y', D_t = 1]$$

$$\begin{aligned}
&= \int_{\mathcal{Y}} y_{0t} f_{Y_{0t}|Y_{0t-1}, D_t=1}(y_{0t} | y') \, dy_{0t} \\
&= \int_{\mathcal{Y}} y_{0t} \frac{f_{Y_{0t}, Y_{0t-1}, D_t=1}(y_{0t}, y')}{f_{Y_{0t-1}|D_t=1}(y')} \, dy_{0t} \\
&= \int_{\mathcal{Y}} y_{0t} c_{Y_{0t}, Y_{0t-1}|D_t=1}(F_{Y_{0t}|D_t=1}(y_{0t}), F_{Y_{0t-1}|D_t=1}(y')) f_{Y_{0t}|D_t=1}(y_{0t}) \, dy_{0t} \\
&= \int_{\mathcal{Y}} y_{0t} c_{Y_{0t-1}, Y_{0t-2}|D_t=1}(F_{Y_{0t}|D_t=1}(y_{0t}), F_{Y_{0t-1}|D_t=1}(y')) f_{Y_{0t}|D_t=1}(y_{0t}) \, dy_{0t} \\
&= \int_{\mathcal{Y}} y_{0t} f_{Y_{0t-1}, Y_{0t-2}|D_t=1}(F_{Y_{0t-1}|D_t=1}^{-1}(F_{Y_{0t}|D_t=1}(y_{0t})), F_{Y_{0t-2}|D_t=1}^{-1}(F_{Y_{0t-1}|D_t=1}(y'))) \\
&\quad \times \frac{f_{Y_{0t}|D_t=1}(y_{0t})}{f_{Y_{0t-1}|D_t=1}(F_{Y_{0t-1}|D_t=1}^{-1}(F_{Y_{0t}|D_t=1}(y_{0t}))) \times f_{Y_{0t-2}|D_t=1}(F_{Y_{0t-2}|D_t=1}^{-1}(F_{Y_{0t-1}|D_t=1}(y')))} \, dy_{0t} \\
&= \int_{\mathcal{Y}} y_{0t} f_{Y_{0t-1}|Y_{0t-2}, D_t=1}(F_{Y_{0t-1}|D_t=1}^{-1}(F_{Y_{0t}|D_t=1}(y_{0t})) | F_{Y_{0t-2}|D_t=1}^{-1}(F_{Y_{0t-1}|D_t=1}(y'))) \\
&\quad \times \frac{f_{Y_{0t}|D_t=1}(y_{0t})}{f_{Y_{0t-1}|D_t=1}(F_{Y_{0t-1}|D_t=1}^{-1}(F_{Y_{0t}|D_t=1}(y_{0t})))} \, dy_{0t}
\end{aligned}$$

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

$$y_{0t} = F_{Y_{0t-1}|D_t=1}^{-1}(F_{Y_{0t-1}|D_t=1}(u))$$

and

$$dy_{0t} = \frac{f_{Y_{0t-1}|D_t=1}(u)}{f_{Y_{0t}|D_t=1}(F_{Y_{0t}|D_t=1}^{-1}(F_{Y_{0t-1}|D_t=1}(u)))}$$

Plugging these back in implies

$$\begin{aligned}
&E[Y_{0t}|Y_{0t-1} = y', D_t = 1] \\
&= \int_{\mathcal{Y}} F_{Y_{0t}|D_t=1}^{-1}(F_{Y_{0t-1}|D_t=1}(u)) f_{Y_{0t-1}|Y_{0t-2}, D_t=1}(u | F_{Y_{0t-2}|D_t=1}^{-1}(F_{Y_{0t-1}|D_t=1}(y'))) \, du \\
&= E[F_{Y_{0t}|D_t=1}^{-1}(F_{Y_{0t-1}|D_t=1}(Y_{0t-1})) \mid Y_{0t-2} = F_{Y_{0t-2}|D_t=1}^{-1}(F_{Y_{0t-1}|D_t=1}(y'))]
\end{aligned}$$

□

A.4 Proofs of Models Satisfying the Conditional Copula Stability Assumption

Recall, the conditional copula is given by

$$C_{Y_{0t}, Y_{0t-1}|X}(u, v|x) = P(F_{Y_{0t}|X}(Y_{0t}|x) \leq u, F_{Y_{0t-1}|X}(Y_{0t-1}|x) \leq v|X = x)$$

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

Model 1 : $Y_{0t} = g(X_i, c_i, v_{it})$. In this case,

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

which implies

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

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

Model 2: $Y_{0t} = g(X_i, c_i, v_{it}) + h_t(X_i)$. In this case,

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

which implies

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

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

Model 3: $Y_{0t} = g(X_i, c_i, v_{it}) + h_t(X_i, v_{it})$. In this case,

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

which implies

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

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

Model 4: $Y_{0t} = g(X_i, c_i, v_{it}) + h_t(X_i, c_i)$. In this case,

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

which implies

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

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

A.5 Proof of Lemma 4

The proof of Lemma 4 is very similar to the proof of Theorem 4 and is omitted.

A.6 Proof of Theorem 5

Before proving the main result, I state the following lemmas

Lemma 5. (*Uniform convergence of empirical distribution function*)

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

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

Lemma 6. (*Uniform convergence of empirical quantiles*)

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

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

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

Step 1: (Accounting for 1st Step Estimation)

The first step is to show that estimating the first step estimates of unconditional distribution and quantile functions do not affect the asymptotic distribution of the ATT-CPO. As a first step, rewrite Equation 4.5 as

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

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

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

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

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

$$\begin{aligned}
& \times \frac{1}{n_T} K_h(Y_{it-2} - F_{Y_{0t-2}|D_t=1}^{-1}(F_{Y_{0t-1}|D_t=1}(y'))) \\
& \times \frac{1}{n_T h} \frac{1}{Y_{it-2} - F_{Y_{0t-2}|D_t=1}^{-1}(F_{Y_{0t-1}|D_t=1}(y'))} F_{Y_{0t}|D_t=1}^{-1}(F_{Y_{0t-1}|D_t=1}(Y_{it-1})) \\
= & \frac{1}{n_T} \frac{1}{i \in \mathcal{T}} K_h(Y_{it-2} - F_{Y_{0t-2}|D_t=1}^{-1}(F_{Y_{0t-1}|D_t=1}(y'))) \\
& \times \frac{1}{Y_{it-2} - F_{Y_{0t-2}|D_t=1}^{-1}(F_{Y_{0t-1}|D_t=1}(y'))} F_{Y_{0t}|D_t=1}^{-1}(F_{Y_{0t-1}|D_t=1}(Y_{it-1})) \\
& \times \frac{1}{n_T} \frac{1}{i \in \mathcal{T}} K_h(Y_{it-2} - F_{Y_{0t-2}|D_t=1}^{-1}(F_{Y_{0t-1}|D_t=1}(y'))) \\
& \times \left(\begin{aligned} & \sqrt{n_T h} \hat{F}_{Y_{0t}|D_t=1}^{-1}(\hat{F}_{Y_{0t-1}|D_t=1}(Y_{it-1})) - F_{Y_{0t}|D_t=1}^{-1}(F_{Y_{0t-1}|D_t=1}(Y_{it-1})) \\ & \sqrt{n_T h} \hat{X}_i(y') \times \hat{F}_{Y_{0t}|D_t=1}^{-1}(\hat{F}_{Y_{0t-1}|D_t=1}(Y_{it-1})) - X_i(y') \times F_{Y_{0t}|D_t=1}^{-1}(F_{Y_{0t-1}|D_t=1}(Y_{it-1})) \end{aligned} \right) \\
& + o_p(1)
\end{aligned}$$

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

$$\overline{n_T h} \hat{F}_{Y_{0t}|D_t=1}^{-1}(\hat{F}_{Y_{0t-1}|D_t=1}(Y_{it-1})) - F_{Y_{0t}|D_t=1}^{-1}(F_{Y_{0t-1}|D_t=1}(Y_{it-1})) \quad (\text{A.1})$$

$$\leq \sup_z \overline{n_T h} \hat{F}_{Y_{0t}|D_t=1}^{-1}(\hat{F}_{Y_{0t-1}|D_t=1}(z)) - F_{Y_{0t}|D_t=1}^{-1}(F_{Y_{0t-1}|D_t=1}(z))$$

$$\leq \sup_z \overline{n_T h} \hat{F}_{Y_{0t}|D_t=1}^{-1}(\hat{F}_{Y_{0t-1}|D_t=1}(z)) - F_{Y_{0t}|D_t=1}^{-1}(\hat{F}_{Y_{0t-1}|D_t=1}(z)) \quad (\text{A.2})$$

$$+ \sup_z \overline{n_T h} F_{Y_{0t}|D_t=1}^{-1}(\hat{F}_{Y_{0t-1}|D_t=1}(z)) - F_{Y_{0t}|D_t=1}^{-1}(F_{Y_{0t-1}|D_t=1}(z)) \quad (\text{A.3})$$

$$\xrightarrow{p} 0$$

where the result follows since

$$\begin{aligned}
& \sup_z \overline{n_T h} \hat{F}_{Y_{0t}|D_t=1}^{-1}(\hat{F}_{Y_{0t-1}|D_t=1}(z)) - F_{Y_{0t}|D_t=1}^{-1}(\hat{F}_{Y_{0t-1}|D_t=1}(z)) \\
& = \sup_{q \in [0,1]} \overline{n_T h} \hat{F}_{Y_{0t}|D_t=1}^{-1}(q) - F_{Y_{0t}|D_t=1}^{-1}(q)
\end{aligned}$$

which converges to 0 by Lemma 6. And, by a Taylor Expansion,

$$\begin{aligned} & \sup_z \overline{n_T h} F_{Y_{0t}|D_t=1}^{-1}(\hat{F}_{Y_{0t-1}|D_t=1}(z)) - F_{Y_{0t}|D_t=1}^{-1}(F_{Y_{0t-1}|D_t=1}(z)) \\ &= \sup_z \frac{1}{f_{Y_{0t}|D_t=1}(F_{Y_{0t}|D_t=1}^{-1}(F_{Y_{0t-1}|D_t=1}(z)))} \times \overline{n_T h} \hat{F}_{Y_{0t-1}|D_t=1}(z) - F_{Y_{0t-1}|D_t=1}(z) \end{aligned}$$

which converges to 0 because $f_{Y_{0t}|D_t=1}(\cdot)$ is bounded away from 0 on its support and the term $\sqrt{n_T h} \hat{F}_{Y_{0t-1}|D_t=1}(z) - F_{Y_{0t-1}|D_t=1}(z)$ converges to 0 by Lemma 5.

Next, I show that $\sqrt{n_T h} \hat{X}_i(y') \times \hat{F}_{Y_{0t}|D_t=1}^{-1}(\hat{F}_{Y_{0t-1}|D_t=1}(Y_{it-1})) - X_i(y') \times F_{Y_{0t}|D_t=1}^{-1}(F_{Y_{0t-1}|D_t=1}(Y_{it-1}))$ converges in probability to 0.

$$\begin{aligned} & \overline{n_T h} \hat{X}_i(y') \times \hat{F}_{Y_{0t}|D_t=1}^{-1}(\hat{F}_{Y_{0t-1}|D_t=1}(Y_{it-1})) - X_i(y') \times F_{Y_{0t}|D_t=1}^{-1}(F_{Y_{0t-1}|D_t=1}(Y_{it-1})) \\ & \leq \sup_z \overline{n_T h} \hat{F}_{Y_{0t-2}|D_t=1}^{-1}(\hat{F}_{Y_{0t-1}|D_t=1}(y')) - F_{Y_{0t-2}|D_t=1}^{-1}(F_{Y_{0t-1}|D_t=1}(y')) \times \hat{F}_{Y_{0t}|D_t=1}^{-1}(\hat{F}_{Y_{0t-1}|D_t=1}(z)) \end{aligned} \tag{A.4}$$

$$+ \sup_z \overline{n_T h} F_{Y_{0t-2}|D_t=1}^{-1}(F_{Y_{0t-1}|D_t=1}(y')) \times \hat{F}_{Y_{0t}|D_t=1}^{-1}(\hat{F}_{Y_{0t-1}|D_t=1}(z)) - F_{Y_{0t}|D_t=1}^{-1}(F_{Y_{0t-1}|D_t=1}(z)) \tag{A.5}$$

where the second line holds by adding and subtracting $z - F_{Y_{0t-2}|D_t=1}^{-1}(F_{Y_{0t-1}|D_t=1}(y')) \times \hat{F}_{Y_{0t}|D_t=1}^{-1}(\hat{F}_{Y_{0t-1}|D_t=1}(z))$ and by the triangle inequality. Equation A.4 converges to 0 because $\hat{F}_{Y_{0t}|D_t=1}^{-1}(\hat{F}_{Y_{0t-1}|D_t=1}(z))$ is bounded for all z and one can show that $\hat{F}_{Y_{0t-2}|D_t=1}^{-1}(\hat{F}_{Y_{0t-1}|D_t=1}(y')) - F_{Y_{0t-2}|D_t=1}^{-1}(F_{Y_{0t-1}|D_t=1}(y'))$ converges to 0 by exactly the same sort of argument as for Equation A.1.

Likewise, Equation A.5 converges to 0 $F_{Y_{0t-2}|D_t=1}^{-1}(F_{Y_{0t-1}|D_t=1}(y'))$ is bounded and

$$\sup_z \overline{n_T h} \hat{F}_{Y_{0t}|D_t=1}^{-1}(\hat{F}_{Y_{0t-1}|D_t=1}(z)) - F_{Y_{0t}|D_t=1}^{-1}(F_{Y_{0t-1}|D_t=1}(z))$$

converges to 0; this is exactly the result coming from Equation A.1.

Step 2: (Asymptotically Linear Representations)

Let $W_i = F_{Y_{0t}|D_t=1}^{-1}(F_{Y_{0t-1}|D_t=1}(Y_{it-1}))$ and $z = F_{Y_{0t-2}|D_t=1}^{-1}(F_{Y_{0t-1}|D_t=1}(y'))$. Then, standard results on estimating conditional distributions (see, for example, Fan and Gijbels, 1996; Li and

Racine, 2007) implies

$$\begin{aligned} \overline{n_T h} \hat{a}_1(y') - \mathbb{E}[Y_{1t}|Y_{0t-1} = y', D_t = 1] - \frac{\kappa_2}{2} f_{Y_{0t-1}|D_t=1}(y') g_1''(y') h_1^2 \\ = \frac{1}{f_{Y_{0t-1}|D_t=1}(y')} \overline{n_T h} \frac{1}{n_T} \sum_{i \in \mathcal{T}} K_{h_1}(y', Y_{0it-1}) \epsilon_{it} + o_p(1) \end{aligned} \quad (\text{A.6})$$

and

$$\begin{aligned} \overline{n_T h} \tilde{a}_0(y') - \mathbb{E}[Y_{0t}|Y_{0t-1} = y', D_t = 1] - \frac{\kappa_2}{2} f_{Y_{0t-2}|D_t=1}(z) g''(z) h^2 \\ = \frac{1}{f_{Y_{0t-2}|D_t=1}(z)} \overline{n_T h} \frac{1}{n_T} \sum_{i \in \mathcal{T}} K_h(z, Y_{0it-2}) u_{it} + o_p(1) \end{aligned} \quad (\text{A.7})$$

Step 3: (Asymptotic Normality)

Combining Equation A.6 and Equation A.7 implies

$$\begin{aligned} \sqrt{n h} \hat{a}_1(y') - \hat{a}_0(y') - \mathbb{E}[Y_{1t} - Y_{0t}|Y_{0t-1} = y', D_t = 1] - \frac{\kappa_2}{2} f_{Y_{0t-1}|D_t=1}(y') g_1''(y') h_1^2 - f_{Y_{0t-2}|D_t=1}(z) g''(z) h^2 \\ = \frac{1}{f_{Y_{0t-1}|D_t=1}(y')} \overline{n_T h} \frac{1}{n_T} \sum_{i \in \mathcal{T}} K_{h_1}(y', Y_{0it-1}) \epsilon_{it} \\ - \frac{1}{f_{Y_{0t-2}|D_t=1}(z)} \overline{n_T h} \frac{1}{n_T} \sum_{i \in \mathcal{T}} K_h(z, Y_{0it-2}) u_{it} + o_p(1) \\ = \left(\frac{1}{f_{Y_{0t-1}|D_t=1}(y')} \right)^\top \left(\begin{array}{c} \sqrt{n_T h} \frac{1}{n_T} \sum_{i \in \mathcal{T}} K_{h_1}(y', Y_{0it-1}) \epsilon_{it} \\ \sqrt{n_T h} \frac{1}{n_T} \sum_{i \in \mathcal{T}} K_h(z, Y_{0it-2}) u_{it} \end{array} \right) + o_p(1) \end{aligned}$$

Since

$$\begin{aligned} \text{Var} \quad \overline{n_T h} \frac{1}{n_T} \sum_{i \in \mathcal{T}} K_h(z, Y_{0it-2}) u_{it} &= h \mathbb{E} \left[k^2 \left(\frac{Y_{0it-2} - z}{h} \right) u_{it}^2 \right] \\ &= \frac{1}{h} \mathbb{E} \left[\sigma_u^2(Y_{0it-2}) k^2 \left(\frac{Y_{0it-2} - z}{h} \right) \right] \\ &= \int \sigma_u^2(z + vh) k^2(v) f_{Y_{0t-2}|D_t=1}(z + vh) dv \\ &= \sigma_u^2(z) f_{Y_{0t-2}|D_t=1}(z) \int k^2(v) dv + o(1) \end{aligned}$$

where $\sigma_u^2(z) = E[u_{it}^2 | Y_{0t-2} = z]$. Similarly, one can show that

$$\text{Var} \quad \overline{\frac{1}{n_T h}} \frac{1}{n_T} \sum_{i \in \mathcal{T}} K_{h_1}(y', Y_{0it-1}) \epsilon_{it} = \sigma_\epsilon^2(y') f_{Y_{0t-1}|D_t=1}(y') \int k^2(v) dv + o(1)$$

and

$$\begin{aligned} \text{Cov} \quad & \overline{\frac{1}{n_T h}} \frac{1}{n_T} \sum_{i \in \mathcal{T}} K_h(z, Y_{0it-2}) u_{it}, \quad \overline{\frac{1}{n_T h}} \frac{1}{n_T} \sum_{i \in \mathcal{T}} K_{h_1}(y', Y_{0it-1}) \epsilon_{it} \\ &= \frac{1}{h} E \left[k_h \left(\frac{Y_{0it-2} - z}{h} \right) u_{it} k_{h_1} \left(\frac{Y_{0it-1} - y'}{h_1} \right) \epsilon_{it} \right] \\ &= \frac{1}{h} E \left[\sigma_{u\epsilon}(z, y') k_h \left(\frac{Y_{0it-2} - z}{h} \right) k_{h_1} \left(\frac{Y_{0it-1} - y'}{h_1} \right) \epsilon_{it} \right] \\ &= \frac{1}{h} \int \int \sigma_{u\epsilon}(z, y') k_h \left(\frac{\tilde{z} - z}{h} \right) k_{h_1} \left(\frac{y_{0t-1} - y'}{h_1} \right) f_{Z, Y_{0t-1}|D_t=1}(\tilde{z}, y_{0t-1}) d\tilde{z} dy_{0t-1} \\ &= h_1 \int \int \sigma_{u\epsilon}(z, y') k_h(\tilde{v}) k_{h_1}(v) f_{Z, Y_{0t-1}|D_t=1}(z + \tilde{v}h, y' + vh_1) d\tilde{v} dv \\ &= O(h_1) \end{aligned}$$

This implies

$$\begin{pmatrix} \sqrt{n_T h} \frac{1}{n_T} \sum_{i \in \mathcal{T}} K_{h_1}(y', Y_{0it-1}) \epsilon_{it} \\ \sqrt{n_T h} \frac{1}{n_T} \sum_{i \in \mathcal{T}} K_h(z, Y_{0it-2}) u_{it} \end{pmatrix} \xrightarrow{d} N(0, \Omega)$$

where

$$\Omega = \begin{pmatrix} \sigma_\epsilon^2(y') f_{Y_{0t-1}|D_t=1}(y') \kappa & 0 \\ 0 & \sigma_u^2(z) f_{Y_{0t-2}|D_t=1}(z) \kappa \end{pmatrix}$$

which implies the result that

$$\begin{aligned} \sqrt{n h} \left[a_1(y') - a_0(y') - E[Y_{1t} - Y_{0t} | Y_{0t-1} = y', D_t = 1] - \frac{\kappa_2}{2} \left(f_{Y_{0t-1}|D_t=1}(y') g_1''(y') h_1^2 - f_{Y_{0t-2}|D_t=1}(z) g''(z) h^2 \right) \right] \\ \xrightarrow{d} N(0, V) \end{aligned}$$

where

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

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

B Tables and Figures

Table 1: Summary Statistics

	Displaced		Non-displaced	
	mean	sd	mean	sd
Earnings \$1000s				
Earnings 2006	43.2	39.9	48.0	42.7
Earnings 2008	42.6	39.9	50.4	57.5
Earnings 2010	30.1	39.3	48.4	44.4
Earnings 2012	35.0	42.6	51.4	55.4
Demographics				
Birth Year	1948.7	6.8	1947.8	6.8
% Male	0.44	0.50	0.44	0.50
% White	0.81	0.39	0.85	0.35
% Black	0.12	0.33	0.09	0.29
% No Degree	0.07	0.26	0.05	0.22
% HS Degree	0.59	0.49	0.56	0.50
% College Degree	0.33	0.47	0.39	0.49
Employment				
Works FT 2006	0.787	0.410	0.768	0.422
Works FT 2008	0.650	0.478	0.731	0.444
Works FT 2010	0.494	0.502	0.683	0.465
Works FT 2012	0.512	0.501	0.589	0.492
Unemployed 2006	0.000	0.000	0.000	0.000
Unemployed 2008	0.069	0.254	0.009	0.095
Unemployed 2010	0.163	0.370	0.011	0.103
Unemployed 2012	0.069	0.254	0.017	0.128
Retired 2006	0.000	0.000	0.000	0.000
Retired 2008	0.006	0.079	0.017	0.128
Retired 2010	0.087	0.283	0.021	0.145
Retired 2012	0.062	0.243	0.100	0.300

Table 2: Correlated Random Effects estimates of the effect of displacement on log(earnings)

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
Displaced	-0.533*	-0.539*	-0.677*	-0.412*	-0.474*	-0.517*	-0.335
	(0.126)	(0.112)	(0.129)	(0.162)	(0.118)	(0.136)	(0.185)
2012 Dummy	-0.053	-0.050	-0.005	-0.142	0.066	0.089*	0.046
	(0.041)	(0.037)	(0.035)	(0.084)	(0.042)	(0.040)	(0.099)
Birth Year		0.051*	0.047*	0.049*	0.004	0.004	0.010
		(0.003)	(0.003)	(0.005)	(0.005)	(0.005)	(0.011)
Female		-0.373*	-0.360*	-0.381*	-0.369*	-0.352*	-0.396*
		(0.036)	(0.042)	(0.071)	(0.041)	(0.048)	(0.085)
African American		-0.127*	-0.067	-0.087	-0.190*	-0.107	-0.172
		(0.061)	(0.062)	(0.102)	(0.068)	(0.070)	(0.116)
HS		0.349*	0.103	0.454*	0.395*	0.238*	0.478*
		(0.081)	(0.084)	(0.148)	(0.096)	(0.100)	(0.191)
BA		0.717*	0.291*	0.522*	0.818*	0.503*	0.608*
		(0.081)	(0.091)	(0.157)	(0.096)	(0.106)	(0.195)
MA/MBA		0.990*	0.638*	0.845*	0.964*	0.797*	0.872*
		(0.088)	(0.101)	(0.172)	(0.104)	(0.118)	(0.229)
Law/MD/PhD		1.392*	0.872*	1.026*	1.416*	1.126*	0.864*
		(0.110)	(0.127)	(0.223)	(0.148)	(0.161)	(0.345)
Subsample					Young	Young	Young
Occupation FE	No	No	Yes	Yes	No	Yes	Yes
Time Varying Occ.	No	No	No	Yes	No	No	Yes
Industry FE	No	No	Yes	No	No	Yes	No
R ²	0.016	0.224	0.349	0.329	0.179	0.305	0.351
Adj. R ²	0.015	0.219	0.334	0.287	0.170	0.278	0.276
Num. obs.	3132	3132	2818	1298	1872	1684	720

* $p < 0.05$

The dependent variable is the log of earnings in 2006 and 2012. The excluded group is white males without a HS degree. The regressions also include an additional dummy variables for Other Race, GED, Associate's Degree, and Other Education Level whose estimated coefficients are not presented for conciseness. The sample size decreases when Occupation and Industry Fixed Effects are included because these are not observed for all individuals in the sample. The last two columns limit the sample to individuals who are 64 or younger in 2012.

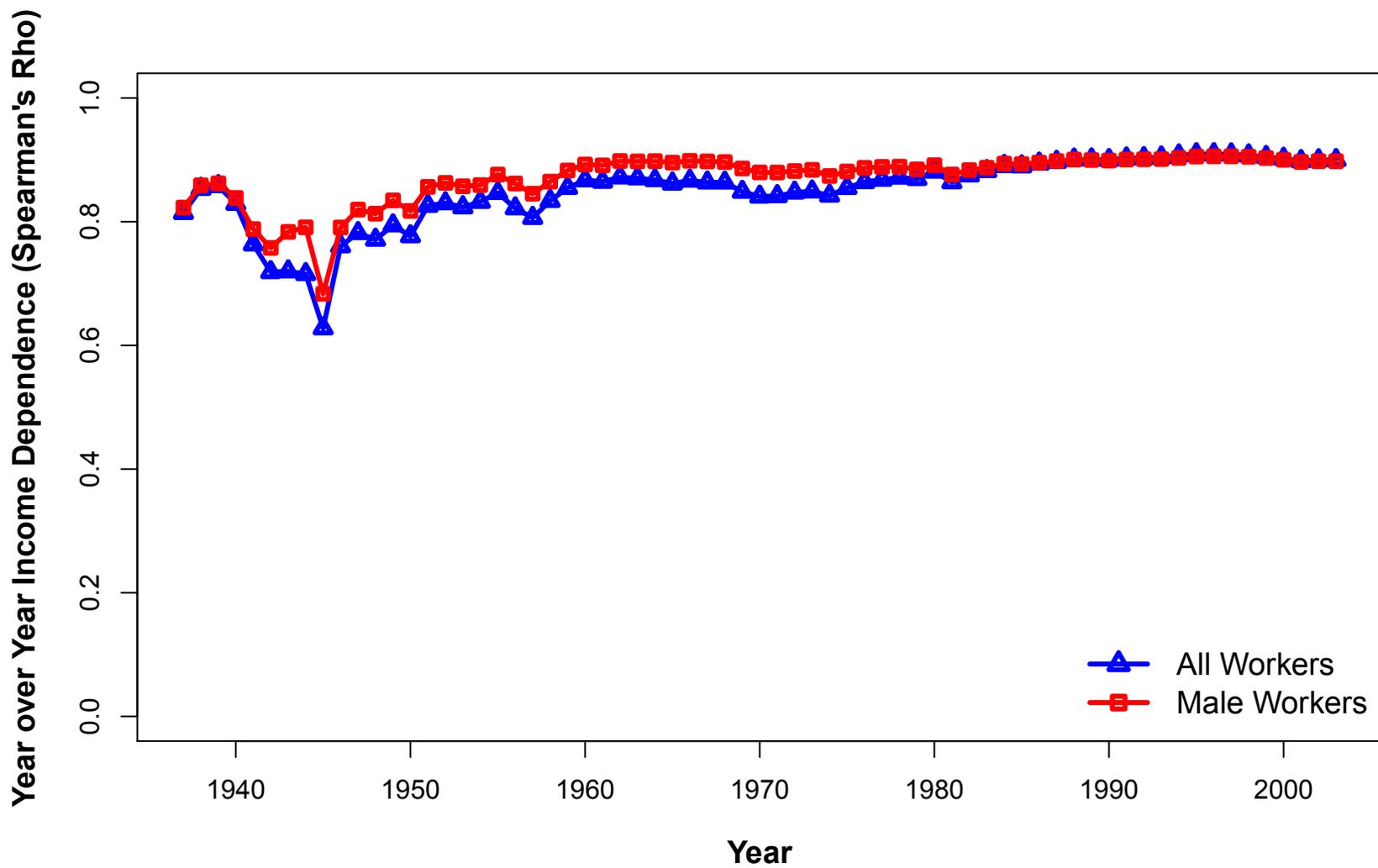


Figure 1: Rank Correlation (Spearman's Rho) of Year over Year Annual Income Dependence for All Workers and Male Workers from 1937-2003. The data comes from Kopczuk, Saez, and Song (2010) and replicates part of Figure 4 in that paper.

Marginal Distributions Potential Outcomes for the Treated Group

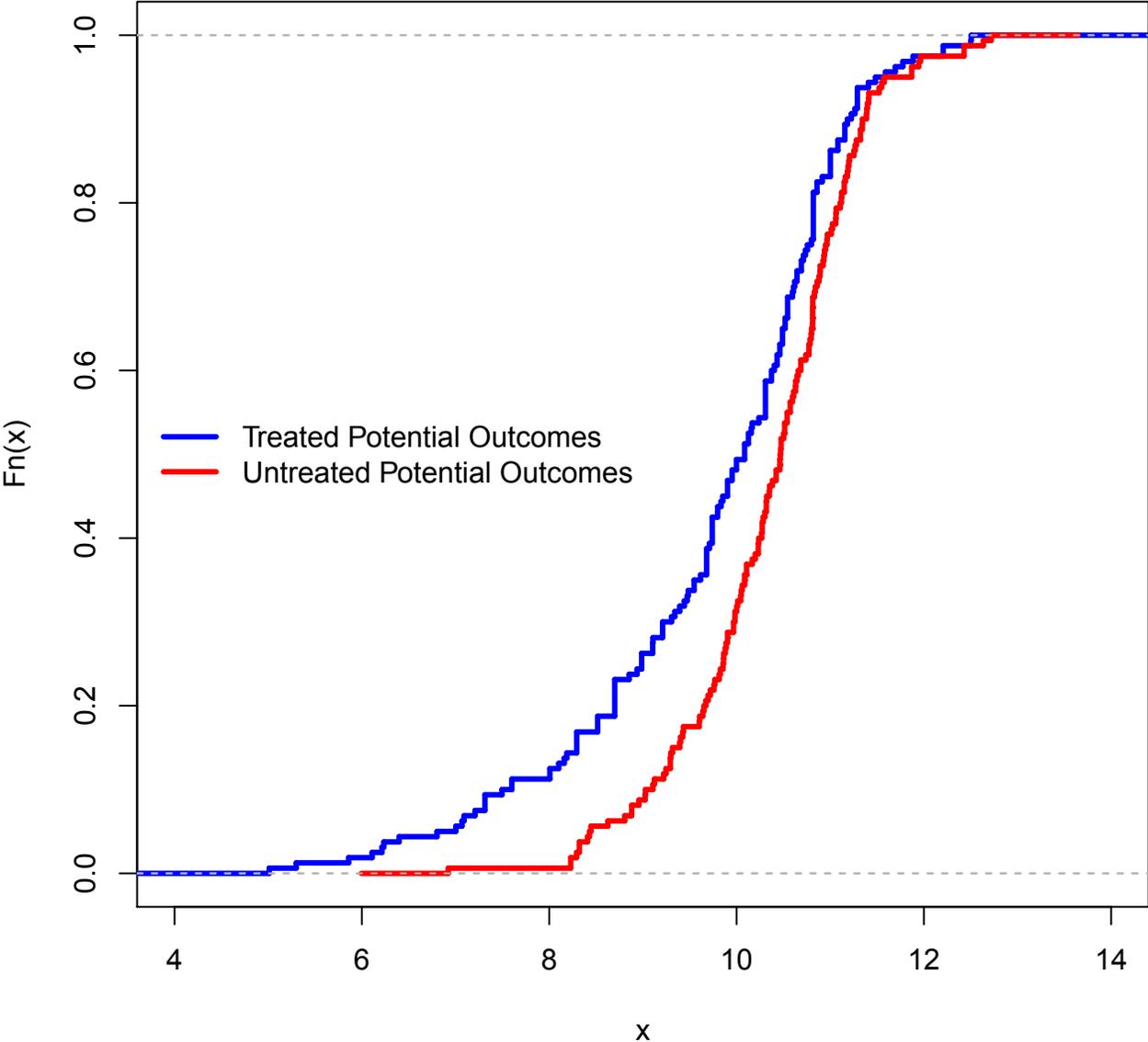


Figure 2: Marginal Distributions of Displaced and Non-displaced Potential Outcomes for the Displaced Group estimated under the Distributional Difference in Differences Assumption.

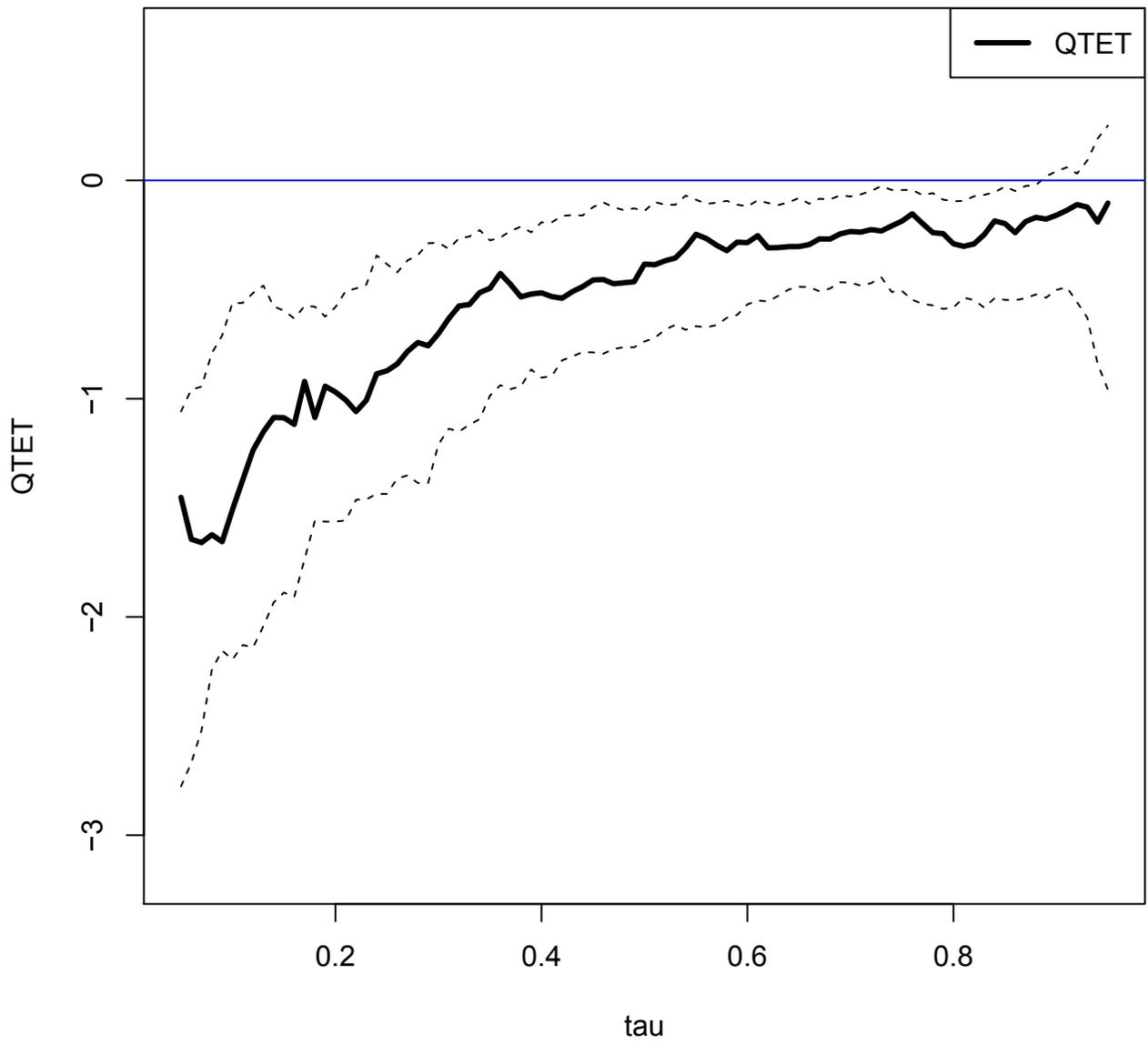


Figure 3: The Quantile Treatment Effect on the Treated estimated under the Distributional Difference in Differences Assumption.

Quantile of the Treatment Effect on the Treated

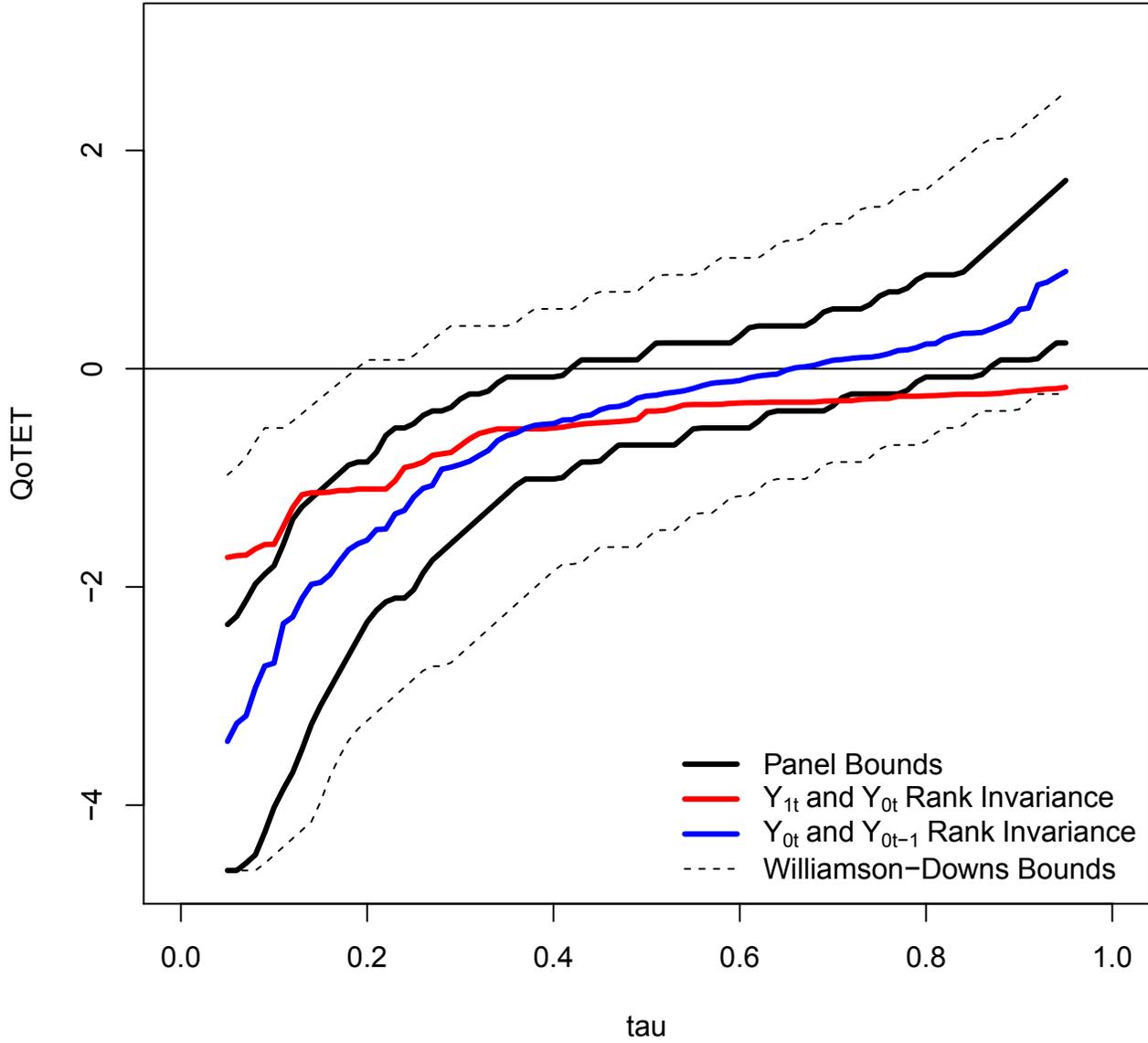


Figure 4: Bounds on the Quantile of the Treatment Effect

The ‘Panel Bounds’ are the estimates coming from the method in the current paper. The ‘ Y_{1t} and Y_{0t} Rank Invariance’ estimates come from employing the cross-sectional perfect positive dependence assumption. The ‘ Y_{0t} and Y_{0t-1} Rank Invariance’ estimates come from applying the assumption of perfect positive dependence in non-displaced potential earnings over time. The ‘Williamson-Downs Bounds’ come from using only information about the marginal distributions of displaced and non-displaced potential earnings without applying any restrictions on their dependence.

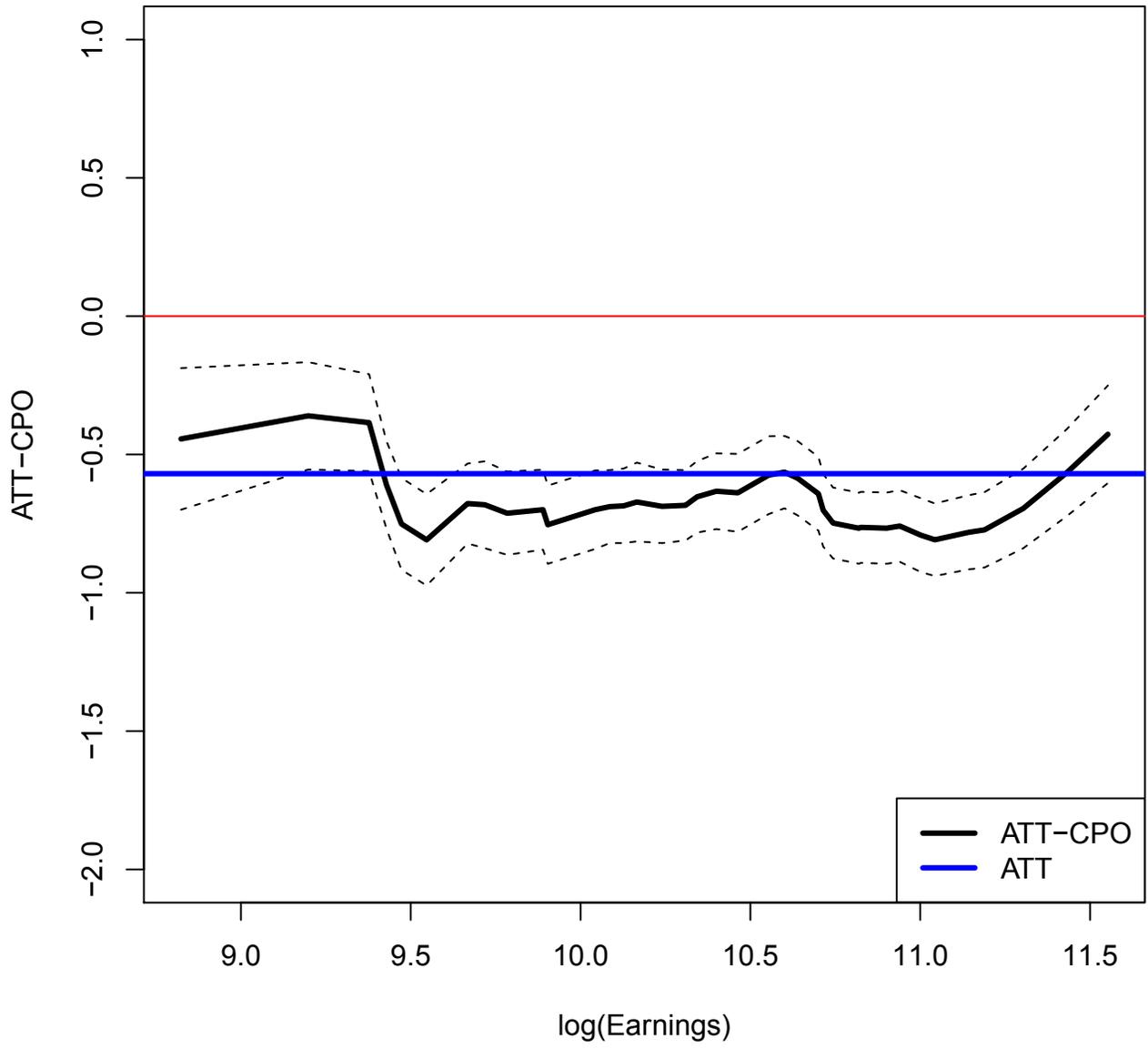


Figure 5: The ATT-CPO