Bounding Treatment Effects in Experimental Studies with Non-Compliance: The Value of Follow Up Surveys

Orville Mondal

October 24, 2022

Abstract

In this paper I study the problem of identifying the causal effect of an experimental treatment when the experiment suffers from non-compliance. In particular, I consider the identifying power of information collected from non-complying participants after the completion of the treatment phase. Follow up surveys often ask study participants why they chose not to accept an offer of treatment despite being assigned to it, and answers to such questions offer insights into the decision process by which agents choose to comply with their assigned treatment status. I propose a model that rationalizes an agent's compliance decision. Based on this model, I characterize the set of values for the average treatment effect that are conformable with data observed in a randomized trial. This model and the implied set of identified values for the average treatment effect rely crucially on the availability of follow up surveys that ask agents why they chose to not comply. This underscores the importance of following up with non-complying agents since the model often leads to substantially tighter identified sets for the average treatment effect than what is possible without this information. I apply the proposed model to data from the Job Training Partnership Act Study to estimate identified sets for the average treatment effect for a number of employment outcomes.

1 Introduction

Program evaluation has been a mainstay of empirical research in economics for decades and encompasses a vast and growing collection of studies which analyze the impact of policy interventions. For example, consider a government sponsored training program designed to improve labor market outcomes for those who are socially or economically disadvantaged. A key question of interest is if this program is actually effective and to what extent it succeeds in attaining its stated goals. If the program is expensive to maintain then the results of an impact evaluation study would allow policy makers to make an informed decision when adjusting the scope or scale of the program. A randomized trial is often the basis for any empirical study which aims to quantify the impact of the training program on labor outcomes. In such a trial, individuals are randomly assigned to one of two groups. Members of one group are allowed training under the program while members of the other group are not. These groups are sometimes referred to as treatment arms with the former and

latter being called the treatment and control arms, respectively. After this initial assignment, members in the treatment arm receive training under the program, following which the study records outcomes for all participants in the trial. Comparing the labor outcomes of individuals in the two groups can then be used to identify the effect of the policy. Randomization of this form has been extensively used in studies which involve human participants and the principles of any such analysis are widely understood and easy to communicate. A crucial requirement for any such analysis is that individuals comply with their assigned treatment status i.e., all individuals placed in the treatment arm must receive training while those placed in the control arm should not receive any training. In practice, noncompliance is extremely common since it is often not possible to enforce perfect compliance of agents with their assigned treatment. If only some individuals in the treatment arm actually accepted training under the program, then it is possible that their self-selection was driven by information or characteristics which were specific to them and therefore rendered them systematically different from other individuals in the treatment group. In effect, the labor outcomes of the control group are no longer an appropriate reference against which to compare the outcomes of those individuals in the treatment arm who actually received training.

Non-compliance in randomized trials is a well recognized problem and it is clear that understanding the process which dictates whether or not members in the treatment arm choose to comply with their assigned status is an important component of any policy evaluation study which relies on data from the trial. Many trials conduct a follow up survey where non-complying members of the treatment arm are asked to list reasons for why they chose not to get trained under the program. In this paper I propose a simple framework which exploits this information to rationalize non-compliance behavior observed in data from a randomized trial as described above. The impact of services provided under a job training program is then analyzed and compared against various standard approaches. It should be noted that most existing policy evaluation studies typically only utilize outcome data and whether an individual complied with their assigned treatment. Reasons for non-compliance are rarely used in a systematic manner when attempting to measure the impact of a policy intervention. To empirically illustrate the importance of this information, I use data from the Job Training Partnership Act Study to estimate the impact of job training and job search assistance provided under the program on various employment outcomes. Data from this study has been used in several papers on policy evaluation though information for the reasons of non-compliance has never been used to systematically explain non-compliance.

1.1 Related Literature

This paper is primarily related to the literature on identifying and estimating the average treatment effect and other causal parameters using either observational data or data from experimental studies. For a review see Abadie and Cattaneo [2018]. In particular, it concerns the identification of average treatment effects in experimental studies with non-compliance where the causal parameter of interest can only be partially identified. The basic framework starts with the causal model of Rubin [1974] but goes back to the analysis of Fisher [1935]. A review of relevant assumptions and key results

in the absence of non-compliance may be found in Imbens [2004]. The issue of non-compliance in experimental studies presents complications for estimating treatment effects which are often mirrored when using data from observational studies. Heckman and Hotz [1989] discusses the issue of estimating treatment effects using data from observational studies. Angrist and Imbens [1991], Imbens and Angrist [1994] are seminal papers which investigate identification of causal parameters in the presence of non-compliance. Imbens and Rubin [1997] analyzes the problem of inference for causal parameters in experimental studies with non-compliance in a Bayesian framework.

Non-compliance, where only some agents eligible for treatment choose to accept is related to the larger problem of selection bias which has been studied extensively, for example in Heckman [1974], Heckman [1976], Manski [1990]. While I will propose an underlying model that describes the pattern of self selection, the analysis in what follows will not directly deal with the problem of estimating conditional means while correcting for bias due to self selection. The issue of non-compliance I consider is distinct from what is typically labeled non-compliance or non-adherence in the medical literature (see for example Kleinsinger [2003], Kleinsinger [2010]). Issues of non-compliance in medicine often involve discussions of how to pursue treatment when a patient does not exactly comply with or follow the recommended treatment strategy, whereas I am concerned with identifying the impact of an experimental intervention after the conclusion of the experiment, or treatment phase.

More relevant to this paper is the extensive literature of partially identified treatment effects as studied in Manski [1990], Balke and Pearl [1997], Heckman and Vytlacil [1999], Manski [2003], Bhattacharya et al. [2008], Chen et al. [2012] ,Demuynck [2015]. By extension, this paper is related to the broader literature on partially identified parameters surveys, reviews of which can be found in Tamer [2010], Molinari [2019].

Much of the analysis relies on features observed in the Job Training Partnership Act (JTPA) Study, though similar features are possibly present in many other experimental datasets which are not publicly available. The JTPA Study has been widely studied and used in several papers. Bloom et al. [1993], Bloom et al. [1997] describe the experiment in great detail and present detailed results for program efficacy. Heckman and Smith [1997] used data from the JTPA Study to examine how estimates of program impact are sensitive to how data is aggregated or processed in the course of statistical analysis. Since these early papers, the JTPA has been used in numerous other studies (Abadie et al. [2002], Courty and Marschke [2004], Kitagawa and Tetenov [2018], Kitagawa and Tetenov [2021]). In what follows I will propose a model which assumes that agents assigned to treatment have some belief about potential program benefits before the treatment phase. Smith et al. [2020] uses the JTPA Study to investigate how agents evaluate the performance of the program but post treatment. The JTPA Study did not directly ask agents how they expected the program to help them, but rather why they chose not to accept services under the JTPA. Their response is assumed to be informative of their subjective expectations regarding program efficacy. This is related to but distinct from studies which directly elicit information about agents' subjective expectations about their outcomes in the future e.g. Dominitz and Manski [1997].

2 Identifying Causal Effects in the Presence of Non-Compliance

I begin with a brief introduction of the potential outcomes framework of Rubin [1974], and as discussed in Imbens and Rubin [2015]. This framework is often referred to as the Rubin Causal Model or Neyman-Rubin causal model and it clearly llustrates the issue of non-compliance and its impact on the identifiability of treatment effects using data from a randomized trial. Consider an outcome of interest Y_i observed for agent i. The causal effect of treatment on outcome Y_i can be defined in terms of potential outcomes. Let $Y_i(1)$ be the outcome for agent i if they had received treatment and let $Y_i(0)$ be their outcome if they had not received treatment. The average treatment effect (ATE) in this framework is defined as,

$$\text{ATE } = \mathbb{E}[Y_i(1) - Y_i(0)], \tag{1}$$

where the expectations operator aggregates over the joint distribution of potential outcomes. Let D_i denote a binary indicator variable equal to 1 if agent i received treatment. Assume that the observed outcome Y_i is related to potential outcomes as $Y_i = D_i \cdot Y_i(1) + (1 - D_i) \cdot Y_i(0)$. This relationship demonstrates a fundamental problem in any study which seeks to determine the impact of a treatment - only one of the two potential outcomes is observed for agent i, as determined by whether or not they received treatment. Now let Z_i denote a binary variable equaling 1 if the agent was assigned to the treatment group. For now it is assumed that assignment is independent of potential outcomes i.e.

$$Z_i \perp (Y_i(1), Y_i(0)),$$

which implies that the assignment decision itself does not influence the potential outcomes. $D_i(z)$ for $z = 0, 1$ then defines two potential outcomes for the treatment indicator. Perfect compliance with the assigned treatment is defined by,

$$D_i(z) = z, \quad \text{for } z = 0, 1,$$

with imperfect or non compliance being defined as a violation of the above. If there is perfect compliance, while only one potential outcome is observed for agent i, the ATE can still be identified from the data of a randomized trial as,

$$\mathbb{E}[Y_i \mid Z_i = 1] - \mathbb{E}[Y_i \mid Z_i = 0]. \tag{2}$$

This follows because for outcome $d \in \{0, 1\}$,

$$\begin{aligned} \mathbb{E}[Y_i \mid Z_i = d] &= \mathbb{E}[Y_i(1) \cdot D_i(d) + (1 - D_i(d)) \cdot Y_i(0) \mid Z_i = d] \\ &= \mathbb{E}[Y_i(d) \mid Z_i = d] = \mathbb{E}[Y_i(d)]. \end{aligned} \tag{3}$$

The first equality above follows by the definition of the observed outcome; the second equality follows under the assumption of perfect compliance, and the final equality follows because assignment to a

treatment group was random. There is also the implicit assumption that merely being assigned to the treatment or control group cannot influence the potential outcomes $Y_i(1), Y_i(0)$. The rationale behind the definition in (2) is quite simple - if assignment was independent of potential outcomes, and everyone complies with their assignment, then any difference in outcomes must be purely due to the treatment. The presence of non-compliance breaks the above relationship and implies that the ATE cannot be identified from the data collected in a randomized trial as described here. To see this return to (3) and notice,

$$\mathbb{E}[Y_i \mid Z_i = d] = \mathbb{E}[Y_i(0)] + \mathbb{E}[Y_i(1) - Y_i(0) \mid D_i(d) = 1, Z_i = d] \cdot \mathbb{P}(D_i(d) = 1 \mid Z_i = d), \tag{4}$$

where $\mathbb{P}(D_i(d) = 1 \mid Z_i = d) < 1$ when there is non-compliance. If we tried to calculate the quantity in (2) then instead of the ATE, we would have:

$$\mathbb{E}[Y_i(1) - Y_i(0) \mid D_i(1) = 1, Z_i = 1] \cdot \mathbb{P}(D_i(1) = 1 \mid Z_i = 1) - \\ \mathbb{E}[Y_i(1) - Y_i(0) \mid D_i(0) = 1, Z_i = 0] \cdot \mathbb{P}(D_i(0) = 1 \mid Z_i = 0). \tag{5}$$

While the ATE cannot be identified using data from a study which suffers from non-compliance, there are other causal effects which still can be. The object in (2) still represents a difference in outcomes between two groups of people who are similar apart from random variation as long as Z_i is independent of potential outcomes. Moreover, it can be argued that it is realistic to expect some degree of non-compliance in real world settings where the authority assigning agents to treatment cannot directly influence an applicant's decision to accept or adhere to their assigned treatment. These observations are why the comparison in (2) is said to measure the ATE from an intention to treat (ITT) perspective. In effect, (2) measures the impact of an *offer of treatment* instead of the impact of the treatment itself. The various methodological benefits of adopting an ITT approach are well known, though the shortcomings are often difficult to ignore (see for example Hewitt et al. [2006]). Imbens and Angrist [1994] show that the data can also be used to estimate the ATE for a sub-group of the population, relying on the fact that an offer of treatment can by itself change the probability that a participant will actually accept and receive treatment. This is known as the local average treatment effect (LATE, or complier average treatment effect as in Hewitt et al. [2006]) and is given by,

$$\frac{\mathbb{E}[Y_i \mid Z_i = 1] - \mathbb{E}[Y_i \mid Z_i = 0]}{\mathbb{P}(D_i = 1 \mid Z_i = 1) - \mathbb{P}(D_i = 1 \mid Z_i = 0)} = \mathbb{E}[Y_i(1) - Y_i(0) \mid D_i(1) - D_i(0) = 1]. \tag{6}$$

The conditions under which the equality in (6) holds are standard and often reasonable (see Imbens and Angrist [1994]). Since D_i is binary, the conditioning event on the right hand side of equation (6) shows that identified quantities may be used to define the average treatment effect for the sub-group of agents who would not receive treatment unless they were offered it - such agents are often called compliers.

While both the LATE and ITT effects are widely used, they suffer from some disadvantages.

The ITT effect captures the effect of an offer of treatment while the LATE measures the impact of treatment on a subset of the whole population of agents. Neither object captures the ATE which measures the impact of actual treatment received on the entire population - this is a fundamental object of interest in many studies and is also the primary focus of this paper. Even if the ATE cannot be point identified using data from a randomized trial with non-compliance, it is often possible to bound its magnitude. Many studies (e.g. Manski [1990]) have pointed out that the ATE can generally be bounded as long as the potential outcomes are bounded. For example, if it is known that the outcomes $Y_i(1), Y_i(0)$ belong to the interval $[0,1]$, then the ATE must lie within $[-1,1]$. The structure of a randomized trial often provides additional information which may be used to derive tighter bounds on the ATE. For example, suppose that only agents who have been assigned to treatment can actually receive treatment i.e $D_i(0) = 0$. Then, letting θ^{LATE} denote the local average treatment effect, the definition in (6) can be used to establish the following relationship:

$$\begin{aligned}\mathbb{E}[Y_i(1) - Y_i(0)] =& \mathbb{E}[Y_i(1) - Y_i(0) \mid D_i(1) - D_i(0) = 1] \cdot \mathbb{P}(D_i(1) - D_i(0) = 1) + \\ & \mathbb{E}[Y_i(1) - Y_i(0) \mid D_i(1) - D_i(0) \neq 1] \cdot \mathbb{P}(D_i(1) - D_i(0) \neq 1) \\ =& \theta^{LATE} \cdot \mathbb{P}(D_i = 1 \mid Z_i = 1) + \\ & \mathbb{E}[Y_i(1) - Y_i(0) \mid D_i = 0, Z_i = 1] \cdot \mathbb{P}(D_i = 0 \mid Z_i = 1) \end{aligned} \tag{7}$$

The second term in the sum above accounts for the average treatment effect for non-compliers. This quantity cannot be identified from data but it must lie within the interval $[-1,1]$ since outcomes lie within $[0,1]$. Therefore, the average treatment effect must be bounded below and above by the following identified quantities[1]:

$$\begin{aligned}&\theta^{LATE} \cdot \mathbb{P}(D_i = 1 \mid Z_i = 1) - \mathbb{P}(D_i = 0 \mid Z_i = 1) \\ &\theta^{LATE} \cdot \mathbb{P}(D_i = 1 \mid Z_i = 1) + \mathbb{P}(D_i = 0 \mid Z_i = 1).\end{aligned} \tag{8}$$

The bounds above imply that the average treatment effect is only partially identified since only an interval containing it can be identified from data, as opposed to its actual value. This paper will focus on deriving bounds on the average treatment effect which are tighter than those in (8) in the sense that the set of identified values for the ATE will correspond to an interval of length shorter that that implied by (8).

A vast literature now exists on analyzing partially identified parameters in various areas of economic interest (see Molinari [2019] for a review). In the case of the average treatment effect, the focus is often on deriving non-parametric bounds on the causal parameter of interest. These bounds usually exploit observable features of the outcome data, the treatment assignment mechanism, and whether or not treatment is actually received or accepted. This is often the only relevant

[1] these bounds are not sharp since they assume $\mathbb{E}[Y_i(1) - Y_i(0) \mid D_i = 0, Z_i = 1] \in [-1,1]$ but do not use the information that $\mathbb{E}[Y_i(1) - Y_i(0) \mid D_i = 0, Z_i = 1] = \mathbb{E}[Y_i(1) \mid D_i = 0, Z_i = 1] - \mathbb{E}[Y_i \mid D_i = 0, Z_i = 1]$ where the latter term in this difference is identified.

information that is available from an experiment. In the absence of additional information, any assumptions on the treatment selection process - the unobserved process by which an agent chooses to accept or reject an offer or treatment- may appear ad hoc or implausible. It is clear however that understanding why an agent would choose to not accept treatment may offer important information which can then be used to bound the ATE. Even a simple model explaining the treatment selection process can have non trivial implications.

2.1 The Treatment Selection Process - A Simple Model

This section illustrates the importance of understanding the treatment selection process by describing a basic model determining when an agent chooses to accept an offer of treatment. Assignment is still assumed to be random and uncorrelated with potential outcomes, and it is still assumed that agents not assigned to treatment cannot receive it. Suppose that when an agent is offered treatment, they form a belief about the potential benefits of treatment using any private information known to them. This is assumed to be,

$$\theta_i^e = \mathbb{E}[Y_i(1) - Y_i(0) \mid \mathcal{I}_i],$$

where $\mathcal{I}_i$ represents information known to agent i when they are offered treatment. For instance, an agent may have a sense of what their outcome will be without treatment, or they may have heard from others about the treatment's efficacy. This information would be included in $\mathcal{I}_i$. The expectations operator in (9) is with respect to the true joint distribution of potential outcomes, given $\mathcal{I}_i$. This assumption is often labeled rational expectations. Now assume that an agent who is offered treatment only accepts if they think it will be beneficial to do so:

$$D_i = \mathbb{1}\left\{\theta_i^e \geq 0\right\}. \tag{9}$$

The compliance decision rule in (9) is very similar in structure to a sector selection rule used in a Roy model of labor choice (see Roy [1951], Heckman and Honoré [1990]). However, a Roy model would generally assume that agents know their exact potential outcomes $Y_i(1), Y_i(0)$ and not just their average difference, conditional on some information. Mourifié et al. [2020] considers an setup as in (9) and some of the discussion which follows will mirror the observations therein. If the true treatment effect for agent i is denoted $\theta_i = Y_i(1) - Y_i(0)$, then the following is true

$$\theta_i = \theta_i^e + \nu_i,$$

where crucially,

$$\mathbb{E}[\nu_i \mid \mathcal{I}_i] = 0.$$

The primary object of interest is the ATE or $\mathbb{E}[\theta_i]$. Recall from (7) that the ATE may be decomposed into a part that is point identified from data and a part that is not. This simple setup already has implications for the identified set for those components of the ATE which can only be partially

identified. For instance, the model so far implies that:

$$\mathbb{E}[\theta_i \mid D_i = 0, Z_i = 1] = \mathbb{E}[\theta_i \mid \theta_i^e < 0] = \mathbb{E}[\theta_i^e + \nu_i \mid \theta_i^e < 0].$$

The law of iterated expectations then delivers the following:

$$\mathbb{E}[\theta_i \mid D_i = 0, Z_i = 1] = \mathbb{E}[\theta_i^e + \nu_i \mid \theta_i^e < 0] = \mathbb{E}[\theta_i^e \mid \theta_i^e < 0] + \mathbb{E}\left[\mathbb{E}[\nu_i \mid \mathcal{I}_i, \theta_i^e < 0]\right] < 0, \tag{10}$$

where the final inequality follows because the first object in the sum above is negative, and the second object must be zero due to the properties of ν_i. Similarly, the decision rule in (9) implies:

$$\mathbb{E}[\theta_i \mid D_i = 1, Z_i = 1] = \mathbb{E}[\theta_i \mid \theta_i^e \geq 0] = \mathbb{E}[\theta_i^e \mid \theta_i^e \geq 0] + \mathbb{E}\left[\mathbb{E}[\nu_i \mid \mathcal{I}_i, \theta_i^e \geq 0]\right] \geq 0. \tag{11}$$

Taken together, the two prior inequalities establish that if agents make their decision to participate as in (9), then the average treatment effect must be non-negative for those agents who accepted the offer for treatment while it must be non-positive for those who reject the offer of treatment. Returning to the decomposition in (7), the following lower bound on the ATE still applies:

$$\mathbb{E}[\theta_i] \geq \mathbb{P}(D_i = 1 \mid Z_i = 1) \cdot \theta^{LATE} - \mathbb{P}(D_i = 0 \mid Z_i = 1), \tag{12}$$

while the upper bound is reduced to:

$$\mathbb{E}[\theta_i] \leq \mathbb{P}(D_i = 1 \mid Z_i = 1) \cdot \theta^{LATE}, \tag{13}$$

where the conditional mean of non-compliers is bounded above by zero due to the relation in (10) implied by the selection model in (9). This shows that imposing some structure to the selection process implies tighter bounds on the ATE relative to the widest possible bounds derived under no additional assumptions.

The definition in (9) is simple but not innocuous. The fact that there is no agent-specific subscript for the expectations operator is important. It means each agent knows the population level joint distribution of potential outcomes given their information $\mathcal{I}_i$. This rules out a more flexible notion of subjective expectations where an agent i could believe that the joint distribution of potential outcomes is different from the true distribution. Nevertheless, the notion that an agent makes their decision to accept treatment based on some private evaluation of the potential benefits is not far fetched.

A far more important consequence of the simple selection model in (9) is that it has testable implications for the LATE. Under the assumptions on treatment assignment and compliance maintained so far, the model implies the inequality in (11) which states that θ^{LATE} must be non-negative. If the identified value of θ^{LATE} violates this, then the model must be rejected. This paper will derive bounds on the ATE using a more general model which will rule out this possible incompatibility of model with data while also leveraging the fact that some randomized trials often collect additional

information from agents which may help to better understand the treatment selection process.

2.2 Understanding Non-Compliance - the JTPA Study

Sometimes, a randomized trial has additional information which can be used to better understand the treatment selection process. The model in (9) may be rejected by observed data - this suggests that an agent's decision to participate may be determined by a more complicated process. For instance, the lower bound in (12) is derived under the implication that the average treatment effect for agents who do not accept treatment is negative. However, this ignores the possibility that an agent may base their decision to accept treatment on factors which are not directly related to the treatment effect. For instance, suppose an agent faces a barrier to participation that is not related to the impact of treatment in any way. This could be difficulty in commuting to the treatment site, or an unforeseen medical emergency which incapacitates them for the duration of treatment. The decision rule in (9) needs to be made richer but it is perhaps not immediately clear how to do so. When designing a randomized trial, researchers can choose to follow up with non-complying agents and ask them why they did not accept treatment. This information may then be used to inform the choice of treatment selection model, and indeed this is exactly the manner of information collected during the Job Training Partnership Act (JTPA) Study.

The study was commissioned in the late 1980's to assess the benefits of training provided under the JTPA to economically disadvantaged individuals and out of school youths with the aim of improving their employment outcomes. Applicants who were eligible for assistance under the program were randomly assigned to a treatment group, which was allowed access to the services of the program, or to a control group, which was not. Non-compliance was prevalent and when agents were contacted in follow up surveys, they were asked to provide reasons for why they did not choose to accept treatment when they were eligible for it. Table 1 lists some of the common reasons given by adult participants in the study who were assigned to the treatment arm but ultimately did not receive any treatment.

Reason	% of respondents
Took Job	18
Changed Mind	7
Needed Job	4
Transport Problem	3
Health Problem	2

Table 1: Reasons for non-compliance in the JTPA study.

The percentages shown in Table 1 correspond to the group of non-compliers who provided a reason (this is approximately half the entire sample of non-compliers). Follow up interviews were performed over the phone if possible, and in person if not. Table 1 does not list all possible answers recorded in the follow up- this complete list and more details about the study can be found in the empirical application section below (see Table 3).

The reasons for non-compliance in Table 1 appear to broadly fall into two groups - one group of reasons indicates that the agent thought treatment would not be beneficial, while the second indicates that the agent thought treatment would be useful but could not accept due to external factors (e.g. medical emergency). This feature of the data effectively splits the non-compliers into two categories. Crucially, it may be reasonable to assume that the average treatment effect should be negative only for those non-complying agents who indicated that they felt treatment would not be beneficial for them. This is a subset of agents who did not accept treatment. If $\mathbb{P}^{NB}$ denotes the probability of a non-complying agent reporting that they thought treatment would not be beneficial (i.e. NB), then the lower bound in (12) would be,

$$\mathbb{P}(D_i = 1 \mid Z_i = 1) \cdot \theta^{LATE} - \mathbb{P}^{NB}. \tag{14}$$

This must be greater than the bound in (12) since $\mathbb{P}^{NB}$ is at most $\mathbb{P}(D_i = 0 \mid Z_i = 1)$. Of course, the result in (14) relies on the plausible but arbitrary assumption that the treatment effect is non-positive for only a subset of those who did not receive treatment whereas the lower bound in (12) followed from a more fundamental model for how agents decide to accept treatment. Information such as that in Table 1 is typically not utilized in studies evaluating program impacts since it is qualitative in nature and cannot be readily used in empirical analysis. To leverage the insight it provides requires a model which explains an agent's treatment selection decision, while accounting for the different reasons an agent may choose to not comply. The following section describes such a model. One immediate conclusion of the analysis here is that following up with non-complying agents is a valuable part of any randomized trial and any auxiliary information which helps explain the treatment selection process should be used in the analysis of treatment effects.

2.3 The Treatment Selection Process - Compliance Model

Motivated by the shortcomings implied by the decision rule in (9), a modified selection model is now presented. A particularly problematic feature of the simple model of (9) was that it led to testable implications which could potentially refute the model. Since the model was driven entirely by the expected treatment effect, this suggests that an agent's decision to participate is driven by more than just an idea of the potential benefits of treatment. The reasons reported in Table 1 seems to indicate that in some instances, agents do not accept an offer of treatment due to factors which are not entirely related to the treatment itself. For example, not participating in treatment due to a health problem does not reflect an applicant's lack of belief in the potential gains from treatment. Rather, it can be thought of as an unexpected shock, unrelated to treatment, which prevented the applicant from accepting.

With this observation, consider the decision problem of an applicant who has been assigned to treatment. At the point of time when the assignment decision is made, the agent forms an idea of the potential treatment effect. This individual specific notion of the benefit of treatment is denoted θ_i^e. This is once again the expected treatment effect, conditional on an agent's information set. In

addition, at the time of assignment, agents are also aware of any external cost $C_i \in \mathbb{R}$ which may affect their ability to participate in treatment. C_i is not directly related to the potential benefits from treatment but does impact whether or not an agent complies with their assigned treatment. The following definition makes this explicit.

Definition INA (Information at Assignment): *Agent's information at the time an agent is assigned to a treatment arm is denoted $\mathcal{I}_i$ and includes knowledge of the variable C_i.*

The final decision to accept an offer of treatment is then based on whether or not the net expected benefit exceeds zero:

$$D_i = \mathbb{1}\left\{\theta_i^e - C_i \geq 0\right\}. \tag{15}$$

Neither θ_i^e nor C_i are observed in the data, though the assumption that agents know C_i at the time they were assigned to treatment is important. The relevance of this timing assumption will be revisited below. The fact that the unobserved cost variable is allowed to take on negative values is also important. If this were not true and costs were non-negative then the following would be true,

$$D_i = 1 \Rightarrow \theta_i^e \geq C_i \geq 0.$$

If θ_i^e is assumed to be the same conditional mean treatment effect as in (9), then the decision rule in (15) would once again imply that the identified LATE must be non-negative. Therefore, it is necessary to allow negative cost values to prevent a model which is potentially refuted by data. A negative cost value may be thought of as an incentive to accept treatment which is unrelated to the treatment itself. For example, an out of work mother may be living with a family member who can offer help with childcare, allowing the applicant to participate in treatment.

Simply adding the unobserved cost term does not lead to a particularly informative framework. While the model need not produce implications which conflict with data, it may not lead to useful insights for those features of the data generating process which cannot be identified or directly observed. For example, the treatment effect conditional on not accepting treatment is now given by:

$$\mathbb{E}[Y_i(1) - Y_i(0) \mid D_i = 0, Z_i = 1] = \mathbb{E}[\theta_i \mid \theta_i^e - C_i < 0], \tag{16}$$

where the latter quantity is not restricted to be positive or negative without additional assumptions on the joint distribution of model fundamentals, many components of which are unobservable. At this stage, the identified set for the ATE is the same as when no selection model is imposed. This is where the information in Table 1 can be used to enrich the model in a way that leads to tighter bounds on the ATE.

To this end, introduce a new variable R_i which can take on three distinct values. For applicants who were assigned to treatment but did not accept, R_i can equal either zero or one. It is defined to equal 1 if θ_i^e is negative, and it equals zero if θ_i^e is strictly positive. In the edge case where θ_i^e exactly equals zero, R_i equals one based on the outcome of a tie breaking rule to be described

below. R_i is assumed to be missing for agents who were assigned to treatment and accepted, or were assigned to the control group. Therefore $R_i \in \{0, 1, \text{ missing }\}$. The variable R_i splits applicants with $Z_i = 1, D_i = 0$ into two groups: those who did not think treatment would be useful, and those who may have wanted to accept treatment but were unable to do so due to an external cost. R_i is actually observed and the selection model is completed by specifying how underlying variables define the observed values of R_i:

$$\begin{aligned} Z_i \cdot (1 - D_i) \cdot R_i &\equiv Z_i \cdot \mathbb{1}\{\theta_i^e < 0, \theta_i^e < C_i\} + Z_i \cdot \mathbb{1}\{0 = \theta_i^e < C_i\} \cdot B_i, \\ Z_i \cdot (1 - D_i) \cdot (1 - R_i) &\equiv Z_i \cdot \mathbb{1}\{0 < \theta_i^e < C_i\} + Z_i \cdot \mathbb{1}\{0 = \theta_i^e < C_i\} \cdot (1 - B_i) \end{aligned} \tag{17}$$

where the variable B_i determines if an agent chooses to reject the offer of treatment when $\theta_i^e = 0$. Similar to R_i, the tie breaking variable is assumed to take one of three values $\{0, 1, \text{missing}\}$, where B_i is defined to be missing for agents who comply with their assigned treatment. It is important to account for the edge case of $\theta_i^e = 0$ with a non-degenerate tie breaking rule to prevent situations where the model cannot rationalize the observed distribution of outcomes. To illustrate, suppose potential outcomes were binary, $Y_i \in \{0, 1\}$ is the observed outcome, $\theta_i^e = Y_i(1) - Y_i(0)$ and B_i in (17) equals one almost surely. This is a situation where an agent knows their exact treatment effect at the time of assignment. While improbable, this "perfect knowledge" assumption can be thought of as an extreme version of the rational expectations formulation for θ_i^e used in the simple model of (9). Now suppose observed data shows a group of non-complying agents for whom R_i equals zero but for whom the observed outcome is 1. The definition of R_i in (17) would require:

$$1 = Z_i \cdot (1 - D_i) \cdot (1 - R_i) \Rightarrow \theta_i^e > 0 \Rightarrow Y_i(1) > Y_i(0),$$

but this, in conjunction with the fact that the observed outcome is 1, would imply $Y_i(1) > 1$ which contradicts the assumption that outcomes are at most one. A more flexible tie breaking variable which equals one with some probability strictly between zero and one would prevent this scenario. Intuitively, the variable B_i captures agent heterogeneity in terms of their response to follow up questions querying their reasons for non-compliance. Some agents who perceive no benefit from treatment might subjectively feel that this was the deciding factor for not complying, while others might feel that their idiosyncratic cost shock was what pushed them to reject treatment.

The new model defined by (15) and (17) naturally captures the two types of reasons reported in Table 1. The relationship between the selection model in (15) and the average treatment effect now rests crucially on the properties of the unobserved variables (θ_i^e, C_i, B_i).

2.4 Understanding Non-Compliance Revisited

Returning to the reasons for non-compliance reported in the JTPA study, it is useful to relate the treatment selection process of (15) and (17) to what is observed in Table 1. The reported reasons for non-compliance do not explicitly fall into two categories as specified in the model. However, it can be argued that some of the reasons listed in Table 1 do indeed divide applicants into one of

two groups. For instance, consider those applicants who did not accept the offer of treatment and reported the reason "took job". Part of the assistance provided under the JTPA involved classroom instruction - the applicant may have felt that the time commitment required for training under the JTPA would not be offset by any potential gains in income. This suggests that they did not think treatment was very beneficial and therefore, this would constitute a case when $R_i = 1$ as defined in (17). Now consider an applicant who reported "changed mind". The pool of potential applicants who were part of the JTPA study all first applied for assistance under the program. Between the time they applied to the program and the time they were offered treatment, an applicant may have re-evaluated the potential benefits of the intervention. This is another example of when the model would specify that $R_i = 1$. By contrast, an applicant who reported that they did not accept treatment because of a transportation problem may not necessarily think that the program would not be beneficial. The fact that they chose to specify a transportation problem as the reason for non-compliance when they had the option to report a reason which may have indicated a lack of confidence in the benefits of the program suggests that they would have accepted treatment if they had easier access to it. The JTPA study did not explicitly ask applicant if they thought the program would not be helpful and it is therefore necessary to divide non-compliers based on hopefully reasonable arguments as in the discussion above. Future studies which do follow up with applicants would benefit from adding such an option.

2.5 Identifying Restrictions

Equipped with a description of the treatment selection process, this section examines the identifying power of the proposed model by revisiting the rational expectations assumption on how agents form their beliefs about the potential benefits of treatment θ_i^e. Throughout the following two assumptions are maintained.

Assumption IA (Independent Assignment) :*Assignment to the treatment is independent of the potential outcomes, the tie breaking variable, and an agent's information set at the time of treatment assignment i.e* $Z_i \perp (Y_i(1), Y_i(0), B_i, \mathcal{I}_i)$.

Assumption CT (Compliance Types) :*Only agents assigned to treatment can receive treatment.*

Assumption IA says that agents were randomly assigned to the treatment or control arm of the intervention, while CT rules out cases when an agent received treatment even if they were not assigned to it. In other words, CT rules out the presence of always takers or defiers (i.e. received treatment despite not being assigned to it) in the population of agents. The simple decision model in (9) specified θ_i^e as the expected treatment effect given an applicant's information at the time of being assigned to treatment. The object θ_i^e in (15) is identical and the following assumption states it's relationship with the true treatment effect.

Assumption RE (Rational Expectations) :*Agents form an estimate of the treatment effect*

at the time of assignment θ_i^e *such that,*

$$\theta_i = \theta_i^e + \nu_i,$$

where $\mathbb{E}[\nu_i \mid \mathcal{I}_i] = 0$, $\theta_i = Y_i(1) - Y_i(0)$ *and* $\mathcal{I}_i$ *is an agent's information set at the time they were assigned to treatment as defined in INA.*

The definition of θ_i^e in RE formalizes the discussion prior to defining the simple treatment selection rule in (9). The unobserved cost shock C_i now ensures that the model is not refuted by data. To see this, suppose assumptions IA, CT and RE hold and notice that the decision rule specified by (15) and (17) implies.

$$\mathbb{E}[\theta_i \mid D_i(1) - D_i(0) = 1] = \mathbb{E}[\theta_i \mid \theta_i^e \geq C_i] = \mathbb{E}[\theta_i^e \mid \theta_i^e \geq C_i] + \mathbb{E}[\nu_i \mid \theta_i^e \geq C_i], \tag{18}$$

where the final term above is zero by assumption RE and the definition of $\mathcal{I}_i$ as specified in INA. The sign of the term $\mathbb{E}[\theta_i^e \mid \theta_i^e \geq C_i]$ in (18) depends on the correlation between θ_i^e and C_i. Since no assumptions are imposed on this joint distribution, the model implied version of the LATE need not contradict the observed value θ^{LATE}. While this is encouraging, the more important implications of this setup are for non-complying agents. For simplicity, the edge case when $\theta_i^e = 0$ is ignored in what follows - this simplifies some of the expressions but does not make substantial changes to the results. Consider the average treatment effect for agents who did not accept an offer of treatment and reported a reason for non-participation which indicated that they did not think treatment would be beneficial for them.

$$\mathbb{E}[\theta_i \mid R_i = 1, D_i = 0, Z_i = 1] = \mathbb{E}[\theta_i^e \mid \theta_i^e \leq \min\{0, C_i\}] \leq 0, \tag{19}$$

where the final inequality follows because assumption RE ensures $\mathbb{E}[\nu_i \mid \theta_i] = 0$. Notice that the inequality in (19) implies,

$$\mathbb{E}[Y_i(1) \mid R_i = 1, D_i = 0, Z_i = 1] \leq \mathbb{E}[Y_i(0) \mid R_i = 1, D_i = 0, Z_i = 1] = \mathbb{E}[Y_i \mid R_i = 1, D_i = 0, Z_i = 1],$$

where the object on the right hand side may be identified from data. This is a non-trivial implication of the model and reflects the intuition that if agents have an informed notion of the treatment effect, then they should (on average) correctly report the fact that treatment would not be useful for them if indeed this is true. A similar line of reasoning can be used to establish the next model implied inequality:

$$\mathbb{E}[Y_i(1) \mid R_i = 0, D_i = 0, Z_i = 1] \geq \mathbb{E}[Y_i(0) \mid R_i = 0, D_i = 0, Z_i = 1] = \mathbb{E}[Y_i \mid R_i = 0, D_i = 0, Z_i = 1]. \tag{20}$$

Taken together, the inequalities in (20),(19) and the decomposition in (7) deliver the following upper

bound on the ATE:

$$\mathbb{P}(D_i = 1 \mid Z_i = 1) \cdot \theta^{LATE} + \\ \mathbb{P}(R_i = 0, D_i = 0 \mid Z_i = 1) \cdot (1 - \mathbb{E}[Y_i \mid R_i = 0, D_i = 0, Z_i = 1]), \tag{21}$$

where the second term utilizes the inequality in (19) to impose an upper bound of zero for the ATE of agents who report a reason such that $R_i = 1$, and uses the inequality in (20) to impose an upper bound of one for the average value of the outcome $Y_i(1)$ when $R_i = 0$. A lower bound for the ATE is given by,

$$\mathbb{P}(D_i = 1 \mid Z_i = 1) \cdot \theta^{LATE} - \\ \mathbb{P}(R_i = 1, D_i = 0 \mid Z_i = 1) \cdot \mathbb{E}[Y_i \mid R_i = 1, D_i = 0, Z_i = 1]. \tag{22}$$

Notice that the upper and lower bounds in (21) and (22) are tighter than those in (8) in the sense that the set of identified values for the ATE is now an interval of smaller length. In fact, the bounds of (21) and (22) are the best possible since an appropriate choice for the distribution of θ_i^e and C_i can lead to an exact ATE which equals the lower or upper bounds, respectively. To see this, notice that the model and assumptions invoked thus far imply the following:

$$\begin{aligned} \theta^{LATE} &= \mathbb{E}[\theta_i^e \mid \theta_i^e \geq C_i] \\ \mathbb{E}[\theta_i \mid R_i = 1, D_i = 0, Z_i = 1] &= \mathbb{E}[\theta_i^e \mid \theta_i^e \leq \min\{0, C_i\}] \\ \mathbb{E}[\theta_i \mid R_i = 0, D_i = 0, Z_i = 1] &= \mathbb{E}[\theta_i^e \mid 0 < \theta_i^e < C_i] \end{aligned} \tag{23}$$

The first equality in (23) ties the model to the identified value of θ^{LATE} through an appropriate choice of the joint distribution of the cost shock and an agent's expected benefit of treatment, conditional on the event $\theta_i^e \geq C_i$. By choosing a distribution such that $E[\theta_i^e \mid \theta_i^e \leq \min\{0, C_i\}]$ is zero and,

$$\mathbb{E}[\theta_i^e \mid 0 < \theta_i^e < C_i] = 1 - \mathbb{E}[Y_i \mid R_i = 0, D_i = 0, Z_i = 1],$$

the model implied ATE will coincide with the upper bound of (21). A data generating process such that the ATE achieves the lower bound in (22) is one which satisfies the following conditions:

$$\begin{aligned} \mathbb{P}(\theta_i^e \leq \min\{0, C_i\}) &= \mathbb{P}(R_i = 1, D_i = 1 \mid Z_i = 1), \\ \mathbb{E}[\theta_i^e \mid \theta_i^e \leq \min\{0, C_i\}] &= -\mathbb{E}[Y_i \mid R_i = 1, D_i = 0, Z_i = 1] \\ \mathbb{P}(0 < \theta_i^e < C_i) &= 0. \end{aligned}$$

This shows the identifying power of the selection model and also establishes that the bounds attainable are an improvement over having no model at all. The following proposition collects these observations.

Proposition 1: *Suppose the treatment participation rule is given by (15), and the variable R_i is*

specified as in (17). Suppose that the observed outcome lies in $\mathcal{Y} = [0, 1]$*. Under assumptions IA, CT, RE the average treatment effect must lie in an interval defined by the following lower and upper bounds:*

$$\begin{aligned}
&\mathbb{P}(D_i = 1 \mid Z_i = 1) \cdot \theta^{LATE} - \\
&\qquad \mathbb{P}(R_i = 1, D_i = 0 \mid Z_i = 1) \cdot \mathbb{E}[Y_i \mid R_i = 1, D_i = 0, Z_i = 1], \\
&\mathbb{P}(D_i = 1 \mid Z_i = 1) \cdot \theta^{LATE} + \\
&\qquad \mathbb{P}(R_i = 0, D_i = 0 \mid Z_i = 1) \cdot (1 - \mathbb{E}[Y_i \mid R_i = 0, D_i = 0, Z_i = 1]).
\end{aligned} \tag{24}$$

Moreover, the bounds in (24) are the best possible in the sense that a distribution of underlying variables may be constructed such that the resulting ATE exactly equals the lower bound, upper bound or any value in between. □

Proposition 1 and the discussion so far has assumed no covariates, but it is often straightforward to include them without substantially complicating the analysis. In particular, if data for a discrete valued covariate X_i is available, then the entire discussion follows assuming that all implications were derived conditional on $X_i = x$ for some value x. The empirical application shown below illustrates that conditioning on a discrete covariate vector is sometimes enough to gain valuable insights. Proposition 1 and the discussion leading up to it has also been somewhat vague about the role of the tie breaking variable B_i as specified in (17). The variable B_i is actually crucial when establishing the sharpness of the identified set for the ATE. The appendix also considers identification when assumption RE is strengthened to one of perfect foresight in the spirit of a standard Roy Model - in this case B_i is crucial to ensure a model which is not falsifiable.

3 Empirical Application: The Job Training Partnership Act (JTPA) Study

3.1 Study Design and Sample for Analysis

Title II-A of the Job Training Partnership Act of 1982 (JTPA) established a federally funded assistance program which provided employment and training opportunities for disadvantaged adults and out of school youths with the goal of reducing barriers to enter the labor force. In 1986, the U.S. Department of Labor commissioned a study to measure the costs and benefits of the program (hereafter referred to as the JTPA study) . The study involved applicant at 16 local JTPA programs referred to as Service Delivery Areas (SDAs) and eventually comprised around twenty thousand sample applicants. The SDAs in the study were volunteers and therefore did not represent a random sample of all SDAs in the nation. However, there was random sampling of applicants into a treatment and control group within the 16 SDAs which volunteered. During sample intake at an SDA site staff would assess if an applicant was eligible for services under the JTPA and would determine their employment and training needs. After this initial screening, applicants were randomly assigned to the treatment or control arm by study staff with 2/3 of all eligible applicants

assigned to treatment. Applicants in the treatment arm were allowed to enroll in the recommended service strategy while control group applicants were not allowed to apply for assistance under the JTPA for 18 months.

At the conclusion of the 18 month study period, applicant were contacted for a follow up survey where key outcome data was collected. This outcome data includes periods of employment, wages, and hours worked. A second follow up survey was also conducted to account for applicants who could not be reached during the first follow up. The analysis here is based on data from 11,204 applicants - it includes information for any applicant successfully reached during the follow ups, and excludes information for youths (applicants aged between 16 and 21). Additional details regarding sampling design and the follow ups may be found in Bloom et al. [1997] and Bloom et al. [1993].

3.2 Non-Compliance

The design of the JTPA study closely follows that of a randomized trial, apart from the fact that SDAs were not randomly chosen. As in many other randomized trials, not every applicant who was allowed to enroll in the JTPA actually complied with their assigned treatment status. Table 2 shows that 2,683 of the 7,487 applicants assigned to treatment did not go on to actually receive treatment.

	Trained	Not Trained	Total
Assigned to Treatment	4,804	2,683	7,487
Assigned to Control	54	3,663	3,717
Total	4,868	6,346	11,204

Table 2: Treatment assigned and treatment received status for adult applicants in the JTPA study.

The data has a small number of applicants who were assigned to control but received treatment regardless. Given the extremely small size of this sub-population (approximately 1% of all applicants who received treatment), they can be ignored with relatively little impact on the analysis. As such, assumption compliance types (CT) still applies.

3.3 Auxiliary Information in the Follow up Surveys

The follow up surveys in the JTPA study explicitly asked applicants who were assigned to treatment but did not enroll in the JTPA why they chose to do so. During follow up interviews, applicants who did not participate in JTPA training when they were assigned to the treatment group were asked to choose a reason from a pre-specified list of reasons. These non-complying applicants were asked to list three reasons in descending order of importance and were allowed to not report a reason, or indicate that the list of reasons did not contain an appropriate choice. Table 3 displays the primary reason reported for non-compliance. A significant number of applicants did not report a reason but the list of reasons is fairly exhaustive.

Reasons	Female	Male	Total
Missing	779	737	1,516
No Choice	41	21	62
Changed Mind	48	41	89
Took Job	96	108	204
Needed Job	24	28	52
Entered Other School	26	7	33
Transport Problem	19	13	32
Drug/Alcohol Problem	0	2	2
Family Problem	30	9	39
Health Problem	16	14	30
Jail	2	9	11
Other Inst.	1	2	3
Moved	19	9	28
Pregnant	12	0	12
Child Care	16	1	17
Can't afford	5	8	13
Family disallow	4	0	4
Not Chosen	80	79	159
Other	182	194	376
Total	1,401	1,282	2,683

Table 3: Primary Reason for non compliance reported across both follow up surveys.

The pattern of missing values in Table 3 needs to be accounted for since the formal analysis does not explicitly account for missing values. For the empirical application, reasons for non compliance are assumed to be missing at random. The definition of a variable being missing at random is as defined in Rubin [1976]. This does not mean that whether or not an observation is missing is completely independent of all other observable and unobservable variables. Rather, it means that while there may be systematic differences between missing and observed values of the stated reasons for non-compliance, these differences can be explained by other observed variables. For example, suppose outcomes are binary and can only take on the values 0,1 and let $O_i \in \{0, 1\}$ equal 1 if applicant $i = 1, 2, ..n$ reported a reason for non-compliance. O_i can be arbitrary for applicants who were not assigned to treatment or who accepted the offer of treatment. The assumption that reasons for non-compliance are missing at random implies:

$$\mathbb{P}(Y = 1 \mid R = 1, O = 1, D = 0, Z = 1, X) = \mathbb{P}(Y = 1 \mid R = 1, D = 0, Z = 1, X).$$

If X can only take on finitely many values, the above equality states that reasons for non-compliance are missing completely at random within each strata of the population defined by the value of X. Since the assumption that reasons for non-compliance is missing at random is not testable, it must be justified based on the particulars of the experiment.

The information in Table 3 is the basis for the variable R in the analysis.

Definition of variable R: *For applicants who reported a reason for non-compliance, R equals one if the reported reason is one of : (i) Changed Mind, (ii) Took Job, (iii) Needed Job. It is zero otherwise.*

Observations for applicants who reported jail are removed from the analysis because it is difficult to interpret in the context of the model described above. Being incarcerated for any amount of time may be thought of as an infinitely costly barrier to participation. The missing at random assumption implies all relevant finite sample quantities which involve the variable R can be estimated using only those observations for which a reported reason is available.

Note on interpretation of R: The translation of the information in Table 3 to a binary variable R involves several caveats. First of all, the list of reasons are not well documented - exactly what constitutes a difference between "took job" and "needed a job" is not known. Considering applicants who reported other reasons in the group defined by $R = 0$ may be difficult to justify. However, considering that applicants chose to not report that they changed their mind or took a job indicates that the reason for not accepting treatment may be that the applicant thought training would have been beneficial to improve their employment prospects. Of course it may be that the applicant did not think the JTPA would be helpful but none of the listed reasons exactly aligned with that assessment.

3.4 Outcomes of Interest and Empirical Results

The objective of training under the JTPA was to improve an applicant's ability to find gainful employment, and there are several outcomes of interest that may be considered. The empirical results here will consider the following outcomes: (i) whether or not an applicant had managed to secure employment within 12 months of the treatment, (ii) number of months an applicant was employed, (iii) earnings of an applicant, and (iv) maximum number of consecutive months an applicant was employed. The latter three of these outcomes were recorded over a period of 30 months post assignment to treatment and are aggregated.

Whether or not an applicant was able to find a job following treatment is a preliminary indication of program efficacy. The period immediately following treatment is when the benefits of the JTPA are the easiest to communicate to potential employers. During this time, it is more likely that an applicant will retain any information learned during classroom training received under the program. If the treatment involved help with the job search directly, then this outcome is a direct measure of whether or not applicants benefited. Table 5 presents various measures of the treatment effect. As mentioned above, the local average treatment effect is a causal object which measures the impact of treatment on the sub-population of compliers i.e.agents who would accept treatment if they were assigned to the treated group. The methods of this paper deliver bounds on the average treatment effect. These are compared against the worst case bounds and the local average treatment effect.

The worst case bounds can be derived without any selection model, and is equivalent to bounds under the selection model of (15) when the variable R_i is not observed. It requires only assumptions IA, CT and that outcomes be bounded. Bounds implied by the falsifiable model are not considered

here because this simple model precludes the possibility that $\theta_i^e < 0$. However, the JTPA study has significant observations where agents reported a reason which suggested they did not think the program would be useful. Table 4 displays some key probabilities estimated from the data. Non-compliance is a substantial issue throughout, and there is a non-trivial proportion of agents who reject the offer of treatment and report $R_i = 1$. Table 5 displays identified bounds and the LATE

Description			N	$\hat{\mathbb{P}}(D_i = 1 \mid Z_i = 1)$	$\hat{\mathbb{P}}(R_i = 1, D_i = 0 \mid Z_i = 1)$	$\hat{\mathbb{P}}(R_i = 0, D_i = 0 \mid Z_i = 1)$
Overall			10826	0.65	0.05	0.30
No income	Age < 30	Did not graduate HS	506	0.69	0.04	0.27
		Graduated HS	541	0.66	0.05	0.29
	Age ≥ 30	Did not graduate HS	656	0.58	0.06	0.36
		Graduated HS	870	0.72	0.02	0.26
Pos. income	Age < 30	Did not graduate HS	1363	0.60	0.06	0.34
		Graduated HS	2396	0.67	0.06	0.27
	Age ≥ 30	Did not graduate HS	1651	0.60	0.07	0.33
		Graduated HS	2843	0.66	0.05	0.29

Table 4: Estimated probabilities.

for the overall sample, and for subsets of the data set based on agents' pre-treatment characteristics. When conditioning on covariates, agents are divided into mutually exclusive groups based on their pre-treatment income, age and education level.

Description			N	ITT	LATE	Worst Case	Model Bounds
Overall			10826	0.01	0.02	[-0.25 , 0.10]	[-0.03 , 0.09]
				(0.00 , 0.03)	(0.00 , 0.05)	(-0.27 , 0.11)	(-0.05 , 0.10)
No income	Age < 30	Did not graduate HS	506	-0.02	-0.04	[-0.20 , 0.11]	[-0.05 , 0.09]
				(-0.11 , 0.06)	(-0.16 , 0.09)	(-0.27 , 0.18)	(-0.12 , 0.16)
		Graduated HS	541	0.12	0.18	[-0.13 , 0.20]	[0.07 , 0.20]
				(0.04 , 0.20)	(0.05 , 0.30)	(-0.21 , 0.27)	(0.00 , 0.27)
	Age ≥ 30	Did not graduate HS	656	0.04	0.07	[-0.18 , 0.24]	[0.01 , 0.22]
				(-0.04 , 0.12)	(-0.07 , 0.20)	(-0.24 , 0.30)	(-0.06 , 0.28)
		Graduated HS	870	0.05	0.07	[-0.11 , 0.17]	[0.03 , 0.17]
				(-0.02 , 0.12)	(-0.02 , 0.17)	(-0.17 , 0.23)	(-0.03 , 0.23)
Pos. income	Age < 30	Did not graduate HS	1363	0.03	0.04	[-0.31 , 0.09]	[-0.03 , 0.09]
				(-0.02 , 0.07)	(-0.03 , 0.11)	(-0.35 , 0.12)	(-0.07 , 0.12)
		Graduated HS	2396	0.01	0.01	[-0.28 , 0.05]	[-0.05 , 0.05]
				(-0.02 , 0.04)	(-0.03 , 0.05)	(-0.30 , 0.07)	(-0.07 , 0.07)
	Age ≥ 30	Did not graduate HS	1651	-0.02	-0.03	[-0.31 , 0.09]	[-0.07 , 0.07]
				(-0.05 , 0.02)	(-0.09 , 0.04)	(-0.35 , 0.11)	(-0.10 , 0.10)
		Graduated HS	2843	-0.01	-0.01	[-0.28 , 0.06]	[-0.05 , 0.06]
				(-0.04 , 0.02)	(-0.05 , 0.03)	(-0.30 , 0.08)	(-0.08 , 0.08)

Table 5: Treatment effect on whether an agent found a job within a year of treatment. Agents split into groups based on earnings in year prior to treatment, whether or not they graduated high school, and their age. (95% CIs in parentheses).

Table 5 shows an overall positive LATE - as such, the model without an unobserved cost is not rejected. None of the displayed identified sets can rule out the possibility that treatment had no impact. Conditioning on agent characteristics shows that the conditional LATE may be negative, in which case the model without cost is indeed the wrong selection process to use. More importantly, while the widest possible bounds cannot rule out no treatment effect, the selection model with cost can lead to an unambiguous treatment effect for sub-sets of the population. For example, the model implies a positive treatment effect for agents who had no earnings in the year prior to treatment and were at least thirty years of age.

Having managed to secure a job, the next aspects of employment to analyze is whether or not an applicant manages to retain their position and how much of an impact treatment had on their earnings. Job stability and wages are both important outcomes for those who are particularly economically disadvantaged, perhaps more so if an applicant has been out of the labor force for considerable amounts of time prior to approaching JTPA services. Job stability is difficult to capture in a single measure so the analysis below considers two different measures. The average number of months an applicant is employed is an obvious measure to consider. However, the number of months an applicant is employed may hide troubling patterns. Two applicants may have managed to find work for an equal number of months in the thirty month period post treatment, but one may have been forced to look for new employers more frequently than the other. This would mean that conditional on having worked for the same amount of time, one applicant did not have as much job stability as another. One way to account for job stability is to analyze the number of consecutive months an applicant was employed. Table 6 reports figures for the treatment effect on the number of months an applicant was employed post the treatment period, while Table 7 replicates the analysis for the maximum number of consecutive months an agent was employed post the treatment period.

As before, Tables 6 and 7 show that asking agents why they chose not to comply can be valuable. Though the selection model is rarely able to exclude zero from the identified set, it does occasionally identify subgroups of agents for whom the treatment is unambiguously beneficial (up to estimation error). Table 6 shows that the JTPA led to greater hours worked for high school graduates who were at most 30 years old at the time of treatment, and failed to earn any income in the twelve months prior. Table 7 suggests the same, and also shows that applicants who did not graduate high school, were at least 30 years old, and did not earn any income in the year prior to the treatment assignment decision benefited from the JTPA.

A deeper look at the total months employed and maximum consecutive months employed shows that treatment may have shifted the the distribution of outcomes - this is a more complicated pattern of change than a simple mean shift in outcomes. Figure 1 displays histograms for these outcomes by an applicant's treatment arm and, in particular, shows that the probability of an applicant failing to find employment post treatment is lower for applicants assigned to treatment. In addition, the probability that an applicant was employed for the full 30 months is higher in the treatment group. These are observations about the impact of an offer of training under the JTPA, not the impact of

Description			N	ITT	LATE	Worst Case	Model Bounds
Overall			10826	0.56	0.87	[-4.81 , 5.71]	[-0.48 , 5.11]
				(0.14 , 0.98)	(0.21 , 1.52)	(-5.19 , 6.09)	(-0.85 , 5.48)
No income	Age $<$ 30	Did not graduate HS	506	0.18	0.26	[-2.87 , 6.37]	[-0.51 , 4.36]
				(-1.64 , 1.99)	(-2.38 , 2.89)	(-4.37 , 7.86)	(-2.04 , 5.85)
		Graduated HS	541	1.97	2.96	[-2.93 , 7.12]	[1.04 , 6.37]
				(0.03 , 3.91)	(0.03 , 5.89)	(-4.62 , 8.79)	(-0.59 , 8.05)
	Age $\geq$ 30	Did not graduate HS	656	1.01	1.75	[-2.68 , 9.93]	[-0.08 , 6.58]
				(-0.63 , 2.66)	(-1.07 , 4.57)	(-4.06 , 11.29)	(-1.46 , 7.95)
		Graduated HS	870	0.23	0.32	[-2.50 , 5.91]	[-0.21 , 4.18]
				(-1.40 , 1.86)	(-1.94 , 2.59)	(-3.87 , 7.28)	(-1.58 , 5.55)
Pos. income	Age $<$ 30	Did not graduate HS	1363	0.62	1.03	[-5.7 , 6.21]	[-0.55 , 5.78]
				(-0.58 , 1.81)	(-0.95 , 3.01)	(-6.77 , 7.27)	(-1.58 , 6.82)
		Graduated HS	2396	0.39	0.58	[-5.67 , 4.15]	[-0.75 , 4.50]
				(-0.45 , 1.23)	(-0.66 , 1.83)	(-6.44 , 4.92)	(-1.48 , 5.25)
	Age $\geq$ 30	Did not graduate HS	1651	0.86	1.42	[-5.03 , 6.87]	[-0.48 , 5.87]
				(-0.20 , 1.91)	(-0.32 , 3.17)	(-5.99 , 7.82)	(-1.37 , 6.83)
		Graduated HS	2843	0.09	0.13	[-5.63 , 4.54]	[-0.89 , 4.50]
				(-0.73 , 0.90)	(-1.10 , 1.36)	(-6.35 , 5.26)	(-1.61 , 5.20)

Table 6: Treatment effect on number of months worked. Agents split into groups based on earnings in year prior to treatment, whether or not they graduated high school, and their age. (95% CIs in parentheses)

treatment itself. The selection model can be used to bound the treatment effect on whether or not an applicant will be unemployed (or continuously employed) following the treatment period.

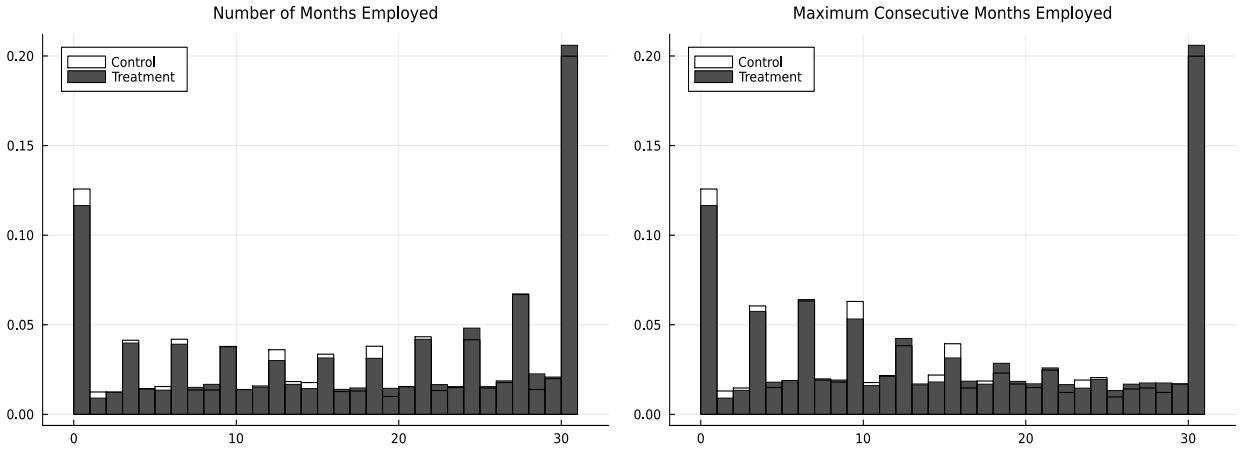

Figure 1: Histograms for number of months employed (left) and maximum consecutive months employed (right) by whether or not applicant was assigned to treatment.

To analyze the impact of treatment on the entire distribution of outcomes, the selection model can be used to derived bounds on the shift in the distribution function of an outcome. Fix any

Description			N	ITT	LATE	Worst Case	Model Bounds
Overall			10826	0.55	0.85	[-4.21 , 6.31]	[-0.40 , 5.62]
				(0.13 , 0.97)	(0.20 , 1.50)	(-4.58 , 6.69)	(-0.76 , 5.98)
No income	Age < 30	Did not graduate HS	506	0.64	0.93	[-1.96 , 7.28]	[0.01 , 5.30]
				(-1.00 , 2.28)	(-1.45 , 3.31)	(-3.31 , 8.62)	(-1.37 , 6.63)
		Graduated HS	541	1.66	2.49	[-2.63 , 7.43]	[0.82 , 6.56]
				(-0.25 , 3.56)	(-0.38 , 5.36)	(-4.23 , 9.02)	(-0.77 , 8.17)
	Age ≥ 30	Did not graduate HS	656	1.20	2.07	[-2.01 , 10.59]	[0.20 , 7.41]
				(-0.32 , 2.72)	(-0.52 , 4.67)	(-3.30 , 11.86)	(-1.07 , 8.68)
		Graduated HS	870	0.04	0.05	[-2.37 , 6.04]	[-0.37 , 4.43]
				(-1.59 , 1.67)	(-2.21 , 2.31)	(-3.74 , 7.40)	(-1.73 , 5.80)
Pos. income	Age < 30	Did not graduate HS	1363	0.48	0.79	[-4.91 , 7.00]	[-0.59 , 6.22]
				(-0.70 , 1.65)	(-1.16 , 2.74)	(-5.95 , 8.03)	(-1.60 , 7.24)
		Graduated HS	2396	0.32	0.48	[-5.08 , 4.74]	[-0.71 , 4.90]
				(-0.55 , 1.20)	(-0.82 , 1.78)	(-5.86 , 5.51)	(-1.47 , 5.66)
	Age ≥ 30	Did not graduate HS	1651	0.93	1.54	[-4.26 , 7.64]	[-0.29 , 6.51]
				(-0.11 , 1.97)	(-0.18 , 3.25)	(-5.19 , 8.56)	(-1.16 , 7.44)
		Graduated HS	2843	0.14	0.21	[-5.06 , 5.11]	[-0.76 , 5.05]
				(-0.70 , 0.97)	(-1.05 , 1.47)	(-5.78 , 5.82)	(-1.48 , 5.76)

Table 7: Treatment effect on maximum number of consecutive months agent was employed. Agents split into groups based on earnings in year prior to treatment, whether or not they graduated high school, and their age. (95% CIs in parentheses)

value $y \in \{0, 1, ..., 30\}$ and consider the following difference,

$$\mathbb{P}(Y_i(1) \leq y) - \mathbb{P}(Y_i(0) \leq y).$$

If this difference is negative, then it implies that the distribution of $Y_i(1)$ is "shifted" to the right, relative to the distribution of $Y_i(0)$ i.e. the probability that $Y_i(1)$ is higher than y is greater than the probability that $Y_i(0)$ is higher than y. Taken together, this means that the mean value of $Y_i(1)$ is greater than that of $Y_i(0)$ i.e. the ATE is positive. Figure 2 displays the distribution treatment effect on these outcomes and shows that treatment may not have had a positive impact on either outcome from Tables 6 and 7. The local average treatment effect implies a rightward shift in the distribution of $Y_i(1)$ for complying agents, which seems to align with the observations from Figure 1 but bounds on the treatment effect do not rule out the possibility of a negative distributional treatment effect. The appendix replicates these figures for sub-groups of applicants.

The final measure of program efficacy is ex-post earnings. Table 8 displays results for this outcome. The scale of this outcome is somewhat different relative to those considered above. Unlike previous outcomes, there is no known upper limit ot how much an applicant can earn so an appropriate limit must be chosen by the researcher. Since there are applicants who reported considerably higher amounts than the mean or median income earnings, the upper bound of the identified sets can be quite high relative to the lower bound. This is because the upper limit for earnings affects the upper limit of the treatment effect for non-complying agents, and an appropriate upper bound is

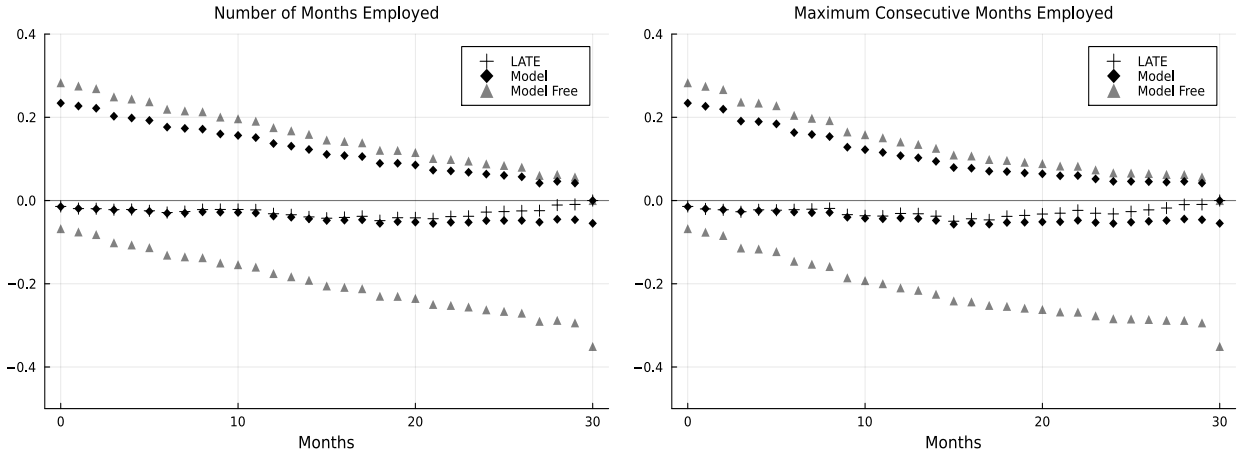

Figure 2: Distributional treatment effect for applicants who had graduated high school, were aged less than 30, and had no prior income during the year prior to assignment.

necessarily one which exceeds the maximum observed earnings. The simple decision model without C_i has much lower upper bounds because the average treatment effect for non complying agents is at most zero under this model. However, this simple model is also seemingly rejected by the data since it implies a positive local average treatment effect which turn is contradicted by the negative value observed for some sub-groups of agents.

Interestingly, Table 8 shows that the selection model with cost implies a positive lower bound for the average treatment effect for the overall sample of agents. This is not so for the other outcomes considered above. Indeed the ATE is estimated to be positive for quite a few sub-sets of agents.

The treatment effect of training on the distribution of a continuously distributed variable such as earnings can be analyzed in greater detail than is presented in Table 8 which only provides information about the mean impact of training. For example, let E_i denote the observed earnings for agent i with corresponding potential outcomes $E_i(1)$, $E_i(0)$ which are assumed to lie in $[0, \overline{y}]$. Consider the following binary variables,

$$Y_i(1) = \mathbb{1}\{E_i(1) \leq y\}, \;\; Y_i(0) = \mathbb{1}\{E_i(0) \leq y\} \quad \text{for } y \in [0, \overline{y}].$$

The binary variable $Y_i = Y_i(1) \cdot D_i + Y_i(0) \cdot (1 - D_i)$ now indicates if the observed earnings for an agent is less than the value y. The analysis of the previous sections can now be applied to measure the impact of training under the JTPA on the entire distribution of earnings. For example, given a value $y \in [0, \overline{y}]$, the average treatment effect for outcome Y_i may be defined as,

$$ATE = \mathbb{P}(Y_i(1) \leq y) - \mathbb{P}(Y_i(0) \leq y).$$

A distribution level treatment effect is of course more informative than a mean impact. Indeed if the ATE as defined above is negative for all possible values $y \in [0, \overline{y}]$ then the mean treatment effect must be positive. Intuitively, a negative treatment effect on the distribution of outcomes means

Description			N	ITT	LATE	Worst Case	Model Bounds
Overall			10826	1.13	1.74	[-3.58 , 24.47]	[0.09 , 21.16]
				(0.53 , 1.73)	(0.82 , 2.66)	(-4.09 , 24.97)	(-0.42 , 21.66)
No income	Age < 30	Did not graduate HS	506	1.08	1.56	[-1.04 , 23.61]	[0.97 , 20.82]
				(-1.08 , 3.23)	(-1.55 , 4.67)	(-2.81 , 25.37)	(-0.84 , 22.58)
		Graduated HS	541	2.07	3.11	[-1.49 , 25.25]	[1.35 , 22.06]
				(-0.45 , 4.59)	(-0.68 , 6.9)	(-3.63 , 27.37)	(-0.76 , 24.2)
	Age ≥ 30	Did not graduate HS	656	1.11	1.92	[-1.41 , 32.37]	[0.57 , 28.25]
				(-0.75 , 2.97)	(-1.29 , 5.12)	(-2.93 , 33.88)	(-0.98 , 29.77)
		Graduated HS	870	-0.37	-0.52	[-2.26 , 20.13]	[-0.67 , 18.54]
				(-2.18 , 1.44)	(-3.03 , 1.99)	(-3.78 , 21.65)	(-2.18 , 20.06)
Pos. income	Age < 30	Did not graduate HS	1363	1.92	3.18	[-3.29 , 28.45]	[0.81 , 24.64]
				(0.27 , 3.57)	(0.43 , 5.93)	(-4.62 , 29.78)	(-0.59 , 25.96)
		Graduated HS	2396	1.16	1.72	[-4.54 , 21.54]	[-0.16 , 18.08]
				(-0.16 , 2.47)	(-0.23 , 3.67)	(-5.64 , 22.63)	(-1.27 , 19.17)
	Age ≥ 30	Did not graduate HS	1651	1.15	1.91	[-3.69 , 27.96]	[-0.07 , 23.68]
				(-0.32 , 2.63)	(-0.52 , 4.34)	(-4.93 , 29.19)	(-1.28 , 24.94)
		Graduated HS	2843	0.79	1.20	[-4.79 , 22.42]	[-0.44 , 19.56]
				(-0.47 , 2.05)	(-0.71 , 3.10)	(-5.85 , 23.47)	(-1.51 , 20.61)

Table 8: Treatment effect on earnings (in '000s) post treatment. Agents split into groups based on earnings in year prior to treatment, whether or not they graduated high school, and their age. (95% CIs in parentheses)

the distribution function of the outcome under treatment is pushed out to the right relative to the distribution function for the outcome without treatment. This means the probability of attaining a higher outcome is greater under treatment. Figure 3 displays upper and lower bounds for the distribution of earnings under treatment for the full sample of individuals. The figure plots the distribution of earnings for agents in the control group and compares it against the distribution of earnings under treatment. Since assignment to treatment was independent in the JTPA study (i.e. assumption (IA)) the distribution of the observed earnings of agents in the control group is equivalent to the distribution of potential earnings without treatment.

Figure 3 shows that the model significantly tightens the upper bound for the treatment effect. This ties in with the results of Table 8. The LATE for mean earnings is positive at all points of the distribution but the upper and lower bounds in Figure 3 show that this can be misleading. Surprisingly, it appears that the treatment effect is more pronounced and positive at higher ends of the income distribution implying that greater earners possibly benefited more from training. This relationship is potentially obscuring a number of factors. For example, agents with greater education pre-treatment are more likely to earn more and also stand to gain the most from classroom training under the JTPA due to their greater educational attainment. A combination of these features might manifest as a positive lower bound for the treatment effect on the distribution of income at the upper end of the distribution.

Deriving distribution treatment effects on a subsets of the study participants shows that training

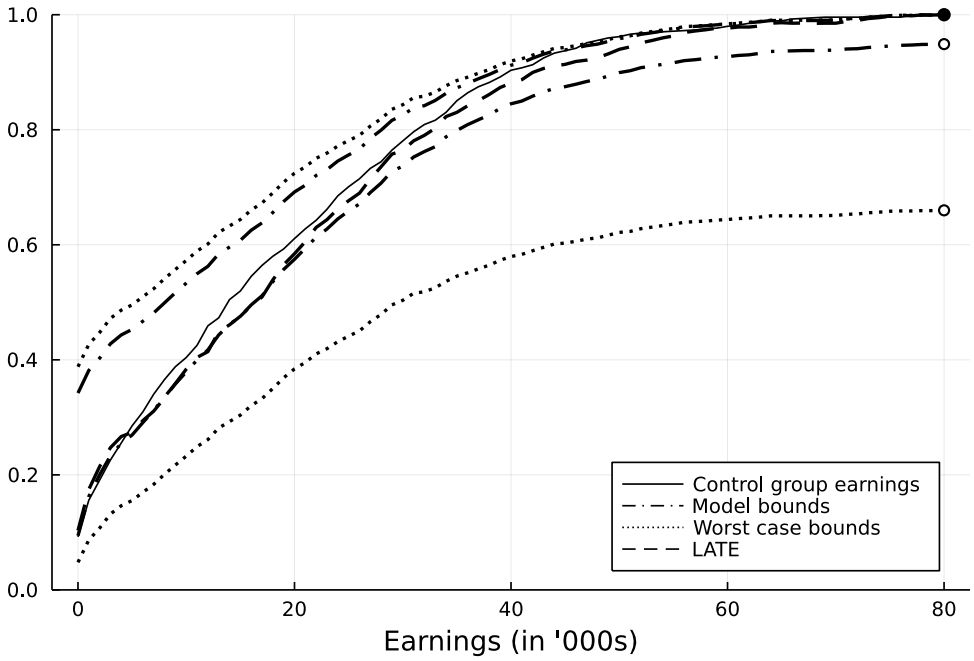

Figure 3: Treatment effect on the distribution of earnings (in '000s).

impacts agents differently based on their pre-treatment characteristics - this conforms with the results in Table 8 For example, Table 8 identifies a positive treatment effect on earnings for agents who were aged at least thirty at the time of assignment but did not graduate high school, as opposed to if they did graduate high school. Figure 4 shows that these mean impacts aggregate over significant differences.

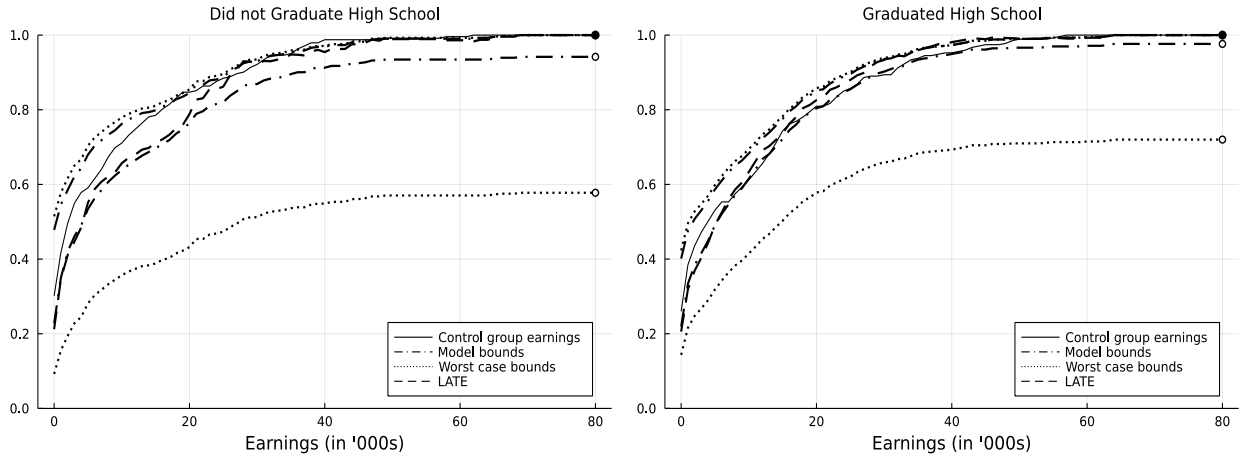

Figure 4: Treatment effect of training under JTPA on the distribution of earnings (in '000s) for agents who were aged at least 30 years, and reported no income in the year prior to time assignment.

4 Extension: Identification with Parametric Assumptions on Cost

No assumptions have been made on the distribution of the unobserved cost variable C_i and in this section I derive additional restrictions on the identified set for the ATE if such assumptions are made. Suppose the distribution of the unobserved C_i is assumed to be known up to a finite dimensional parameter $\mu \in \mathbb{R}^c$. For the sake of simplicity, the discussion here will assume that cost is uniformly distributed $C_i \sim U[l, u]$ for fixed, finite values $l, u \in \mathbb{R}$. It is assumed that the interval $[l, u]$ contains at least all possible values of the difference in potential outcomes $Y_i(1) - Y_i(0)$. If G denotes the distribution function of cost, then notice that the following functions are both convex,

$$\frac{G(x)}{1-G(x)} = \frac{x-l}{u-x}, \quad \frac{1-G(x)}{G(x)} = \frac{u-x}{x-l},$$

for values $x \in (l, u)$. This feature is a key property which will deliver useful identifying restrictions from the treatment selection model of (15). Many studies have exploited such convexity properties to derive identifying restrictions - notably Dickstein and Morales [2018] employ such assumptions to tackle a similar problem involving a firms decision to enter a foreign market. To illustrate the argument, suppose assumptions IA, CT and RE hold and that the cost shock is independent of all other model variables. Consider the probability that an agent assigned to treatment actually accepts the offer. The decision rule of (15) requires that agents accept only if the net expected benefit from treatment at the time of assignment is non-negative,

$$\mathbb{P}(D_i = 1 \mid Z_i = 1) = \mathbb{P}(C_i \leq \theta_i^e) = \mathbb{E}\left[\frac{\theta_i^e - l}{u-l}\right] = \mathbb{E}[G(\theta_i^e)] \tag{25}$$

where the final expectations operator above averages over the distribution of agent information sets $\mathcal{I}_i$. Notice the following,

$$\mathbb{E}\left[\frac{\theta_i^e - l}{u-\theta_i^e}\right] \geq \mathbb{E}\left[\frac{\theta_i^e - l}{u-\theta_i^e} \cdot (1-D_i)\right] = \mathbb{E}\left[\frac{G(\theta_i^e)}{1-G(\theta_i^e)} \cdot \mathbb{E}\left[1-D_i \mid \mathcal{I}_i\right]\right] = \mathbb{E}[G(\theta_i^e)]. \tag{26}$$

The final equality above follows from the model specification of D_i. This observation establishes the following link between the model implied value of the probability of complying with assignment to treatment, and its identified value,

$$\mathbb{P}(D_i = 1 \mid Z_i = 1) \leq \mathbb{E}\left[\frac{G(\theta_i^e)}{1-G(\theta_i^e)}\right] = \mathbb{E}\left[\frac{\theta_i^e - l}{u-\theta_i^e}\right].$$

This inequality connects agents' subjective expectation of the possible treatment effect with the observed probability of an agent accepting the offer of treatment. However, it does not directly connect the distribution of potential outcomes $(Y_i(1), Y_i(0))$ with an identified quantity. This is where the convexity of the function $x \mapsto G(x)/(1-G(x))$ is useful. Since the object θ_i^e is defined

as a (conditional) mean value in assumption RE, an application of Jensen's inequality establishes,

$$\mathbb{P}(D_i = 1 \mid Z_i = 1) \leq \mathbb{E}\left[\frac{\mathbb{E}[Y_i(1) - Y_i(0) \mid \mathcal{I}_i] - l}{u - \mathbb{E}[Y_i(1) - Y_i(0) \mid \mathcal{I}_i]}\right] \leq \mathbb{E}\left[\frac{Y_i(1) - Y_i(0) - l}{u - Y_i(1) + Y_i(0)}\right], \tag{27}$$

where the final object above involves only the distribution of potential outcomes, and the known parameters u, l. This follows regardless of what lies in an agent's information set $\mathcal{I}_i$. Different agents may have distinct sets of private information which they use when forming their expected treatment effect θ_i^e as defined in RE. A similar argument can be used to establish,

$$\mathbb{P}(D_i = 0 \mid Z_i = 1) \leq \mathbb{E}\left[\frac{u - Y_i(1) + Y_i(0)}{Y_i(1) - Y_i(0) - l}\right], \tag{28}$$

by relying on the convexity of the function $x \mapsto (1 - G(x))/G(x)$. These inequalities provide binding restrictions on the distribution of potential outcomes. For example, any joint distribution of potential outcomes conformable with the observed data must satisfy the restrictions in (27) and (28), and any such valid distribution leads to a candidate identified value for the ATE.

The discussion above is encouraging in that it describes a potential method by which the unobserved joint distribution of potential outcomes can be related to identified quantities. While the assumptions on the distribution of cost is particularly strong, the specific requirement that cost is uniformly distributed can be weakened as long as the chosen distribution has the necessary convexity properties. In principle, this class of possible distributions for C_i is very large, but empirical tractability will often restrict the choice of distribution. Outcome information is also not used in inequalities (27), (28) but this is for ease of exposition. The complete list of model related restrictions on the joint distribution of potential outcomes presented below does use data for observed outcomes.

Unlike the discussion involving rational expectations (RE), the inequalities in (27) and (28) do not directly involve the ATE. As a consequence, the identified set for the average treatment effect does not have an explicit upper and lower bound as in proposition 1. Instead, it is possible to derive a complete list of model implied restrictions which implicitly define a set of joint distributions for the potential outcomes which are conformable with data. Each member of this implicitly defined set may be used to define an identified value for the ATE. Before stating this identification result, the following assumption first states the exact conditions on a candidate class for the distribution of the unobserved cost term.

Assumption Independent Cost (IC) : *The cost variable C_i is independent of all other variables, has support $\mathcal{C}$ and its distribution is characterized by a distribution function G such that the following are convex,*

$$x \mapsto \frac{G(x)}{1 - G(x)}, \quad x \mapsto \frac{1 - G(x)}{G(x)},$$

for any $x \in Int(\mathcal{C})$.

An immediate consequence of the convexity requirements in assumption IC is that the cost variable must be continuously distributed. Such a restriction wasn't necessary under the rational

expectations (RE) framework. IC also meaningfully restricts the class of admissible distributions beyond just requiring them to be continuous. For example, not all parameterizations of a Beta distribution satisfy the convexity requirements of IC (e.g. the distribution function of a variable $X \sim$Beta$(0.1, 0.1)$ will violate IC). Nevertheless, many commonly used parametric families do satisfy this assumption - for example, the Normal, Cauchy and uniform distributions all possess the necessary convexity properties.

The following proposition describes conditions on the joint distribution of potential outcomes $(Y_i(1), Y_i(0))$ which must be satisfied when the cost variable is assumed to follow a distribution G which possesses the convexity requirements in IC.

Proposition 2: *Suppose IA, CT, RE, IC hold and potenial outcomes lie within* $\mathcal{Y} = [0, 1]$. *The following relations must then be true for any* $y \in [0, 1]$*:*

$$\mathbb{P}\left(Y_i(1) \leq y\right)^{1/2} \cdot \mathbb{E}\left[\left(\frac{G\left(Y_i(1) - Y_i(0)\right)}{1 - G\left(Y_i(1) - Y_i(0)\right)}\right)^2\right]^{1/2} \geq \mathbb{P}\left(Y_i \leq y, D_i = 1 \mid Z_i = 1\right),$$

$$\mathbb{P}\left(Y_i(0) \leq y\right)^{1/2} \cdot \mathbb{E}\left[\left(\frac{1 - G\left(Y_i(1) - Y_i(0)\right)}{G\left(Y_i(1) - Y_i(0)\right)}\right)^2\right]^{1/2} \geq \mathbb{P}\left(Y_i \leq y, D_i = 0 \mid Z_i = 1\right),$$

$$\mathbb{E}\left[\frac{G\left(Y_i(1) - Y_i(0)\right)}{1 - G\left(Y_i(1) - Y_i(0)\right)}\right] \geq \mathbb{P}(D_i = 1 \mid Z_i = 1),$$

$$\mathbb{E}\left[\frac{1 - G\left(Y_i(1) - Y_i(0)\right)}{G\left(Y_i(1) - Y_i(0)\right)}\right] \geq \mathbb{P}(D_i = 0 \mid Z_i = 1),$$

$$\mathbb{P}(Y_i(0) \leq y) = \mathbb{P}(Y_i \leq y \mid Z_i = 0). \tag{29}$$

Moreover, the average treatment effect implied by any joint distribution of potential outcomes $(Y_i(1), Y_i(0))$ *which satisfies the relations in(29) must lie within the upper and lower bounds defined in display (24) of Proposition 1.* □

The conditions in (29) all relate a model implied quantity to an identified value. Proposition 2 defines restrictions on the joint distribution of the potential outcomes but does not directly define an identified set for the ATE. However, as mentioned above, the average treatment is straightforward to derive given any joint distribution of potential outcomes which satisfy the conditions in (29). There are more substantive disadvantages inherent in this approach when compared to the results in Proposition 1.

Leaving aside the practical question of how an appropriate class of distributions for the cost variable C_i may be chosen, the relations in (29) make no use of the variable R_i though information in R_i is embedded in the requirement that the identified region for the ATE in Proposition 2 must be at most that defined in Proposition 1. The identified set in Proposition 2 may be difficult to calculate in finite sample since the restrictions are on the joint distribution of potential outcomes directly instead of on just the implied conditional means of outcomes. This issue is mitigated somewhat when outcomes are discrete. For example, Table 5 pertains to the binary outcome for whether or

not an agent managed to find employment within a year of treatment. The joint distribution of potential outcomes in this case can be described by a vector of four non-negative numbers which add up to one and satisfy the restrictions of Proposition 2. The ATE with binary outcomes is given by:

$$\begin{aligned} ATE =& \mathbb{P}(Y_i(1) = 1) - \mathbb{P}(Y_i(0) = 1) \\ =& \mathbb{P}(Y_i(1) = 1, Y_i(0) = 0) - \mathbb{P}(Y_i(1) = 0, Y_i(0) = 1). \end{aligned}$$

While the identified set for the average treatment effect defined in Proposition 2 must be smaller than the same from Proposition 1 the extent of this shrinking will depend on the exact parametric specification on the cost variable C_i. As an example, Figure 5 displays lower/upper bounds for the ATE when C_i is assumed to follow a $\mathcal{N}(\mu_c, \sigma^2)$ distribution. The flat portions of the dashed line correspond to the bounds from Proposition 1.

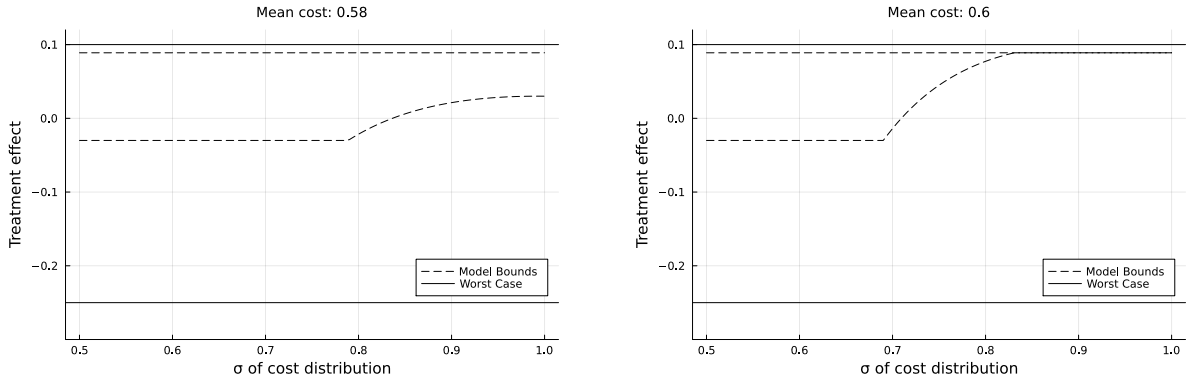

Figure 5: Estimated bounds for treatment effect on whether or not agent found employment within 12 months post completion of treatment phase when $C_i \sim \mathcal{N}(\mu_c, \sigma^2)$.

While the model significantly increases the lower bound for the ATE (as shown in Table 5), only some values of the parameter vector (μ_c, σ^2) which governs the distribution of C_i lead to a shrinking of the identified set. This suggests that even if the researcher is willing to impose strong parametric assumptions on C_i, the gain may not be substantial. For instance, suppose we were willing to accept that indeed $C_i \sim \mathcal{N}(\mu_c, \sigma^2)$ but allowed the mean and variance parameters to lie in some pre-specified set of admissible values instead of fixing them to a particular value. If this set were sufficiently large, a lower bound for the ATE conformable with the characterization in Proposition 2 and the set of admissible parameters would coincide with the bounds of Proposition 1. Figure 6 displays the estimated joint distribution of potential outcomes which lead to the ATE values in Figure 5.

Figure 5 shows that certain parameter values for the distribution of C_i will lead to point identification of the ATE. This is driven entirely by assumptions on the cost variable and not by additional assumptions on the underlying compliance behavior of agents. The non-differentiable nature of the estimated lower bound in Figure 5 may also pose additional challenges for inference. Taken

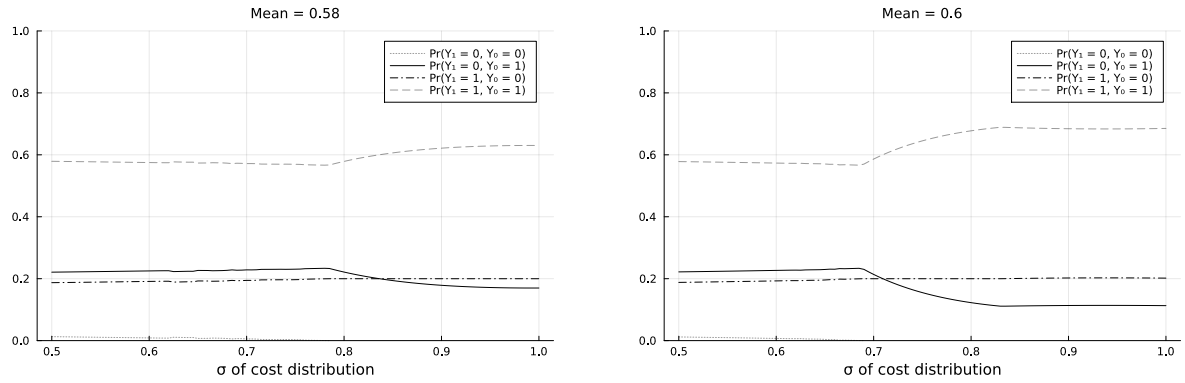

Figure 6: Estimated bounds for treatment effect on whether or not agent found employment within 12 months post completion of treatment phase when $C_i \sim \mathcal{N}(\mu_c, \sigma^2)$.

together, this discussion suggests that relying on parametric assumptions to tighten the identified set for the ATE on an outcome may not be ideal.

5 Conclusion

In this paper, I propose a decision model rationalizing an agent's compliance behavior. The model is fairly parsimonious and leads to a characterization of the identified set for the ATE that is intuitive and may be estimated in finite samples using existing techniques. The empirical application highlights the significant value of information collected from non-complying agents - this highlights the importance of follow up surveys which explicitly ask agents why they refused to adhere to their assigned treatment. It is sometimes argued that analyzing the identified set for the ATE when there is non-compliance does not necessarily suggest a useful policy recommendation. This reasoning is driven by the observation that the identified set for the ATE is often wide enough to contain both positive and negative values, rendering the impact of treatment ambiguous even if other causal parameters such as the LATE are significantly different from zero. Indeed the model proposed here does not always lead to a clear answer for whether or not treatment was beneficial for the applicant. Nevertheless, the results of Tables 5, 6 and 8 show that it is possible to identify sub-groups of agents for whom the intervention led to an increase in average outcome. This is true even though the bounds derived without the model were much wider, once again emphasizing the usefulness of follow up surveys.

References

A. Abadie and M. Cattaneo. Econometric Methods for Program Evaluation. *Annual Review of Economics*, 10(1):465–503, 2018. URL https://EconPapers.repec.org/RePEc:anr:reveco:v:10:y:2018:p:465-503.

A. Abadie, J. Angrist, and G. Imbens. Instrumental Variables Estimates of the Effect of Subsidized

Training on the Quantiles of Trainee Earnings. *Econometrica*, 70(1):91–117, January 2002. URL https://ideas.repec.org/a/ecm/emetrp/v70y2002i1p91-117.html.

J. D. Angrist and G. W. Imbens. Sources of Identifying Information in Evaluation Models. NBER Technical Working Papers 0117, National Bureau of Economic Research, Inc, Dec. 1991. URL https://ideas.repec.org/p/nbr/nberte/0117.html.

A. Balke and J. Pearl. Bounds on Treatment Effects from Studies with Imperfect Compliance. *Journal of the American Statistical Association*, 92(439):1171–1176, 1997. doi: 10.1080/01621459.1997.10474074. URL https://doi.org/10.1080/01621459.1997.10474074.

J. Bhattacharya, A. Shaikh, and E. Vytlacil. Treatment Effect Bounds under Monotonicity Assumptions: An Application to Swan-Ganz Catheterization. *American Economic Review*, 98(2):351–56, 2008. URL https://EconPapers.repec.org/RePEc:aea:aecrev:v:98:y:2008:i:2:p:351-56.

H. Bloom, L. Orr, S. Bell, G. Cave, and Doolittle. The National JTPA Study: Title II-A Impacts on Earnings and Employment at 18 Months. *Bethesda, Md.: Abt Associates Inc*, 1993.

H. Bloom, L. Orr, S. Bell, G. Cave, F. Doolittle, W. Lin, and J. Bos. The Benefits and Costs of JTPA Title II-A Programs: Key findings from the National Job Training Partnership Act study. *Journal of Human Resources*, 32:549–576, 06 1997. doi: 10.2307/146183.

X. Chen, C. A. Flores, and A. Flores-Lagunes. Bounds on Population Average Treatment Effects with an Instrumental Variable. 2012.

P. Courty and G. Marschke. An Empirical Investigation of Gaming Responses to Explicit Performance Measures. *Journal of Labor Economics*, 22:23–56, 01 2004. doi: 10.2139/ssrn.131328.

T. Demuynck. Bounding Average Treatment Effects: A Linear Programming Approach. *Economics Letters*, 137(C):75–77, 2015. doi: 10.1016/j.econlet.2015.09. URL https://ideas.repec.org/a/eee/ecolet/v137y2015icp75-77.html.

M. J. Dickstein and E. Morales. What do Exporters Know? *The Quarterly Journal of Economics*, 133(4):1753–1801, 2018.

J. Dominitz and C. F. Manski. Using Expectations Data To Study Subjective Income Expectations. *Journal of the American Statistical Association*, 92(439):855–867, 1997. ISSN 01621459. URL http://www.jstor.org/stable/2965550.

R. A. Fisher. *The Design of Experiments*. Oliver and Boyd, Edinburgh, 1935.

J. Heckman. Shadow Prices, Market Wages, and Labor Supply. *Econometrica*, 42:679–94, 02 1974. doi: 10.2307/1913937.

J. Heckman. The Common Structure of Statistical Models of Truncation, Sample Selection and Limited Dependent Variables and a Simple Estimator for Such Models. In *Annals of Economic*

and Social Measurement, Volume 5, number 4, pages 475–492. National Bureau of Economic Research, Inc, 1976. URL https://EconPapers.repec.org/RePEc:nbr:nberch:10491.

J. Heckman and B. E. Honoré. The Empirical Content of the Roy Model. *Econometrica*, 58(5): 1121–49, 1990. URL https://EconPapers.repec.org/RePEc:ecm:emetrp:v:58:y:1990:i:5:p:1121-49.

J. J. Heckman and V. J. 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):862–874, 1989. ISSN 01621459. URL http://www.jstor.org/stable/2290059.

J. J. Heckman and J. A. Smith. The Sensitivity of Experimental Impact Estimates: Evidence from the National JTPA Study. Working Paper 6105, National Bureau of Economic Research, July 1997. URL http://www.nber.org/papers/w6105.

J. J. Heckman and E. J. Vytlacil. Local Instrumental Variables and Latent Variable Models for Identifying and Bounding Treatment Effects. *Proceedings of the National Academy of Sciences*, 96(8):4730–4734, 1999. doi: 10.1073/pnas.96.8.4730. URL https://www.pnas.org/doi/abs/10.1073/pnas.96.8.4730.

C. Hewitt, D. Torgerson, and J. Miles. Is There Another Way to Take Account of Noncompliance in Randomized Controlled Trials? *CMAJ : Canadian Medical Association Journal = Journal de l'Association Medicale Canadienne*, 175:347, 09 2006.

G. Imbens and J. Angrist. Identification and Estimation of Local Average Treatment Effects. *Econometrica*, 62:467–75, 02 1994. doi: 10.2307/2951620.

G. Imbens and D. Rubin. Bayesian Inference for Causal Effects in Randomized Experiments with Noncompliance. *Ann. Statist.*, 25, 02 1997. doi: 10.1214/aos/1034276631.

G. W. Imbens. Nonparametric Estimation of Average Treatment Effects Under Exogeneity: A Review. *The Review of Economics and Statistics*, 86(1):4–29, February 2004. URL https://ideas.repec.org/a/tpr/restat/v86y2004i1p4-29.html.

G. W. Imbens and C. F. Manski. Confidence Intervals for Partially Identified Parameters. *Econometrica*, 72(6):1845–1857, November 2004. URL https://ideas.repec.org/a/ecm/emetrp/v72y2004i6p1845-1857.html.

G. W. Imbens and D. B. Rubin. *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction*. Cambridge University Press, 2015. doi: 10.1017/CBO9781139025751.

T. Kitagawa and A. Tetenov. Who Should Be Treated? Empirical Welfare Maximization Methods for Treatment Choice. *Econometrica*, 86(2):591–616, March 2018. doi: 10.3982/ECTA13288. URL https://ideas.repec.org/a/wly/emetrp/v86y2018i2p591-616.html.

T. Kitagawa and A. Tetenov. Equality-Minded Treatment Choice. *Journal of Business & Economic Statistics*, 39(2):561–574, March 2021. doi: 10.1080/07350015.2019.168. URL https://ideas.repec.org/a/taf/jnlbes/v39y2021i2p561-574.html.

F. Kleinsinger. Understanding Noncompliant Behavior: Definitions and Causes. *The Permanente Journal*, 7(4):18, 2003.

F. Kleinsinger. Working with the Noncompliant Patient. *The Permanente Journal*, 14 1:54–60, 2010.

C. Manski. *Partial Identification of Probability Distributions: Springer Series in Statistics*. Springer, 2003. ISBN 9780387004549.

C. F. Manski. Nonparametric Bounds on Treatment Effects. *The American Economic Review*, 80 (2):319–323, 1990. ISSN 00028282. URL http://www.jstor.org/stable/2006592.

F. Molinari. Econometrics with Partial Identification. CeMMAP working papers CWP25/19, Centre for Microdata Methods and Practice, Institute for Fiscal Studies, May 2019. URL https://ideas.repec.org/p/ifs/cemmap/25-19.html.

I. Mourifié, M. Henry, and R. Méango. Sharp Bounds and Testability of a Roy Model of STEM Major Choices. *Journal of Political Economy*, 128:3220 – 3283, 2020.

A. D. Roy. Some Thoughts on the Distribution of Earnings. *Oxford Economic Papers*, 3(2): 135–146, 1951. URL https://EconPapers.repec.org/RePEc:oup:oxecpp:v:3:y:1951:i:2:p:135-146.

D. Rubin. Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies. *Journal of Educational Psychology*, 66(5):688–701, 1974.

D. B. Rubin. Inference and Missing Data. *Biometrika*, 63(3):581–592, 1976. ISSN 00063444. URL http://www.jstor.org/stable/2335739.

J. A. Smith, A. Whalley, and N. T. Wilcox. Are Program Participants Good Evaluators? IZA Discussion Papers 13584, Institute of Labor Economics (IZA), Aug. 2020. URL https://ideas.repec.org/p/iza/izadps/dp13584.html.

J. Stoye. More on Confidence Intervals for Partially Identified Parameters. *Econometrica*, 77(4):1299–1315, July 2009. URL https://ideas.repec.org/a/ecm/emetrp/v77y2009i4p1299-1315.html.

E. Tamer. Partial Identification in Econometrics. *Annual Review of Economics*, 2(1):167–195, 2010. doi: 10.1146/annurev.economics.050708.143401. URL https://doi.org/10.1146/annurev.economics.050708.143401.

A Proof of Proposition 1

Define the following events,

$$R_1 = \left[\theta_i^e < C_i;\ \theta_i^e < 0\right] \cup \left[0 = \theta_i^e < C_i;\ B_i = 1\right] \tag{30}$$

$$R_0 = \left[0 < \theta_i^e < C_i\right] \cup \left[0 = \theta_i^e < C_i;\ B_i = 0\right], \tag{31}$$

where B_i denotes the tie breaking random variable. Similar to (19) the following is true,

$$\begin{aligned}\mathbb{E}[\theta_i \mid R_i = 1, D_i = 0, Z_i = 1] &= \mathbb{E}[\theta_i^e + \nu_i \mid R_1] \\ &= \mathbb{E}[\theta_i^e \mid R_1] + \mathbb{E}[\nu_i \mid R_1] \\ &= \mathbb{E}[\theta_i^e \mid R_1] \leq 0,\end{aligned} \tag{32}$$

where $\mathbb{E}[\nu_i \mid R_1] = 0$ follows from rational expectations (RE) and the assumption that agents know their realized cost at the time they were assigned to treatment i.e.

$$\mathbb{E}[\nu_i \mid R_1] = \mathbb{E}\left[\mathbb{E}[\nu_i \mid \mathcal{I}_i, C_i, \theta_i^e]\right] = 0.$$

Similar to inequality (20), and following from the above observations, the following is true:

$$\begin{aligned}\mathbb{E}[\theta_i \mid R_i = 0, D_i = 0, Z_i = 1] &= \mathbb{E}[\theta_i^e + \nu_i \mid R_0] \\ &= \mathbb{E}[\theta_i^e \mid R_0] \geq 0.\end{aligned} \tag{33}$$

It remains to be shown that the bounds for the average treatment effect in proposition 1 are tight. This amounts to showing that the upper and lower bounds in (24) can be attained by some distribution of underlying variables which are conformable with the inequalities above, and which do not contradict identified quantities. By assumptions IA and CT the distribution of potential outcome $Y_i(0)$ is identified since,

$$\mathbb{P}(Y_i(0) \leq y) = \mathbb{P}(Y_i \leq y \mid Z_i = 0),$$

so any distribution of underlying variables must possess a marginal distribution of $Y_i(0)$ which coincides with its identified version. Moreover, any valid data generating process (DGP) must satisfy the following,

$$\begin{aligned}\mathbb{E}[Y_i(1) - Y_i(0) \mid \theta_i^e \geq C_i] &= \theta^{LATE}, \\ \mathbb{E}[Y_i(0) \mid R_1] &= \mathbb{E}[Y_i \mid R_i = 1, D_i = 0, Z_i = 1], \\ \mathbb{E}[Y_i(0) \mid R_0] &= \mathbb{E}[Y_i \mid R_i = 0, D_i = 0, Z_i = 1], \\ \mathbb{P}(D_i = 1 \mid Z_i = 1) &= \mathbb{P}(\theta_i^e \geq C_i), \\ \mathbb{P}(R_i = 1, D_i = 0 \mid Z_i = 1) &= \mathbb{P}(\theta_i^e < C_i;\ \theta_i^e < 0) + \mathbb{P}(0 = \theta_i^e < C_i;\ B_i = 1) \\ \mathbb{P}(R_i = 0, D_i = 0 \mid Z_i = 1) &= \mathbb{P}(0 < \theta_i^e < C_i) + \mathbb{P}(0 = \theta_i^e < C_i;\ B_i = 0)\end{aligned} \tag{34}$$

Consider any DGP i.e. a joint distribution of the vector $(Y_i(1), Y_i(0), \mathcal{I}_i, C_i, B_i)$ which has a conformable marginal distribution for $Y_i(0)$ and satisfies the relations in (34). This implies the following relation,

$$\begin{aligned}\mathbb{E}[\theta_i \mid R_1] = \mathbb{E}[\theta_i^e \mid R_1] = \mathbb{E}[Y_i(1) \mid R_1] - \mathbb{E}[Y_i \mid R_i = 1, D_i = 0, Z_i = 1] \\ \mathbb{E}[\theta_i \mid R_0] = \mathbb{E}[\theta_i^e \mid R_0] = \mathbb{E}[Y_i(1) \mid R_0] - \mathbb{E}[Y_i \mid R_i = 0, D_i = 0, Z_i = 1].\end{aligned} \tag{35}$$

The lower bound in (24) can be attained by a DGP such that the following is true:

$$\mathbb{E}[Y_i(1) \mid R_1] = 0, \quad \mathbb{E}[Y_i(1) \mid R_0] = \mathbb{E}[Y_i \mid R_i = 0, D_i = 0, Z_i = 1]. \tag{36}$$

It should be noted that the second condition above requires the following to be true:

$$\mathbb{P}(0 < \theta_i^e < C_i) = 0.$$

It is only possible to satisfy the above and the equality for $\mathbb{P}(R_i = 0, D_i = 0 \mid Z_i = 1)$ in (34) if there is a non-degenerate tie breaking rule B_i which equals zero with positive probability.

Similarly, the upper bound in (24) can be attained by any DGP which satisfies (34), has conformable marginal distribution of $Y_i(0)$ and also satisfies:

$$\mathbb{E}[Y_i(1) \mid R_1] = \mathbb{E}[Y_i \mid R_i = 1, D_i = 0, Z_i = 1], \quad \mathbb{E}[Y_i(1) \mid R_0] = 1, \tag{37}$$

which is again only possible if B_i equals one with positive probability and,

$$\mathbb{P}(\theta_i^e < C_i;\ \theta_i^e < 0) = 0.$$

Similar arguments show that it is possible to construct an underlying joint distribution which attains any value of the ATE between the lower and upper bounds.

B Derivation of Inequalities in Proposition 2

Assume IA, CT, RE and IC hold. The inequalities for $\mathbb{P}(D_i = 0 \mid Z_i = 1)$ are already derived in the discussion preceding proposition 2. Let $y \in [0, 1]$. The claim that the upper and lower bounds for the ATE derived under rational expectations (RE) in proposition 1 still apply in proposition 2 may be established by replicating the proof of proposition 1 while conditioning throughout on the type variable τ_i. The decision rule definitions in (15) and (17) imply the following:

$$
\begin{aligned}
\mathbb{P}\left(Y_i \leq y, D_i = 1 \mid Z_i = 1\right) &= \mathbb{E}\left[\mathbb{P}\left(Y_i(1) \leq y \mid \mathcal{I}_i\right) \cdot G\left(\theta_i^e\right)\right] \\
&= \mathbb{E}\left[\frac{\mathbb{P}\left(Y_i(1) \leq y \mid \mathcal{I}_i\right) \cdot G\left(\theta_i^e\right)}{1 - G\left(\theta_i^e\right)} \cdot (1 - D_i) \middle| Z_i = 1\right] \\
&\leq \mathbb{E}\left[\mathbb{E}\left[\frac{\mathbb{P}\left(Y_i(1) \leq y \mid \mathcal{I}_i\right) \cdot G\left(\theta_i^e\right)}{1 - G\left(\theta_i^e\right)} \middle| \mathcal{I}_i, Z_i = 1\right]\right] \\
&\leq \mathbb{E}\left[\mathbb{P}\left(Y_i(1) \leq y \mid \mathcal{I}_i\right) \cdot \mathbb{E}\left[\frac{G\left(Y_i(1) - Y_i(0)\right)}{1 - G\left(Y_i(1) - Y_i(0)\right)} \middle| \mathcal{I}_i\right]\right] \\
&= \mathbb{E}\left[\mathbb{P}\left(Y_i(1) \leq y \mid \mathcal{I}_i\right) \cdot \frac{G\left(Y_i(1) - Y_i(0)\right)}{1 - G\left(Y_i(1) - Y_i(0)\right)} \middle| \mathcal{I}_i\right] \\
&= \mathbb{E}\left[\mathbb{P}\left(Y_i(1) \leq y \mid \mathcal{I}_i\right) \cdot \frac{G\left(Y_i(1) - Y_i(0)\right)}{1 - G\left(Y_i(1) - Y_i(0)\right)}\right] \\
&\leq \mathbb{E}\left[\mathbb{P}\left(Y_i(1) \leq y \mid \mathcal{I}_i\right)^2\right]^{1/2} \cdot \mathbb{E}\left[\left(\frac{G\left(Y_i(1) - Y_i(0)\right)}{1 - G\left(Y_i(1) - Y_i(0)\right)}\right)^2\right]^{1/2} \\
&\leq \mathbb{E}\left[\mathbb{P}\left(Y_i(1) \leq y \mid \mathcal{I}_i\right)\right]^{1/2} \cdot \mathbb{E}\left[\left(\frac{G\left(Y_i(1) - Y_i(0)\right)}{1 - G\left(Y_i(1) - Y_i(0)\right)}\right)^2\right]^{1/2} \\
&= \mathbb{E}\left[\mathbb{P}\left(Y_i(1) \leq y\right)\right]^{1/2} \cdot \mathbb{E}\left[\left(\frac{G\left(Y_i(1) - Y_i(0)\right)}{1 - G\left(Y_i(1) - Y_i(0)\right)}\right)^2\right]^{1/2}
\end{aligned}
\tag{38}
$$

The first inequality above follows because $(1 - D_i) \in \{0, 1\}$. The second inequality follows from Jensen's inequality and the assumption that the mapping $t \mapsto G(t)/(1 - G(t))$ is convex. The third inequality follows from Cauchy-Schwarz. The fourth inequality follows because probabilities lie between 0 and 1.

Similarly, the inequality for $\mathbb{P}(Y_i \leq y, D_i = 0 \mid Z_i = 1)$ may be derived as:

$$
\begin{aligned}
\mathbb{P}\left(Y_i \leq y, D_i = 0 \mid Z_i = 1\right) &= \mathbb{E}\left[\mathbb{P}\left(Y_i(0) \leq y \mid \mathcal{I}_i\right) \cdot \left(1 - G\left(\theta_i^e\right)\right)\right] \\
&= \mathbb{E}\left[\frac{\mathbb{P}\left(Y_i(0) \leq y \mid \mathcal{I}_i\right) \cdot \left(1 - G\left(\theta_i^e\right)\right)}{G\left(\theta_i^e\right)} \cdot D_i \middle| Z_i = 1\right] \\
&\leq \mathbb{E}\left[\mathbb{E}\left[\frac{\mathbb{P}\left(Y_i(0) \leq y \mid \mathcal{I}_i\right) \cdot \left(1 - G\left(\theta_i^e\right)\right)}{G\left(\theta_i^e\right)} \middle| \mathcal{I}_i, Z_i = 1\right]\right] \\
&\leq \mathbb{E}\left[\mathbb{P}\left(Y_i(0) \leq y \mid \mathcal{I}_i\right) \cdot \mathbb{E}\left[\frac{1 - G\left(Y_i(1) - Y_i(0)\right)}{G\left(Y_i(1) - Y_i(0)\right)} \middle| \mathcal{I}_i\right]\right] \\
&= \mathbb{E}\left[\mathbb{P}\left(Y_i(0) \leq y \mid \mathcal{I}_i\right) \cdot \frac{1 - G\left(Y_i(1) - Y_i(0)\right)}{G\left(Y_i(1) - Y_i(0)\right)} \middle| \mathcal{I}_i\right] \\
&= \mathbb{E}\left[\mathbb{P}\left(Y_i(0) \leq y \mid \mathcal{I}_i\right) \cdot \frac{1 - G\left(Y_i(1) - Y_i(0)\right)}{G\left(Y_i(1) - Y_i(0)\right)}\right] \\
&\leq \mathbb{E}\left[\mathbb{P}\left(Y_i(0) \leq y \mid \mathcal{I}_i\right)^2\right]^{1/2} \cdot \mathbb{E}\left[\left(\frac{1 - G\left(Y_i(1) - Y_i(0)\right)}{G\left(Y_i(1) - Y_i(0)\right)}\right)^2\right]^{1/2} \\
&\leq \mathbb{E}\left[\mathbb{P}\left(Y_i(0) \leq y \mid \mathcal{I}_i\right)\right]^{1/2} \cdot \mathbb{E}\left[\left(\frac{1 - G\left(Y_i(1) - Y_i(0)\right)}{G\left(Y_i(1) - Y_i(0)\right)}\right)^2\right]^{1/2} \\
&= \mathbb{E}\left[\mathbb{P}\left(Y_i(0) \leq y\right)\right]^{1/2} \cdot \mathbb{E}\left[\left(\frac{1 - G\left(Y_i(1) - Y_i(0)\right)}{G\left(Y_i(1) - Y_i(0)\right)}\right)^2\right]^{1/2}
\end{aligned}
\tag{39}
$$

C Estimation and Inference

This section describes the procedure to estimate and conduct inference on identified sets when a finite sample is observed. Suppose the data is composed of i.i.d. draws of the vector $S_i = (Y_i, D_i, Z_i, R_i, X_i)$ for $i = 1, 2, .., n$ for some finite integer n. I first describe the estimation outline for bounds which don't condition on covariates X_i. The lower and upper bounds in (24) and the worst case bounds depend on several identified probabilities, each of which can be estimated from

data in the following way:

$$\hat{\mathbb{P}}(D_i = 1 \mid Z_i = 1) = \frac{\sum_{i=1}^n D_i \cdot Z_i}{\sum_{i=1}^n Z_i},$$

$$\hat{\mathbb{P}}(R_i = r, D_i = 0 \mid Z_i = 1) = \frac{\sum_{i=1}^n \mathbb{1}\{R_i = r\} \cdot (1 - D_i) \cdot Z_i}{\sum_{i=1}^n Z_i}, \quad \text{for } r = 0, 1,$$

$$\hat{\mathbb{E}}[Y_i \mid Z_i = z] = \frac{\sum_{i=1}^n Y_i \cdot \mathbb{1}\{Z_i = z\}}{\sum_{i=1}^n \mathbb{1}\{Z_i = z\}}, \qquad \text{for } z = 0, 1,$$

$$\hat{\theta}^{LATE} = \frac{\hat{\mathbb{E}}[Y_i \mid Z_i = 1] - \hat{\mathbb{E}}[Y_i \mid Z_i = 0]}{\hat{\mathbb{P}}(D_i = 1 \mid Z_i = 1)}$$

$$\hat{\mathbb{E}}[Y_i \mid R_i = r, D_i = 0, Z_i = 1] = \frac{\sum_{i=1}^n Y_i \cdot \mathbb{1}\{R_i = r\} \cdot (1 - D_i) \cdot Z_i}{\sum_{i=1}^n \mathbb{1}\{R_i = r\} \cdot (1 - D_i) \cdot Z_i}, \quad \text{for } r = 0, 1. \tag{40}$$

The estimated identified set for the ATE is defined by replacing the population level quantities with their estimated versions. For example, the estimated lower and upper bounds for the ATE under the model are given by:

$$\hat{l}_n^m = \hat{\mathbb{P}}(D_i = 1 \mid Z_i = 1) \cdot \hat{\theta}^{LATE} - \hat{\mathbb{P}}(R_i = 1, D_i = 0 \mid Z_i = 1) \cdot \hat{\mathbb{E}}[Y_i \mid R_i = 1, D_i = 0, Z_i = 1]$$

$$\hat{u}_n^m = \hat{\mathbb{P}}(D_i = 1 \mid Z_i = 1) \cdot \hat{\theta}^{LATE} + \hat{\mathbb{P}}(R_i = 0, D_i = 0 \mid Z_i = 1) \cdot (1 - \hat{\mathbb{E}}[Y_i \mid R_i = 0, D_i = 0, Z_i = 1]). \tag{41}$$

Similarly, the estimated worst case bounds are given by,

$$\hat{l}_n^w = \hat{\mathbb{P}}(D_i = 1 \mid Z_i = 1) \cdot \hat{\theta}^{LATE} - \hat{\mathbb{P}}(D_i = 0 \mid Z_i = 1) \cdot \hat{\mathbb{E}}[Y_i \mid D_i = 0, Z_i = 1]$$

$$\hat{u}_n^w = \hat{\mathbb{P}}(D_i = 1 \mid Z_i = 1) \cdot \hat{\theta}^{LATE} + \hat{\mathbb{P}}(D_i = 0 \mid Z_i = 1) \cdot (1 - \hat{\mathbb{E}}[Y_i \mid D_i = 0, Z_i = 1]). \tag{42}$$

For either the worst case bounds, or the model implied bounds, the following is true by standard asymptotic results,

$$\sqrt{n} \begin{bmatrix} \hat{l}_n - l \\ \hat{u}_n - u \end{bmatrix} \xrightarrow{d} \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_l^2 & \rho\sigma_l\sigma_u \\ \rho\sigma_l\sigma_u & \sigma_u^2 \end{bmatrix} \right), \tag{43}$$

where the standard deviations σ_l, σ_u and correlation ρ can all be consistently estimated. The true lower and upper bounds are assumed to be fixed value i.e. l^m, u^m, l^w, u^w are not indexed by the sample size n. Turning now to inference, given an estimated identified set of the form $\hat{\Theta}_I = [\hat{l}_n, \hat{u}_n]$, the objective is to construct a confidence set $\hat{C}_n$ such that the following is true:

$$\lim_{n \to \infty} \mathbb{P}(\theta_0 \in \hat{C}_n) \geq 1 - \alpha,$$

for some specified value $\alpha \in (0,1)$. The confidence regions used here will be of the form:

$$\hat{C}_n = \left[\hat{l}_n - \frac{\hat{\sigma}_l \hat{c}_l}{\sqrt{n}},\ \ \hat{u}_n + \frac{\hat{\sigma}_u \hat{c}_u}{\sqrt{n}}\right],$$

where $\hat{\sigma}_l, \hat{\sigma}_u$ are the estimated standard deviations for the respective bounds and the critical values $\hat{c}_l, \hat{c}_u$ are derived following the procedures in Stoye [2009] which are themselves refinements of procedures in Imbens and Manski [2004]. Without replicating any proofs, the remainder of this section briefly describes how the critical values $\hat{c}_l, \hat{c}_u$ are calculated in finite samples and why they lead to the correct asymptotic coverage for the unknown average treatment effect θ. By definition, the critical values values are chosen such that for any possible value $\theta \in [l, u]$, the following is true:

$$\lim_{n\to\infty} \mathbb{P}\left(\hat{l}_n - \frac{\hat{\sigma}_l \hat{c}_l}{\sqrt{n}} \ \leq\ \theta \ \leq\ \hat{u}_n + \frac{\hat{\sigma}_u \hat{c}_u}{\sqrt{n}}\right) \geq 1 - \alpha.$$

As has been detained in Stoye [2009], Imbens and Manski [2004], at the limit as $n \to \infty$, this probability is minimized at $\theta \in \{l, u\}$. given the asymptotic normality results in (43). In finite samples, $\hat{c}_l$ and $\hat{c}_u$ are jointly chosen such that the probability of the lower bound l (upper bound u) lying in the confidence set is at least $1 - \alpha$. The following conditions must therefore be satisfied for any valid c_l, c_u:

$$\mathbb{P}\left(\hat{l}_n - \frac{\hat{\sigma}_l c_l}{\sqrt{n}} \leq l \leq \hat{u}_n + \frac{\hat{\sigma}_u c_u}{\sqrt{n}}\right) \geq 1 - \alpha \tag{44}$$

$$\mathbb{P}\left(\hat{l}_n - \frac{\hat{\sigma}_l c_l}{\sqrt{n}} \leq u \leq \hat{u}_n + \frac{\hat{\sigma}_u c_u}{\sqrt{n}}\right) \geq 1 - \alpha. \tag{45}$$

Consider the inequality in (44) and define $\Delta = u - l$, the unknown length of the population level identified set for θ. The joint asymptotic normality in (43) implies that:

$$\sqrt{n}(\hat{u}_n - u) \mid \sqrt{n}(\hat{l}_n - n) \xrightarrow{d} \mathcal{N}\left(\rho\frac{\sigma_u}{\sigma_l}\sqrt{n}(\hat{l}_n - n),\ \ \sigma_u^2(1-\rho^2)\right). \tag{46}$$

The following rough argument illustrates how to approximate the unknown sampling distribution

of the estimated quantities in (44):

$$\begin{aligned}
\mathbb{P}\left(\hat{l}_n - \frac{\hat{\sigma}_l c_l}{\sqrt{n}} \leq l \leq \hat{u}_n + \frac{\hat{\sigma}_u c_u}{\sqrt{n}}\right) &= \mathbb{P}\left(\frac{\sqrt{n}}{\hat{\sigma}_l}\left(\hat{l}_n - l\right) \leq c_l,\ \frac{\sqrt{n}}{\hat{\sigma}_u}\left(\hat{u}_n - u\right) \geq -\frac{\sqrt{n}\Delta}{\hat{\sigma}_u} - c_u\right) \\
&= \mathbb{E}\left(\mathbb{P}\left(\frac{\sqrt{n}}{\hat{\sigma}_u}\left(\hat{u}_n - u\right) \geq -\frac{\sqrt{n}\Delta}{\hat{\sigma}_u} - c_u \middle| \sqrt{n}\left(\hat{l}_n - l\right)\right) \times \right. \\
&\qquad \left. \mathbb{1}\left\{\frac{\sqrt{n}}{\hat{\sigma}_l}\left(\hat{l}_n - l\right) \leq c_l\right\}\right) \\
&\approx \mathbb{E}\left(\Phi\left(\frac{\hat{\rho}}{\sqrt{1-\hat{\rho}^2}} \cdot \frac{\sqrt{n}}{\hat{\sigma}_l}\left(\hat{l}_n - l\right) + \frac{c_u}{\sqrt{1-\hat{\rho}^2}} + \frac{\sqrt{n}\Delta}{\hat{\sigma}_u\sqrt{1-\hat{\rho}^2}}\right) \times \right. \\
&\qquad \left. \mathbb{1}\left\{\frac{\sqrt{n}}{\hat{\sigma}_l}\left(\hat{l}_n - l\right) \leq c_l\right\}\right) \\
&\approx \int_{-\infty}^{c_l} \Phi\left(\frac{\hat{\rho}}{\sqrt{1-\hat{\rho}^2}} \cdot z + \frac{c_u}{\sqrt{1-\hat{\rho}^2}} + \frac{\sqrt{n}\Delta}{\hat{\sigma}_u\sqrt{1-\hat{\rho}^2}}\right)\phi(z)\mathrm{d}z \qquad (47)
\end{aligned}$$

The first approximation relation above uses the conditional convergence in distribution result in (46) while the second approximation uses the asymptotic normality result in (43). By a similar argument, the following approximate relationship can be established for $\theta = u$:

$$\mathbb{P}\left(\hat{l}_n - \frac{\hat{\sigma}_l c_l}{\sqrt{n}} \leq u \leq \hat{u}_n + \frac{\hat{\sigma}_u c_u}{\sqrt{n}}\right) \approx \int_{-\infty}^{c_u} \Phi\left(\frac{\hat{\rho}}{\sqrt{1-\hat{\rho}^2}} \cdot z + \frac{c_l}{\sqrt{1-\hat{\rho}^2}} + \frac{\sqrt{n}\Delta}{\hat{\sigma}_l\sqrt{1-\hat{\rho}^2}}\right)\phi(z)\mathrm{d}z. \qquad (48)$$

The approximate equalities in (47), (48) then suggests choosing $\hat{c}_l, \hat{c}_u$ by minimizing the amount by which the estimated identified set is enlarged, $\hat{\sigma}_l \cdot c_l + \hat{\sigma}_u \cdot c_u$ subject to the following restrictions,

$$\begin{aligned}
&\int_{-\infty}^{c_l} \Phi\left(\frac{\hat{\rho}}{\sqrt{1-\hat{\rho}^2}} \cdot z + \frac{c_u}{\sqrt{1-\hat{\rho}^2}} + \frac{\sqrt{n}\hat{\Delta}}{\hat{\sigma}_u\sqrt{1-\hat{\rho}^2}}\right)\phi(z)\mathrm{d}z \geq 1-\alpha \\
&\int_{-\infty}^{c_u} \Phi\left(\frac{\hat{\rho}}{\sqrt{1-\hat{\rho}^2}} \cdot z + \frac{c_l}{\sqrt{1-\hat{\rho}^2}} + \frac{\sqrt{n}\hat{\Delta}}{\hat{\sigma}_l\sqrt{1-\hat{\rho}^2}}\right)\phi(z)\mathrm{d}z \geq 1-\alpha, \qquad (49)
\end{aligned}$$

where $\hat{\Delta} = \hat{u} - \hat{l}$. The validity of the empirically calculated critical values $\hat{c}_l, \hat{c}_u$ does not require any additional assumptions apart from the asymptotic normality of the estimated bounds and that the variances $\underline{\sigma} \leq \sigma_l^2, \sigma_u^2 \leq \overline{\sigma}$ for strictly positive, finite values $\underline{\sigma}, \overline{\sigma}$.

D Identification under Perfect Foresight

This section examines the limits of identification under the rational expectations assumption (RE). I do this by considering the strongest version of RE , namely that of perfect foresight. Under this

assumptions, agents exactly know the potential impact of treatment.

Assumption PF (Perfect Foresight) : $\theta_i^e = Y_i(1) - Y_i(0)$.

This is a strengthened version of rational expectations (RE) where the error term ν_i is assumed to equal zero with probability one. For simplicity, this section will assume that outcomes are discrete and binary i.e. $Y_i \in \{0, 1\}$. An extension to multiple discrete outcomes or continuous outcomes is straightforward though the case with continuous outcomes poses computational challenges which are not addressed here. To begin, perfect foresight (PF) clearly illustrates the importance of accounting for the case when $\theta_i^e = 0$ beyond its utility to prove sharpness of the identified intervals in Proposition 1. For example, suppose there was no tie breaking variable B_i as specified in (17). Without a non-degenerate B_i, perfect foresight may fail to rationalize observed data due to the manner in which agents choose to participate in treatment. To see this, suppose the observed distribution of outcomes identifies a positive probability of a non-complying agent who reported $R_i = 0$ and for whom $Y_i = 1$ i.e. :

$$\mathbb{P}(Y_i = 1, D_i = 0, R_i = 0 \mid Z_i = 1) > 0.$$

The model states that these agents would have accepted treatment if not for an unexpectedly high cost shock. Under perfect foresight (PF), it must be that,

$$\mathbb{P}(Y_i(0) = 1,\ \ 0 < \theta_i^e < C_i) = \mathbb{P}(Y_i(0) = 1,\ \ 1 < Y_i(1) < 1 + C_i) = 0,$$

since the highest possible value of $Y_i(1)$ is 1. However, the observed data then rejects the model. The assumption could then contradict the observed distribution of outcomes. The unobserved variable B_i prevents this contradiction. In principle, B_i may be arbitrarily correlated with other observed and unobserved elements of the DGP. For the sake of simplicity, it is assumed that B_i is independent of the cost shock C_i.

Assumption ITB (Independent Tie Breaking) : $B_i \perp C_i$.

Independent Tie Breaking (ITB) allows the probability that B_i equals one to vary based on the values of the potential outcomes. For example, consider the event that $Y_i(1)$ and $Y_i(0)$ both equal $y \in \{0, 1\}$. Under PF and ITB, θ_i^e would equal zero and a fixed proportion of agents in the population, p_y will choose to reject treatment. With this structure the joint distribution of potential outcomes and the variable B_i may be described by the combination of two objects. The first is a vector of real numbers $\{p_0, p_1\}$ where $p_y \in [0, 1]$ measures $\mathbb{P}(B_i = 1 \mid Y_i(1) = Y_i(0) = y)$. The second is a joint probability density function π which describes the distribution of the vector $(Y_i(1), Y_i(0), C_i)$. There are no additional assumptions on C_i other than the requirement that its support includes an interval $(-a, a)$ for some $a \in \mathbb{R}$ i.e. C_i takes on both positive and negative values with positive probability. For any such candidate density, define the following convenient

shorthand:

$$\pi\left(y_1, y_0, C_i \leq y_1 - y_0\right) = \int_{-\infty}^{y_1 - y_0} \pi\left(Y_i(1) = y_1, Y_i(0) = y_0, c\right) \mathrm{d}c$$

$$\pi\left(y_1, y_0, C_i > y_1 - y_0\right) = \int_{y_1 - y_0}^{\infty} \pi\left(Y_i(1) = y_1, Y_i(0) = y_0, c\right) \mathrm{d}c. \tag{50}$$

Under assumptions IA, CT, PF, and ITB, the observed data and the model are related as:

$$\begin{aligned}
\mathbb{P}\left(Y_i = 1, D_i = 1 \mid Z_i = 1\right) =& \pi(1,1,C \leq 0) + \pi(1,0,C_i \leq 1) \\
\mathbb{P}(Y_i = 0, D_i = 1 \mid Z_i = 1) =& \pi(0,0,C \leq 0) + \pi(0,1,C_i \leq -1) \\
\mathbb{P}\left(Y_i = 1, R_i = 1, D_i = 0 \mid Z_i = 1\right) =& \pi(1,1,C > 0) \cdot p_1 + \pi(0,1,C_i > -1) \\
\mathbb{P}\left(Y_i = 0, R_i = 1, D_i = 0 \mid Z_i = 1\right) =& \pi(0,0,C > 0) \cdot p_0 \\
\mathbb{P}\left(Y_i = 1, R_i = 0, D_i = 0 \mid Z_i = 1\right) =& \pi(1,1,C > 0) \cdot (1 - p_1) \\
\mathbb{P}\left(Y_i = 0, R_i = 0, D_i = 0 \mid Z_i = 1\right) =& \pi(0,0,C > 0) \cdot (1 - p_0) + \pi(1,0,C_i > 1) \\
\mathbb{P}(Y_i = 1 \mid Z_i = 0) =& \pi(1,1,C_i \leq 0) + \pi(1,1,C_i > 0) + \\
& \pi(0,1,C_i \leq -1) + \pi(0,1,C_i > -1)
\end{aligned} \tag{51}$$

The average treatment effect with binary outcomes is simply $\mathbb{P}(Y_i(1) = 1) - \mathbb{P}(Y_i(0) = 1)$. Given any density π, the maximum identified value is given by the solution to the following problem:

$$\max_{\pi, p_1, p_0} \pi(1,0,C_i \leq 1) + \pi(1,0,C_i > 1) - \pi(0,1,C_i \leq -1) - \pi(0,1,C_i > -1), \quad s.t. \tag{52}$$

$$\sum_{y_1=0}^{1} \sum_{y_0=0}^{1} \pi(y_1, y_0, C_i \leq y_1 - y_0) + \pi(y_1, y_0, C_i > y_1 - y_0) = 1,$$

$$0 \leq \pi(y_1, y_0, C_i \leq y_1 - y_0), \pi(y_1, y_0, C_i > y_1 - y_0) \leq 1, \quad 0 \leq p_1, p_0, \leq 1,$$

the conditions in (51), (53)

while the minimum identified value is given by a problem which minimizes the objective function in (52) subject to the same set of constraints. These are computationally tractable problems since they are simple non-linear programming problems where a linear objective function is optimized over a convex set.

As an empirical illustration, consider the outcome of interest in Table 5 i.e. whether or not an applicant managed to find employment within 12 months of the treatment phase. Table 9 shows the estimated lower bound on the average treatment effect when perfect foresight is assumed. It also displays the estimated probability vectors which led to the attained lower bound. To save on notation, the following shorthand is used for the column names:

$$\pi_{y_1 y_0, l} = \mathbb{P}(Y_i(1) = y_1, Y_i(0) = y_0, C_i \leq y_1 - y_0), \quad \pi_{y_1 y_0, g} = \mathbb{P}(Y_i(1) = y_1, Y_i(0) = y_0, C_i > y_1 - y_0).$$

Table 10 displays similar results for the upper bound on the ATE. Bounds derived under the setup of Proposition 1 are labeled ATE (RE) to signify that rational expectations (RE) is the major behavioral assumption on agents. Bounds under the stronger perfect foresight assumption (PF) are labeled ATE (PF). Most surprisingly, the stronger version of perfect foresight offers no benefit in terms of a smaller identified set for the ATE. Secondly, almost all of the difference in the lower and upper bounds is driven by changes in the underlying probabilities $\pi_{y_1y_0,g}$ i.e. when cost prohibits agents from accepting treatment.

Description			ATE (RE)	ATE (PF)	$\pi_{00,l}$	$\pi_{01,l}$	$\pi_{10,l}$	$\pi_{11,l}$	$\pi_{00,g}$	$\pi_{01,g}$	$\pi_{10,g}$	$\pi_{11,g}$	p_0	p_1
Overall			-0.03	-0.03	0.04	0.07	0.08	0.47	0.08	0.05	0.00	0.22	0.09	0.00
No income	Age < 30	Did not graduate HS	-0.05	-0.05	0.07	0.14	0.12	0.36	0.13	0.02	0.00	0.16	0.12	0.00
		Graduated HS	0.07	0.07	0.07	0.08	0.20	0.32	0.08	0.05	0.00	0.21	0.03	0.00
	Age ≥ 30	Did not graduate HS	0.01	0.01	0.07	0.10	0.14	0.28	0.20	0.03	0.00	0.18	0.12	0.00
		Graduated HS	0.03	0.03	0.09	0.12	0.18	0.33	0.12	0.02	0.00	0.14	0.03	0.00
Pos. income	Age < 30	Did not graduate HS	-0.03	-0.03	0.02	0.04	0.07	0.46	0.06	0.06	0.00	0.28	0.05	0.00
		Graduated HS	-0.05	-0.05	0.02	0.05	0.06	0.55	0.04	0.06	0.00	0.23	0.09	0.00
	Age ≥ 30	Did not graduate HS	-0.07	-0.07	0.03	0.06	0.04	0.48	0.10	0.06	0.00	0.24	0.14	0.00
		Graduated HS	-0.05	-0.05	0.03	0.06	0.05	0.53	0.07	0.04	0.00	0.22	0.10	0.00

Table 9: Estimated lower bounds for ATE on whether applicant found a job within 12 months of the treatment phase, under perfect foresight.

Description			ATE (RE)	ATE (PF)	$\pi_{00,l}$	$\pi_{01,l}$	$\pi_{10,l}$	$\pi_{11,l}$	$\pi_{00,g}$	$\pi_{01,g}$	$\pi_{10,g}$	$\pi_{11,g}$	p_0	p_1
Overall			0.09	0.09	0.04	0.07	0.08	0.47	0.01	0.00	0.08	0.27	1.00	0.18
No income	Age < 30	Did not graduate HS	0.09	0.09	0.07	0.14	0.12	0.36	0.02	0.00	0.11	0.18	1.00	0.12
		Graduated HS	0.20	0.20	0.07	0.08	0.20	0.32	0.00	0.00	0.08	0.25	1.00	0.18
	Age ≥ 30	Did not graduate HS	0.22	0.22	0.07	0.10	0.14	0.28	0.02	0.00	0.18	0.22	1.00	0.16
		Graduated HS	0.17	0.17	0.09	0.12	0.18	0.33	0.00	0.00	0.12	0.16	1.00	0.13
Pos. income	Age < 30	Did not graduate HS	0.09	0.09	0.03	0.04	0.07	0.47	0.00	0.00	0.06	0.33	1.00	0.17
		Graduated HS	0.05	0.05	0.03	0.05	0.05	0.55	0.00	0.00	0.04	0.29	1.00	0.20
	Age ≥ 30	Did not graduate HS	0.07	0.07	0.03	0.06	0.04	0.48	0.01	0.00	0.09	0.29	1.00	0.19
		Graduated HS	0.06	0.06	0.03	0.06	0.05	0.53	0.01	0.00	0.06	0.27	1.00	0.17

Table 10: Estimated upper bounds for ATE on whether applicant found a job within 12 months of the treatment phase, under perfect foresight.

By construction, perfect foresight in the case with binary outcomes cannot lead to tighter bounds for the ATE. Perfect foresight does lead to substantially smaller identified sets if the outcome takes on more than two values. Table 11 displays the estimated identified sets for the ATE on the total number of months an applicant was employed post the treatment phase. The bounds under perfect foresight are now significantly tighter than those under just rational expectations (as shown in Table 6). In some cases, perfect foresight (PF) almost leads to point identification[2].

[2]Estimated probabilities which lead to the corresponding bounds are not shown as the number of parameters is too large; agents can be employed between 0 to 30 months which leads to $2 \cdot 31^2$ positive real numbers needed to define the joint distribution of potential outcomes. In addition, there are 31 real numbers in $[0, 1]$ which determine the tue breaking rule.

Description			ATE (RE)	ATE (PF)
Overall			[-0.48 , 5.11]	[0.81 , 1.04]
No income	Age < 30	Did not graduate HS	[-0.51 , 4.36]	[0.73 , 0.73]
		Graduated HS	[1.04 , 6.37]	[2.00 , 2.01]
	Age ≥ 30	Did not graduate HS	[-0.08 , 6.58]	[2.57 , 2.57]
		Graduated HS	[-0.21 , 4.18]	[0.04 , 1.38]
Pos. income	Age < 30	Did not graduate HS	[-0.55 , 5.78]	[1.02 , 1.04]
		Graduated HS	[-0.75 , 4.50]	[0.56 , 1.01]
	Age ≥ 30	Did not graduate HS	[-0.48 , 5.87]	[1.03 , 1.03]
		Graduated HS	[-0.89 , 4.50]	[0.37 , 0.80]

Table 11: Estimated identified sets for the ATE on total months employed post treatment under perfect foresight.

E Additional Figures

E.1 Distribution Treatment Effects for months employed and maximum consecutive months employed

The following figures display treatment effects on the distribution of months employed and maximum consecutive months employed over the 30 month period after treatment on various sub-groups of participants in the JTPA study. Agents are split into mutually exclusive groups based on their education, age and whether or not they reported any income in the year prior to being assigned to the treated or control group.

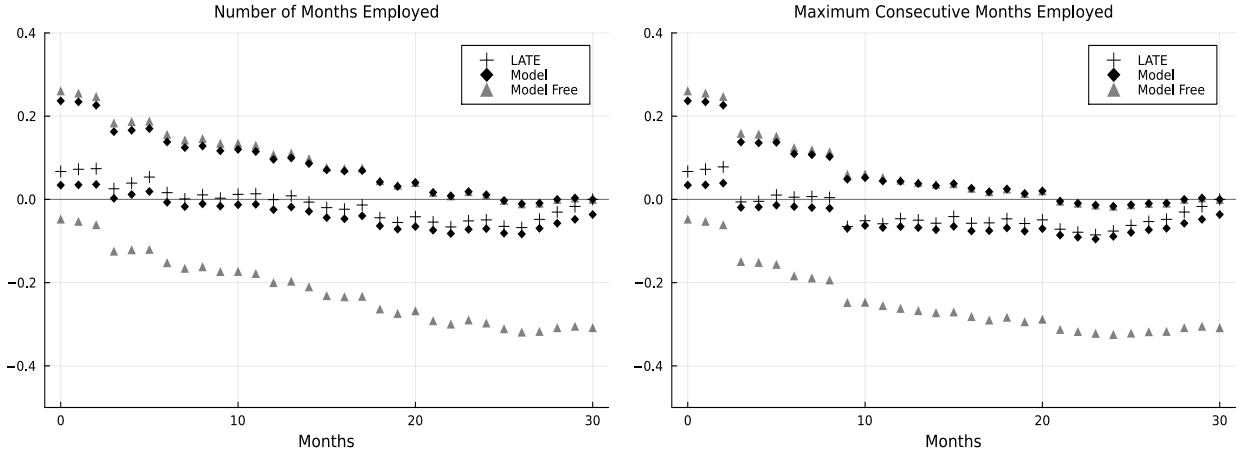

Figure 7: Distributional treatment effect for applicants who did not graduate high school, were aged less than 30, and had no prior income during the year prior to assignment.

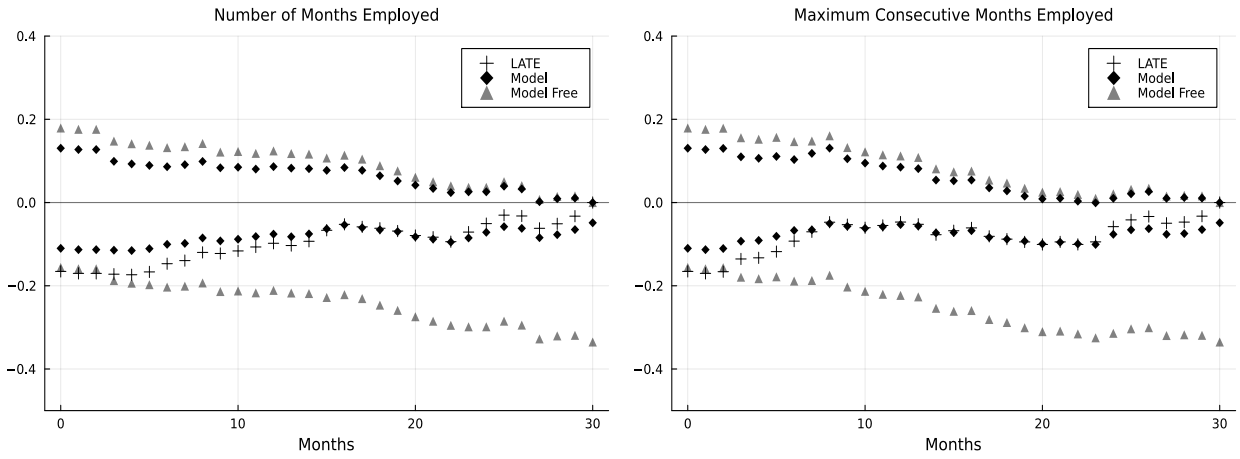

Figure 8: Distributional treatment effect for applicants who had graduated high school, were aged less than 30, and had no prior income during the year prior to assignment.

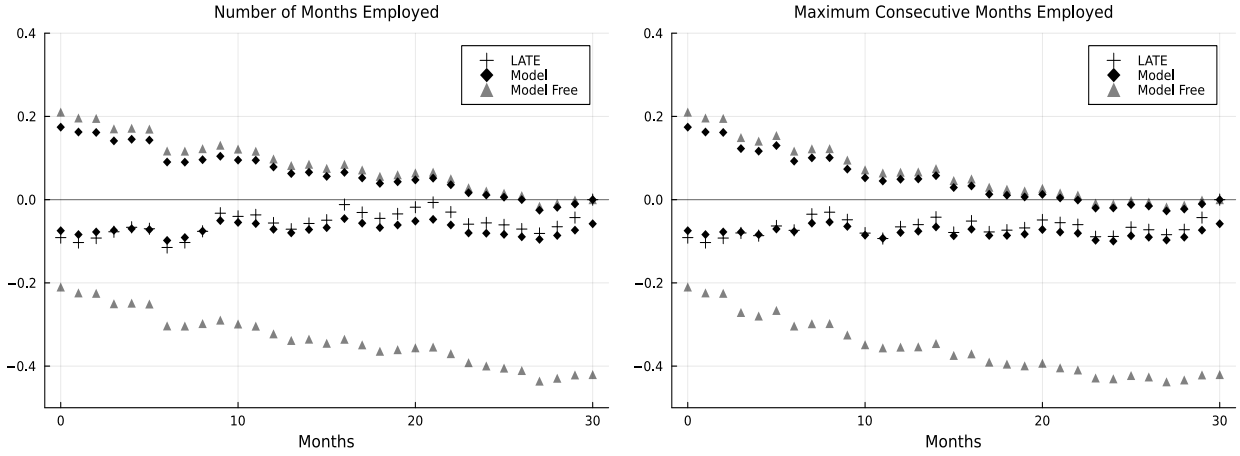

Figure 9: Distributional treatment effect for applicants who did not graduate high school, were aged at least 30, and had no prior income during the year prior to assignment.

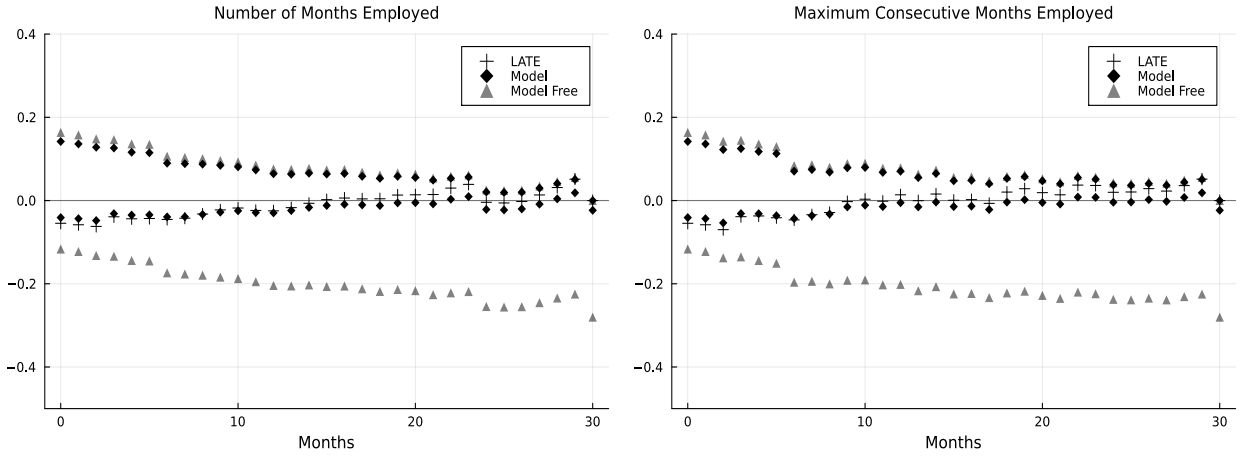

Figure 10: Distributional treatment effect for applicants who had graduated high school, were aged at least 30, and had no prior income during the year prior to assignment.

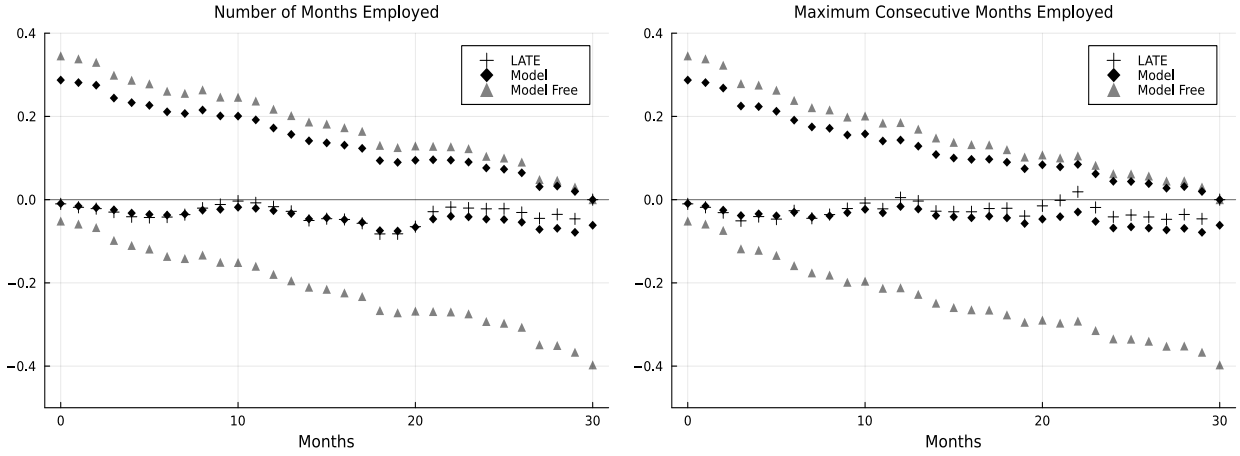

Figure 11: Distributional treatment effect for applicants who did not graduate high school, were aged less than 30, and had positive prior income during the year prior to assignment.

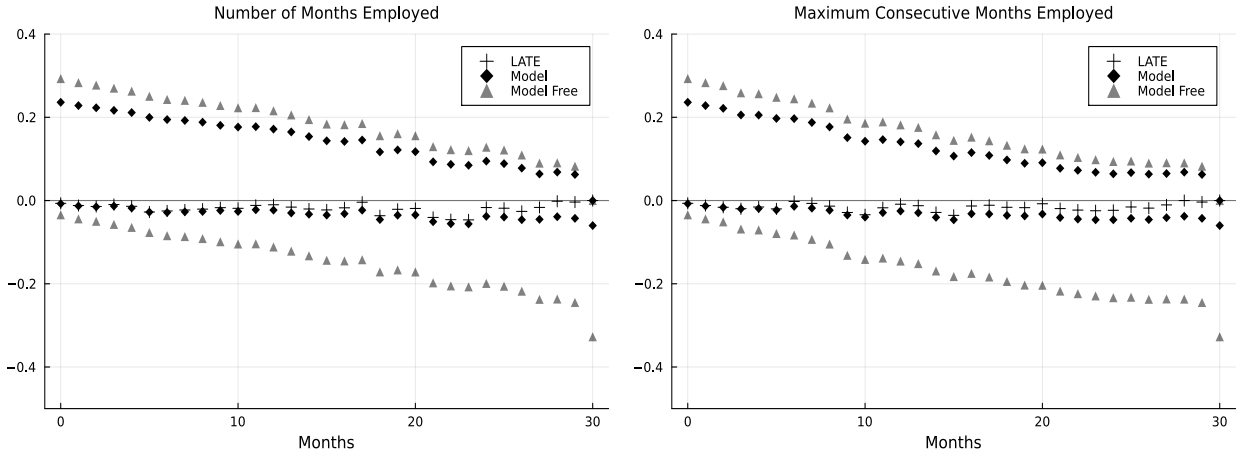

Figure 12: Distributional treatment effect for applicants who had graduated high school, were aged less than 30, and had positive prior income during the year prior to assignment.

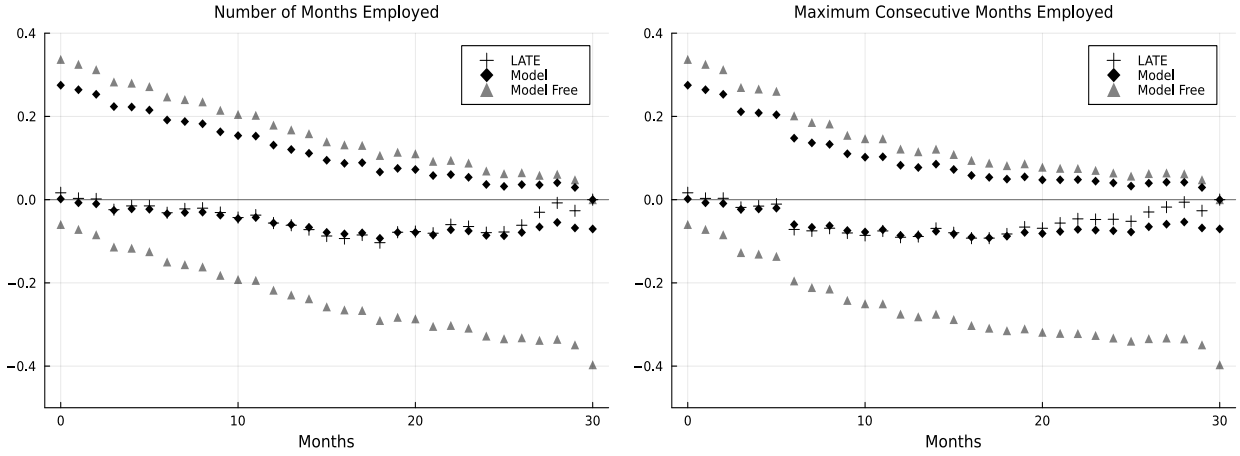

Figure 13: Distributional treatment effect for applicants who did not graduate high school, were aged at least 30, and had positive prior income during the year prior to assignment.

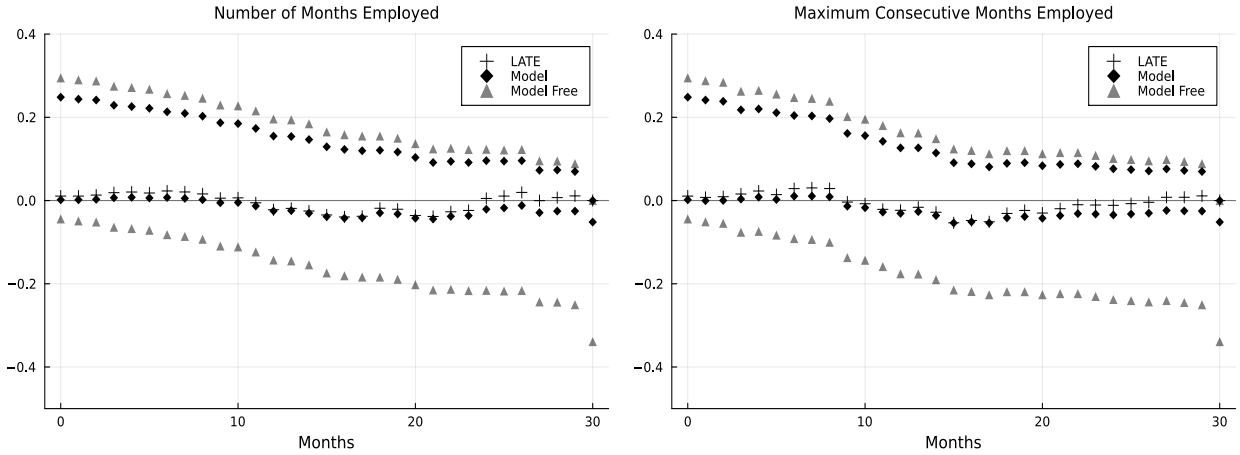

Figure 14: Distributional treatment effect for applicants who had graduated high school, were aged at least 30, and had positive prior income during the year prior to assignment.

E.2 Distribution Treatment Effects for Earnings

The following figures display treatment effects on the distribution of earnings over the 30 month period after treatment on various sub-groups of participants in the JTPA study. Agents are split into mutually exclusive groups based on their education, age and whether or not they reported any income in the year prior to being assigned to the treated or control group.

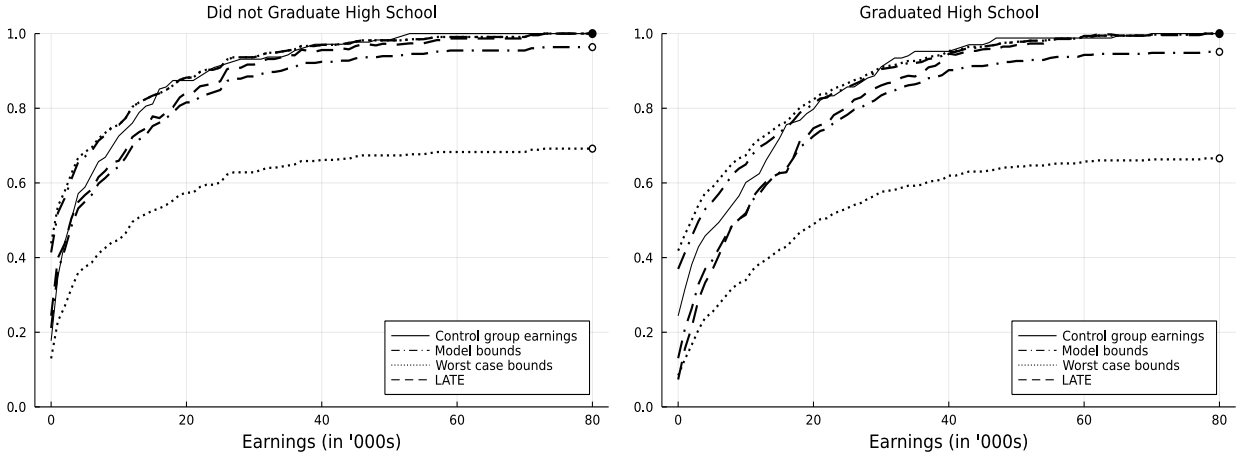

Figure 15: Treatment effect of training under JTPA on the distribution of earnings (in '000s) for agents who were aged at most 30 years, and reported no income in the year prior to time assignment.

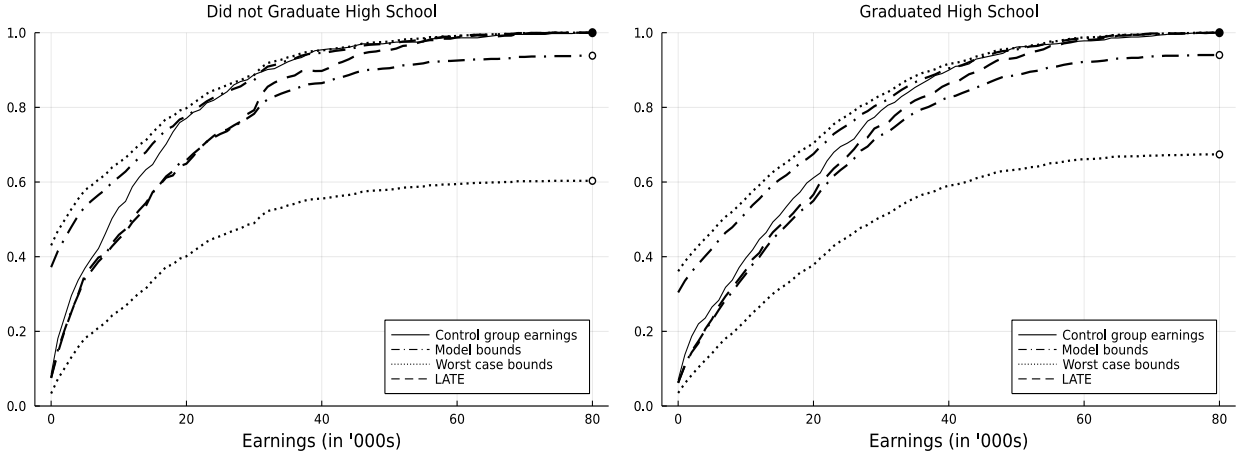

Figure 16: Treatment effect of training under JTPA on the distribution of earnings (in '000s) for agents who were aged at most 30 years, and reported positive income in the year prior to time assignment.

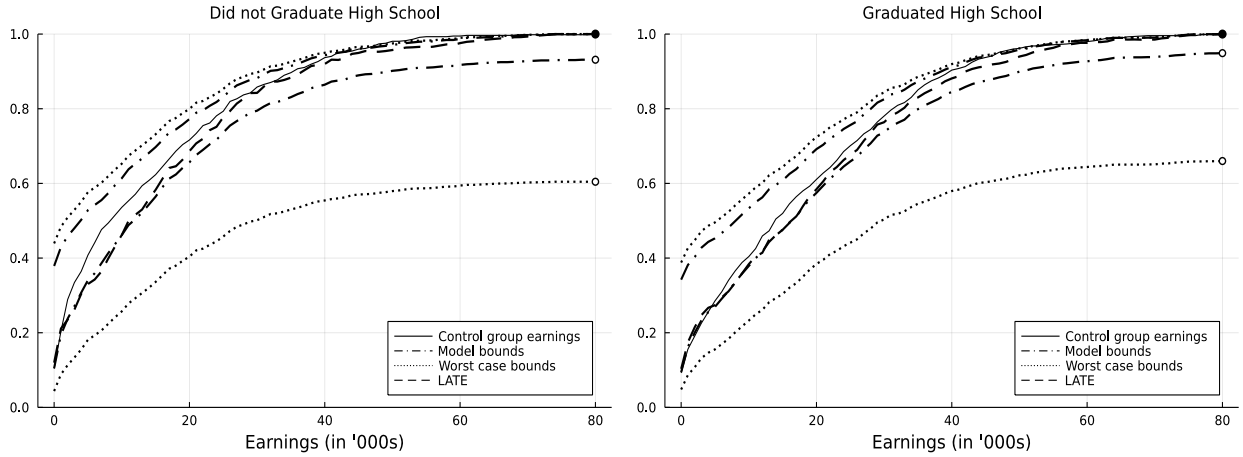

Figure 17: Treatment effect of training under JTPA on the distribution of earnings (in '000s) for agents who were aged at least 30 years, and reported positive income in the year prior to time assignment.