NBER WORKING PAPER SERIES

HOROWITZ-MANSKI-LEE BOUNDS WITH MULTILAYERED SAMPLE SELECTION

Kory Kroft
Ismael Mourifié
Atom Vayalinkal

Working Paper 32952
http://www.nber.org/papers/w32952

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
September 2024, Revised June 2026

This present version of the paper is June 3, 2026. We thank David Lee and seminar audiences at Arizona State University, Brown University, Carnegie Mellon University, National University of Singapore, New York University, Princeton University, Society of Labor Economics Meeting, University of Toronto, as well as the Rochester Labor Economics and Southern Economic Association conferences. We thank Stephen Claassen for providing outstanding research assistance. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Horowitz-Manski-Lee Bounds with Multilayered Sample Selection
Kory Kroft, Ismael Mourifié, and Atom Vayalinkal
NBER Working Paper No. 32952
September 2024, Revised June 2026
JEL No. C01, C21, C26, H0, H50, J0, J24, J32

ABSTRACT

This paper studies partial identification of treatment effects in the presence of sample selection, where treatment affects both selection into the sample and sorting across layers with heterogeneous outcomes. We show that canonical Lee bounds identify a total effect that combines the within-layer causal effect of treatment with a sorting effect reflecting outcome differences across layers. We derive sharp bounds on the within-layer causal effect using a two-step approach that extends Horowitz and Manski (1995) to a system of mixture equations with cross-equation dependence. Further, we show that under additional restrictions, these within-layer effects are sufficient for welfare analysis. Two empirical applications to job training experiments illustrate the approach; our estimates show that even when Lee bounds are strictly positive, within-firm bounds can be tight around zero, suggesting that Lee bounds capture a pure sorting effect.

Kory Kroft
University of Toronto
Department of Economics
and NBER
kory.kroft@utoronto.ca

Ismael Mourifié
Washington University in St. Louis
Department of Economics
and NBER
ismaelyacoub@gmail.com

Atom Vayalinkal
University of Toronto
Department of Economics
atom.vayalinkal@mail.utoronto.ca

1. Introduction

Social scientists are often interested in estimating the causal effect of a treatment Z on an outcome Y when, (i) the outcome is observed only for a selected subpopulation $D = 1$ and (ii) treatment affects selection into that subpopulation. A large literature has developed methods to address the selection bias that arises from conditioning on observed outcomes. A prominent approach, introduced by Lee (2009), partially identifies the causal effect of Z on Y for the subpopulation of always-selected individuals.

In many empirical settings however, selection occurs along multiple margins. In labor economics, job training can affect earnings directly through human capital accumulation and indirectly through sorting to heterogeneous firms or occupations with different wage premia. In education, college admission affects earnings both through attainment and through sorting to schools or majors with heterogeneous value-added. In health economics, insurance eligibility affects health outcomes both directly and through sorting to different physicians or hospitals. In immigration, refugee resettlement policies affect earnings directly and through assignment to locations with heterogeneous labor market conditions.

In each case, the selection problem is *multilayered*: the outcome is observed only for a selected subpopulation, and treatment affects both selection into that subpopulation and sorting within it.[1] Existing frameworks do not accommodate multilayered selection, nor do they provide tools to disentangle the distinct channels through which treatment affects outcomes. This gap has practical consequences: in a survey of papers published in top-5 general-interest journals that cite Lee (2009), we find that 6 of 42 papers implementing Lee bounds feature multilayered selection but nonetheless collapse the selection problem to a single dimension to apply standard methods.[2]

This paper fills this gap by developing a general framework for partial identification of treatment effects in settings with multilayered selection. We contribute to the literature in three ways. First, we extend the standard sample selection model to

[1]By *multilayered*, we do not mean a hierarchical or dynamic structure; rather, we use this terminology to refer to the case where selection is polychotomous.

[2]Further details on our literature survey are provided in the Supplemental Appendix.

a setting where selection is multilayered, and show that Lee bounds set identify a total effect that combines a weighted average of the causal effect of Z on Y across D (we label this the "within-layer effect") with a weighted average of the contrast in Y between different layers D for a fixed level of the treatment Z (we label this the "sorting effect").

Second, we derive sharp bounds on the within-layer effect. Our bounding approach proceeds in two steps. In the first step, we derive sharp closed-form bounds on the *response type* probabilities.[3] In deriving these bounds, we exploit a unique feature of our setting, which is that (unlike in the traditional instrumental variables framework) the exclusion restriction does not hold since treatment can have a direct causal effect on the outcome. We show that this feature implies that the distribution of response types does not depend on the distribution of Y (which may have continuous and/or unbounded support) and uses only the distribution on (D, Z) which has finite support (and can therefore be solved using a linear programming approach).

The second step provides closed-form bounds on the within-layer effect as a function of the sharp bounds on the response types derived in the first step. This step involves extending the Horowitz and Manski (1995) approach (which involves a single-equation mixture model with two components) to our setting which involves two mixture model equations with unknown weights that are interdependent across the equations. Importantly, we show that while this two-step approach provides an easy and tractable way to construct closed-form bounds, it does not entail any loss of information and provides sharp bounds. We also consider a set of additional restrictions on response types and show that they naturally lead to tighter bounds.

While understanding the role of D in mediating the effect of Z on Y is often of interest to shed light on mechanisms, a natural question is whether the within-layer effect is policy relevant. When both the treatment and the layer are directly manipulable, this makes the within-layer effect a natural object for mechanism-based policy design. Even when the layer is not directly manipulable, the parameter remains informative because it reveals whether the effectiveness of the policy depends on access to particular layers, an issue that is central for scaling or transporting the policy to

[3]The *response type* represents the pair of layers that an individual would choose if she were externally assigned to the control group and the treatment group, respectively.

new environments.[4] The policy relevance of this parameter motivates our welfare analysis: under additional restrictions, we derive sharp bounds on the welfare gain that depend exclusively on the within-layer effect, not the total effect inclusive of sorting. This connects to the literature on "sufficient statistics" for welfare analysis (Chetty 2009, Kline and Walters 2016, and Hendren and Sprung-Keyser 2020), and constitutes our third contribution.

As a proof of concept, we consider the causal effect of job training on wages where the layer corresponds to a firm type. Our empirical application focuses on two randomized experiments. The first experiment is based on the Job Corps Study and builds on the evaluation in Lee (2009). Our main finding is that in the cases where conventional Lee bounds are strictly positive (i.e., $[0.047, 0.048]$), our multilayered bounds for the within-firm wage effect, which hold the sorting effect constant, include zero. This suggests that Lee bounds may capture a pure sorting effect of job training rather than a direct wage effect. Our second empirical application is based on the WorkAdvance experiment recently examined in Katz et al. (2022). This is a sectoral employment program that targets high-quality jobs in specific industries with strong labor demand. Consistent with the Job Corps Study results, we find that Lee bounds are strictly positive and tight, while the within-firm wage effects include zero. We discuss the welfare implications of these findings.

The remainder of the paper is organized as follows. Section 2 introduces our multilayered sample selection model and defines the key causal estimands of interest. Section 3 discusses the causal interpretation of Lee bounds in the presence of multilayered sample selection and presents a general decomposition. Section 4 derives the sharp bounds on the within-layer causal effect in the multilayered sample selection model and sharp bounds for the welfare gain of job training. Section 5 presents empirical applications that implement the sharp bounds for Job Corps and WorkAdvance. The last section concludes. All the proofs are presented in the Appendix and additional results are presented in the Supplementary Appendix.

[4]For example, President Biden's workforce training initiative, the *American Rescue Plan's Good Jobs Challenge*, explicitly prioritized job quality and was designed to ensure that workers gained access to good jobs. This is also important for training programs that are aimed at "reskilling" workers, i.e., training them to work in different occupations.

Related Literature. Our paper builds on and contributes to the following literatures. First, it relates to econometric approaches that address sample selection. The Heckman (1979) sample selection model has been extended along various dimensions. Lee (2009) extends Heckman (1979) by relaxing the exclusion restriction of instrumental variables and derives bounds on the parameters of interest. Honoré and Hu (2020) study a semiparametric version of Lee's model, and Semenova (2020) and Olma (2021) propose inference methods for Lee bounds conditional on (potentially continuous) covariates. To our knowledge, this is the first paper to extend Heckman (1979) to a multilayered setting.

Second, one can view the layer D as a "mediator" in the context of mediation analysis (Robins and Greenland, 1992; Pearl, 2001). Most of this literature ignores sample selection where the outcome is unobserved at some mediator values. For example, recently Kwon and Roth (2024) develop a test for the presence of a mediator but abstract from sample selection. A rare exception is Zuo et al. (2022), who study identification of direct and indirect effects within a mediation analysis framework when both the outcome and the mediator are missing. However, they focus on point identification under strong assumptions including the non-falsifiable completeness condition.[5] In particular, their framework rules out the case where observability of the outcome depends on the mediator, which is central in our setting. Our paper complements Zuo et al. (2022) by establishing partial identification of the direct and indirect effects without imposing completeness, and allows for an endogenous mediator and the outcome to be missing non-randomly, even conditional on covariates. Our assumptions are transparent and apply directly to the primitives of our model. We also contribute to the mediation literature by conducting a welfare analysis and deriving informative bounds using revealed preference restrictions.[6]

Third, there is a large literature on active labor market programs reviewed in Heckman et al. (1999) and Card et al. (2010, 2018b). Our contribution is to examine whether worker sorting to firms affects the wage impacts of job training. Andersson et al. (2022) find suggestive evidence of that training affects firm characteristics and

[5]For an in-depth review of completeness, see D'Haultfoeuille (2011) and Canay et al. (2013).

[6]For further discussion, see the Supplemental Appendix which establishes a formal connection between our model and the literature on mediation analysis.

industry of employment. Katz et al. (2022) find substantial earnings gains from sector-based training programs and interpret them as partly driven by sorting to higher-paying industries, but do not provide a framework for isolating sorting as a causal mechanism. Schochet et al. (2008) document positive impacts of Job Corps on sorting to jobs with better amenities but do not disentangle these effects from the impact of Job Corps on employment itself.

Finally, our paper relates to the literature that has documented firm heterogeneity in wages and worker-firm sorting. Firms have been shown to be important for wage inequality (Abowd et al. 1999), the cyclicality of wages and early career progression (Card et al. 2013), the earnings losses of displaced workers (Lachowska et al. 2020; Schmieder et al. 2023), and gender (Card et al. 2016) and racial wage gaps (Gerard et al. 2021). Our contribution to this literature is to examine the role of firms in understanding the wage effect and sorting impact of job training. We do not impose any assumptions on potential wages, such as additive separability in worker and firm effects, nor do we impose exogenous mobility.

2. Analytical Framework

2.1. **Binary Sample Selection.**

To study the causal effect of job training on wage rates, Lee (2009) considered the following extension of Heckman's (1979) seminal sample-selection model:

$$Y = \begin{cases} (Y_1 - Y_0)Z + Y_0, & \text{if } D = 1, \\ \text{unobserved}, & \text{if } D = 0, \end{cases} \tag{2.1}$$

$$D = D_1 Z + D_0(1 - Z), \tag{2.2}$$

where D is a binary sample-selection indicator, equal to 1 if the individual is employed and 0 otherwise. The wage rate Y is observed only when $D = 1$. For each $z \in \{0,1\}$, Y_z denotes the potential outcome that would be realized under treatment assignment z, while D_z denotes the corresponding potential selection status.

The vector (Y_1, Y_0, D_1, D_0) collects the latent variables of the model, X is a vector of observed exogenous covariates, and $Z \in \{0,1\}$ is a binary treatment indicator satisfying:

$$(Y_1, Y_0, D_1, D_0) \perp Z \mid X.$$

Thus, conditional on X, treatment assignment is independent of the latent variables. This would hold, for example, if job training were randomly assigned.

In Lee (2009), job training raises human capital and directly affects wages. The contrast $Y_1 - Y_0$ is the individual causal effect of training on the wage rate. Lee shows how to partially identify the average causal effect for a specific subpopulation (the always-employed) when sample selection is present (i.e., when training can affect labor supply through $D_1 \neq D_0$).

2.2. **Multilayered Sample Selection.** A key assumption in Lee (2009) is that sample selection problem is binary: individuals are either employed or unemployed. However, job training can also affect worker-firm sorting. In this case, the selection problem is multilayered. We therefore generalize the sample selection model (2.1, 2.2) to allow for a richer model of labor supply where individuals choose layers (firms) and refer to it as the *multilayered selection model*:

$$Y = \begin{cases} (Y_{1,K} - Y_{0,K})Z + Y_{0,K} & \text{if } D = K, \\ \vdots & \vdots \\ (Y_{1,1} - Y_{0,1})Z + Y_{0,1} & \text{if } D = 1, \\ \text{unobserved} & \text{if } D = 0, \end{cases} \tag{2.3}$$

$$D = D_1 Z + D_0(1 - Z) \,. \tag{2.4}$$

Each layer D represents a distinct firm, and $Y_{z,d}$ denotes the potential wage when an agent is externally assigned to training status $z \in \{0, 1\}$ and firm $d \in \{0, 1, ..., K\}$, where $d = 0$ indicates non-employment so that $Y_{z,0}$ is unobserved.[7] While we focus on firms as the primary layer, our framework naturally applies to other settings with multilayered sample selection.[8] Throughout, we assume that for all $z \in \{0, 1\}$ and

[7]In our empirical application, we assume that the layer corresponds to a firm's type, where the type is constructed based on a firm's observable characteristics.

[8]Throughout, we use "within-layer" and "within-firm" interchangeably.

$d \in \{1, \ldots, K\}$, $Y_{z,d}$ is integrable and has a density with respect to some common dominating σ-finite measure μ.[9]

2.3. **Response Types.** In the context of sample selection, the outcomes are observed only when $D \neq 0$. We can partition the population into four groups: $\{D_0 = 0, D_1 = 0\} \cup \{D_0 > 0, D_1 = 0\} \cup \{D_0 = 0, D_1 > 0\} \cup \{D_0 > 0, D_1 > 0\}$. For the first three groups, at least one potential outcome is always missing, so without missing-at-random, selection-on-observables, or parametric assumptions, the observed data are uninformative about treatment effects. We do not impose such restrictions and instead focus exclusively on the subpopulation $\{D_0 > 0, D_1 > 0\}$.

Because individuals may select into different layers under treatment and control, it is useful to further partition the subpopulation $\{D_0 > 0, D_1 > 0\}$ into exogenous *response types*: $\{D_0 > 0, D_1 > 0\} = \bigcup_{d,d' \in \{1,\ldots,K\}} \{D_1 = d, D_0 = d'\}$.[10] A response type is defined as the pair of firms that an individual would select if she were externally assigned to the control group and the treatment group, respectively. Formally, the response type is the random variable $T = (D_0, D_1)$ with support $\mathcal{T}$.

2.4. **Target Parameter.** In the multilayered selection model, $Y_{1,d} - Y_{0,d}$ is the individual causal effect of job training on the wage rate at firm d. We refer to this as the *within-firm effect* since it holds the firm d fixed. Our target parameter of interest is the average causal effect of job training on wages within a specific firm d for a given response type $T = t$:

$$\mathbb{E}[Y_{1,d} - Y_{0,d} | T = t], d \in \{1, ..., K\}, \text{ and } t \in \mathcal{T} \tag{2.5}$$

In the mediation literature, the unconditional version $\mathbb{E}[Y_{1,d} - Y_{0,d}]$ is commonly referred to as the *controlled direct effect*; see, for example, Robins and Greenland (1992) and Pearl (2001). Analyzing the conditional version allows the within-firm effect to vary across response types, in a way that parallels the distinction between the average treatment effect (ATE) and local average treatment effect (LATE) in the

[9] Note that this does not imply that $Y_{z,d}$ is continuous, since μ is allowed to be any arbitrary σ-finite measure, and therefore can be chosen to dominate discrete, continuous, or mixed distributions.

[10] See Heckman and Pinto (2018) for a detailed discussion of the advantages of such a partition.

IV literature. In settings where the instrument corresponds to the policy of interest, the local parameter may be more policy relevant than the unconditional version.

2.5. **Policy Relevance of within-layer effect.** The policy relevance of the within-layer effect stems from the fact that it corresponds to a joint intervention on the treatment and the layer, or mediator, in the terminology of the mediation literature. It answers the question: would the treatment still affect outcomes if the layer were held fixed at a policy-relevant value? When the policymaker can directly intervene on the treatment and the layer, this interpretation is immediate. In this case, the within-layer effect is useful for mechanism-based policy design.

Even when the layer is not directly manipulable, as with firm assignment, the parameter remains informative: it reveals whether the policy's effectiveness depends on the availability of particular firms, which is especially important for scaling or transporting the policy to new environments. For example, if training effects are concentrated in firms offering flexible working arrangements, scaling up the program without ensuring access to such firms may substantially reduce its effectiveness.

Finally, we formally show in Section 4.3 that the within-layer effects are the key ingredients for welfare analysis: under appropriate conditions, they suffice for evaluating the welfare consequences of job training. This provides motivation for the separation of within-firm effects and sorting effects.

Remark 1. *Our framework remains valid even when there is no sample selection (i.e. Y is always observed, $P(D=0)=0$). In this case, our approach generalizes the IV model to settings where the instrument does not satisfy the exclusion restriction.*

3. The Causal Interpretation of Lee's bounds in the presence of Multilayered Sample Selection

To connect the generalized multilayered selection model (2.3)-(2.4) to the one in Lee (2009), it is useful to rewrite it equivalently as:

$$\begin{aligned} Y &= \begin{cases} Y_{1,D_1}Z + Y_{0,D_0}(1-Z) & \text{if } D>0, \\ \text{unobserved} & \text{if } D=0, \end{cases} \\ \mathbf{1}\{D>0\} &= \mathbf{1}\{D_1>0\}Z + \mathbf{1}\{D_0>0\}(1-Z)\ , \end{aligned}$$

where $Y_{z,D_{z'}} \equiv \sum_{d=0}^{K} Y_{z,d}1\{D_{z'} = d\}$ for $z, z' \in \{0,1\}$. This reduces to the setting of Lee (2009), in the special case where potential outcomes do not depend on the layer, i.e. $Y_{z,d} = Y_z$ for $d \in \{1, ..., K\}$.

We first introduce the following two assumptions maintained in Lee (2009). The first is a conditional independence assumption, which will be maintained throughout the remainder of this paper.

Assumption 1 (Conditional Random Assignment)**.** *Z is randomly assigned conditional on X, i.e.* $\{(Y_{z,d}, D_z) : d \in \{0,1,...,K\}, z \in \{0,1\}\} \perp Z|X$.

An implication of Assumption 1 is that the response type T is independent of Z.

The next assumption is Lee's monotonicity restriction, which we impose when studying Lee bounds and in our empirical applications.[11]

Assumption 2 ((Conditional) Lee's Monotonicity Assumption)**.** *We impose the following restriction:* $\mathbb{P}\left[\mathbf{1}\{D_1 > 0\} \geq \mathbf{1}\{D_0 > 0\} \middle| X\right] = 1$ *a.s.*

This assumption requires that treatment never increases selection into the unemployment layer $D = 0$, i.e., anyone who would be employed under control ($Z = 0$) must also be employed under treatment ($Z = 1$). Formally, Assumption 2 restricts the response type support, such that $\mathbb{P}[T = (d, 0)] = 0$ for $d \in \{1, ..., K\}$. Since monotonicity is imposed conditional on X, the direction can be allowed to vary across covariate values without affecting identification, though inference methods must be adapted accordingly, especially when X is continuous. For further details on such adaptations, see Słoczyński (2020) and Semenova (2020).

All remaining analysis implicitly conditions on $X = x$ for some value x of the vector of observed covariates, X. We suppress this dependence for notational ease.

Lemma 1 (Lee (2009) Bounds)**.** *Under* (2.3)-(2.4) *and Assumptions 1 and 2, with* $\mathbb{P}(D > 0|Z = 0) > 0$*, sharp bounds on* $\mathbb{E}[Y_{1,D_1} - Y_{0,D_0}|D_0 > 0, D_1 > 0]$ *are given by:*

$$\underline{\theta}^{\ell} \leq \mathbb{E}[Y_{1,D_1} - Y_{0,D_0}|D_0 > 0, D_1 > 0] \leq \overline{\theta}^{\ell} \tag{3.1}$$

[11]This assumption is not required by our general multilayered framework but can be sharply incorporated when desired; see the discussion following Lemma 3 below.

where $\underline{\theta}^\ell$ *and* $\overline{\theta}^\ell$ *are defined as in Appendix A.3. In particular, with* $p \equiv \frac{\mathbb{P}(D>0|Z=0)}{\mathbb{P}(D>0|Z=1)}$ *and* $F_W^{-1}(u) \equiv \inf\{w \in \mathbb{R} : \mathbb{P}(W \leq w) \geq u\}$ $\forall u \in [0,1]$*:*
(i) if Y *is continuous,*

$$\underline{\theta}^\ell \equiv \mathbb{E}[Y|D>0, Z=1, Y \leq F^{-1}_{Y|D>0,Z=1}(p)] - \mathbb{E}[Y|D>0, Z=0], \tag{3.2}$$

$$\overline{\theta}^\ell \equiv \mathbb{E}[Y|D>0, Z=1, Y \geq F^{-1}_{Y|D>0,Z=1}(1-p)] - \mathbb{E}[Y|D>0, Z=0], \tag{3.3}$$

(ii) if Y *is binary,*

$$\underline{\theta}^\ell \equiv \max\left\{0, 1 - \frac{1}{p}P[Y=0|D>0, Z=1]\right\} - \mathbb{E}[Y|D>0, Z=0], \tag{3.4}$$

$$\overline{\theta}^\ell \equiv \min\left\{1, \frac{1}{p}\mathbb{P}[Y=1|D>0, Z=1]\right\} - \mathbb{E}[Y|D>0, Z=0]. \tag{3.5}$$

In Appendix A.3 we provide universal expressions for the bounds (i.e., for any type of outcome variable), from which the two special cases follow.

Lemma 1 shows that in the presence of heterogeneous firms, Lee's identification approach bounds the estimand $\mathbb{E}[Y_{1,D_1} - Y_{0,D_0}|D_0 > 0, D_1 > 0]$. What is the causal interpretation of this estimand? The following lemma sheds light on this.

Lemma 2 (Decomposition)**.** *Assuming the generalized multilayered sample selection model, we have the following decomposition:*

$$\begin{aligned}
&\mathbb{E}[Y_{1,D_1} - Y_{0,D_0}|D_0 > 0, D_1 > 0] \\
&\quad = \sum_{d=1}^{K}\sum_{d'=1}^{K} \mathbb{E}[Y_{1,d} - Y_{0,d}|T=(d',d)] \times \mathbb{P}[T=(d',d)|D_0>0, D_1>0] \\
&\quad + \sum_{d=1}^{K}\sum_{d'=1:d\neq d'}^{K} \mathbb{E}[Y_{0,d} - Y_{0,d'}|T=(d',d)] \times \mathbb{P}[T=(d',d)|D_0>0, D_1>0]
\end{aligned} \tag{3.6}$$

Lemma 2 provides a general decomposition showing that, in the presence of firm heterogeneity, Lee bounds target a *total effect* which aggregates two components (conditional on $D_0 > 0$ and $D_1 > 0$). The first component is a sample-selection analogue of a within-firm training effect:

$$\text{(a)} \quad \sum_{d=1}^{K}\sum_{d'=1}^{K} \mathbb{E}[Y_{1,d} - Y_{0,d} \mid T=(d',d)]\ \mathbb{P}[T=(d',d) \mid D_0>0, D_1>0],$$

which averages the causal effect of job training on wages *holding the firm d fixed*, weighted by the share of response types that choose firm d under training. The second component captures a sorting (or firm-composition) effect:

$$\text{(b)} \quad \sum_{d=1}^{K} \sum_{d'=1:\, d\neq d'}^{K} \mathbb{E}[Y_{0,d} - Y_{0,d'} \mid T = (d', d)]\, \mathbb{P}[T = (d', d) \mid D_0 > 0, D_1 > 0]\,,$$

which averages wage differences across firms in the no-training counterfactual, weighted by the fraction of response types who sort into different firms. Accordingly, without additional assumptions, Lee bounds do not separately identify the *within-firm wage effect* of training from the *labor-supply/sorting channel* operating through D.

Lemma 2 also suggests special cases where Lee bounds admit a sharper interpretation. First, if wages do not vary across firms—i.e. $Y_{z,d} = Y_{z\bullet}$ (equivalently, no mediation via firm choice), as in Lee (2009)—the sorting component disappears and the total effect collapses to the causal effect of job training on the wage rate, which is Lee's target parameter:

$$\mathbb{E}[Y_{1,D_1} - Y_{0,D_0} | D_0 > 0, D_1 > 0] = \mathbb{E}[Y_{1\bullet} - Y_{0\bullet} | D_0 > 0, D_1 > 0]\ ,$$

Second, if job training has no direct effect on wages—i.e., $Y_{z,d} = Y_{\bullet d}$—Lee's bounds capture only the effect of training operating through *sorting into different firms*:

$$\mathbb{E}[Y_{1,D_1} - Y_{0,D_0} | D_0 > 0, D_1 > 0] = \mathbb{E}[Y_{\bullet D_1} - Y_{\bullet D_0} | D_0 > 0, D_1 > 0]\ .$$

Taken together, these results show that Lee bounds identify the direct wage effect of training only if mediation through firm choice can be ruled out, which is a restriction Lee's framework cannot test. The next section develops an alternative approach that overcomes this limitation.

4. Sharp bounds in the multilayered sample selection model

In this section, we partially identify the target parameters $\mathbb{E}[Y_{1,d} - Y_{0,d} | T = t]$. Let $f_{Y_{z,d}|D,Z}(y|d', z')$ denote the conditional density of $Y_{z,d}$ given $\{D = d', Z = z'\}$ and assume that it is absolutely continuous with respect to a dominating measure μ on the support of $Y_{z,d}$. We note that $f_{Y_{z,d},D|Z}(y, d|z) \equiv f_{Y_{z,d}|D,Z}(y|d, z)\mathbb{P}(D = d|Z = z)$.

For $d, d' \in \{1, ..., K\}$ and $z \in \{0, 1\}$, and any $y \in \mathcal{Y}$ we have the following:

$$f_{Y|D=d,Z=z}(y) = f_{Y_{z,d}|D_z}(y|d) = \sum_{d'=0}^{K} \frac{\mathbb{P}(D_z = d, D_{1-z} = d')}{\mathbb{P}(D = d|Z = z)} \times f_{Y_{z,d}|D_z,D_{1-z}}(y|d, d') \quad (4.1)$$

where the first equality holds under Assumption 1. More precisely, under Assumption 1, the following system of equations characterize the empirical content of the *multilayered selection model*:

$$f_{Y,D=d|Z=1}(y) = \sum_{d'=0}^{K} \mathbb{P}[T = (d', d)] \times f_{Y_{1,d}|T}(y|d', d) \quad (4.2)$$

$$f_{Y,D=d|Z=0}(y) = \sum_{d'=0}^{K} \mathbb{P}[T = (d, d')] \times f_{Y_{0,d}|T}(y|d, d') \quad (4.3)$$

and this holds for any $d, d' \in \{1, ..., K\}$ and $y \in \mathcal{Y}$. The left-hand side of equations (4.2) and (4.3) are observed while the individual types $\mathbb{P}[T = (d, d')]$ and the conditional potential outcome distributions $f_{Y_{z,d}|T}(y|d', d)$ on the right-hand side are unknown. For a given d, the system of equations (4.2)-(4.3) is under-determined: the number of unknowns $2K + 1 + 2(K + 1)|\mathcal{Y}|$ exceeds the number of equations $2|\mathcal{Y}|$, so the parameters are set identified.[12] When $\mathcal{Y}$ is finite, the system can in principle be solved by linear programming, but this approach is computationally intractable for large or continuous support —as in our empirical application — and provides little identification intuition.[13]

We instead develop a two-step identification approach. The first step derives sharp bounds on response-type probabilities using only the distribution on (D, Z) which has finite support (and can therefore be solved using a linear programming approach) and is independent of Y (which may have continuous and/or unbounded support). The second step derives closed-form bounds on the treatment effects as functions of the sharp bounds on the response-type probabilities. We show that these two steps provide sharp bounds on our target parameters of interest.

[12]The identified set of unknown parameters could naturally shrink if the researcher is willing to impose additional assumptions, such as Assumption 2 or the additional ones we introduce below.

[13]If the researcher is interested in analyzing a discrete outcome and wishes to explore this avenue further, she could employ the inferential method developed by Fang et al. (2023).

4.1. **Step 1: Sharp bounds on response-type probabilities.** In this step, we focus on the partial identification of the distribution of the response-type vector T, as captured by the vector of response-type probabilities $(\mathbb{P}[T=(d,d')] : d,d' \in \{0,...,K\})$.

Integrating equations (4.2) and (4.3) over $\mathcal{Y}$, we obtain the following system of equations for each d:

$$\mathbb{P}(D=d|Z=1) = \sum_{d'=0}^{K} \mathbb{P}[T=(d',d)] \tag{4.4}$$

$$\mathbb{P}(D=d|Z=0) = \sum_{d'=0}^{K} \mathbb{P}[T=(d,d')] \tag{4.5}$$

In the standard IV model, the distribution of the response types depends on the full joint distribution of the observed data (Y,D,Z), not just the distribution of (D,Z).[14] This complexity arises because the IV framework imposes the exclusion restriction, $Y_{z,d} = Y_{\bullet d}$. When this restriction holds, the response-type conditional density of $Y_{\bullet d}$ appears in both equations (4.2) and (4.3), so integrating each equation separately can lead to a loss of information on the response-type probabilities and non-sharp bounds. Without the exclusion restriction, each response-type conditional density $f_{Y_{z,d}|T}$ in the system of equations (4.2) and (4.3) appears in only one equation, so the integration step can be performed without losing any information on the response-type probabilities. Therefore, the response-type probabilities are entirely characterized by the distribution of (D,Z) which justifies proceeding in two steps. We say that a vector $\mathbf{v}$ satisfies (4.4, 4.5) if $(\mathbb{P}[T=(d,d')] : d,d' \in \{0,...,K\}) = \mathbf{v}$ is a solution to (4.4, 4.5) for all d.

Lemma 3. *Under the model* (2.3)-(2.4) *and Assumption 1, the (sharp) identified set for response-type probabilities is the set of non-negative vectors that satisfy* (4.4)-(4.5).

The approach we propose can incorporate other restrictions on response types instead of, or in addition to, Assumption 2. In the remainder, we will denote by $\mathcal{R}_T$ the

[14]This has been pointed out by Huber et al. (2017) and is also implicit in the results of Kitagawa (2021). See Theorem 3 in Vayalinkal (2024) for a result characterizing the relationship between the outcome distributions and the identified set of response-type probabilities.

restrictions on the response-type probability vector, $(\mathbb{P}[T=(d,d')] : d,d' \in \{0,...,K\})$, specified by the researcher.

As mentioned in Lemma 3, equations (4.4, 4.5) sharply characterize the restrictions on the distribution of T imposed by the model (2.3)-(2.4). Therefore, the identified set for response-type probabilities under model (2.3)-(2.4), Assumptions 1, and response-type restrictions $\mathcal{R}_T$, denoted $\Theta_I(\mathcal{R}_T)$, is simply the set of non-negative vectors that jointly satisfy both $\mathcal{R}_T$ and (4.4, 4.5), i.e.,

$$\Theta_I(\mathcal{R}_T) \equiv \left\{ \mathbf{v} \in [0,1]^{(K+1)^2} : \begin{array}{l} (\mathbb{P}[T=(d,d')] : d,d' \in \{0,...,K\}) = \mathbf{v} \\ \text{satisfies } (4.4)-(4.5) \text{ and } \mathcal{R}_T. \end{array} \right\}.$$

Remark 2. *In the case where the restrictions imposed by $\mathcal{R}_T$ are linear, they can be seamlessly integrated into equations (4.4)-(4.5) as supplementary linear constraints, and optimization over $\Theta_I(\mathcal{R}_T)$ remains a linear program. Some examples of such linear restrictions include the following, given some fixed (potentially partial) ordering of the layers (e.g., firms):*

(i) *("Strong Monotonicity") $D_1 \geq D_0$ with probability 1 or, equivalently, $\mathbb{P}[T = (d,d')] = 0$ for all $d > d' \in \{1,\ldots,K\}$.*
(ii) *("More Upward Switchers than Downward Switchers")*
$\mathbb{P}[T=(d',d)] \geq \mathbb{P}[T=(d,d')]$ *for all* $d > d' \in \{1,\ldots,K\}$.
(iii) *("More Stayers than Downward Switchers"]*
$\mathbb{P}[T=(d,d)] \geq \mathbb{P}[T=(d,d')]$ *for all* $d > d' \in \{1,\ldots,K\}$.
(iv) *(No Switchers) $\mathbb{P}[T=(d,d')] = 0$ for all $d \neq d' \in \{1,\ldots,K\}$*

The researcher may impose these assumptions in place of Assumption 2, or in addition to it. For example, researchers may choose $\mathcal{R}_T = \{$Assumption 2$\}$, $\mathcal{R}_T = \{$Strong Monotonicity$\}$, or $\mathcal{R}_T = \{$Assumption 2, No Switchers$\}$, etc.

The model and assumptions impose testable restrictions on the joint distribution of observables which are characterized by the following lemma:

Lemma 4. *Let some response-type restriction $\mathcal{R}_T$ be given. The model* (2.3)-(2.4), *Assumption 1 and $\mathcal{R}_T$ are jointly rejected by the data if and only if $\Theta_I(\mathcal{R}_T) = \emptyset$.*

Lemma 4 has two practical implications. First, falsification of the model and response-type restrictions $\mathcal{R}_T$ requires information on only the joint distribution

(D, Z), not the outcome distribution. In the leading case where $\mathcal{R}_T$ consists of linear restrictions, $\Theta_I(\mathcal{R}_T)$ is a finite-dimensional convex polytope, and so testing whether $\Theta_I(\mathcal{R}_T) = \emptyset$ reduces to checking feasibility of a linear program. This test can be implemented using existing methods for testing the feasibility of linear systems or linear moment inequalities (see, e.g., Fang et al. (2023) and Andrews and Soares (2010)). Since we do not need to solve an infinite-dimensional problem involving the outcome distribution, the test is much simpler than comparable tests in settings where an exclusion restriction holds; for example, implementing a test of our model (and any additional linear restrictions $\mathcal{R}_T$) is much simpler than the sharp tests for instrument validity and monotonicity proposed by Kitagawa (2015) and Mourifié and Wan (2017). Second, this computational reduction is not conservative: despite depending only on the distribution of (D, Z), the test is sharp. If the joint distribution of (D, Z) is such that some feasible response-type distribution $\boldsymbol{p} \in \Theta_I(\mathcal{R}_T)$ exists, then there also exist response-type-specific joint potential outcome distributions that, together with $\boldsymbol{p}$, rationalize the joint distribution of (Y, D, Z).

Remark 3 (Parametric $\mathcal{R}_T$: Logit and Probit)**.** *A complementary way to restrict response types is to impose a parametric selection model. Suppose there exists constants η_d, θ_d, for $d \in \{0, \ldots, K\}$, with $\eta_0 = \theta_0 = 0$, such that*

$$D_z \in \underset{d \in \{0,\ldots,K\}}{\operatorname{argmax}} \{\eta_d z + \theta_d + V_d\} \ ,$$

with $(V_d : d \in \{0, \ldots, K\})$ independent of Z. If the V_d are i.i.d. type-I generalized extreme value (Gumbel) distributed, then this imposes a multinomial logit discrete choice model on D_z; if V_d are i.i.d. standard Normal, the model is multinomial probit. In both cases, the observed joint distribution of (D, Z) (i.e. choice probabilities) point-identifies (θ_d, η_d) for all layers (under standard regularity conditions), and therefore, the full vector of response-type probabilities is also point-identified. Including such assumptions in $\mathcal{R}_T$ leads to the special case where $\Theta_I(\mathcal{R}_T)$ is a singleton.

To simplify the exposition in the next step, we introduce the shorthand notation, $p_{d,d'} \equiv \mathbb{P}[T = (d, d')]$, and $\gamma^z_{d_0,d_1} \equiv \frac{p_{d_0,d_1}}{\mathbb{P}(D=d_z|Z=z)}$. When $\Theta_I(\mathcal{R}_T) \neq \emptyset$, let $\underline{p}^r_{d,d'}$ denote the infimum of $p_{d,d'}$ over all distributions in $\Theta_I(\mathcal{R}_T)$. Subsequently, we can define: $\underline{\gamma}^{z,r}_{d_0,d_1} = \frac{\underline{p}^r_{d_0,d_1}}{\mathbb{P}(D=d_z|Z=z)}$. The superscript "$r$" reflects the dependence on $\mathcal{R}_T$. Computation of $\underline{\gamma}^{z,r}_{d,d'}$ is trivial when $\Theta_I(\mathcal{R}_T)$ is a singleton, as in the parametric choices of $\mathcal{R}_T$ considered in Remark 3. On the other hand, for all the potential choices of $\mathcal{R}_T$ considered in Remark 2, $\Theta_I(\mathcal{R}_T)$ is the set of nonnegative solutions to a linear system. Therefore, in such cases, $\underline{\gamma}^{z,r}_{d,d'}$ for $d, d' \in \{0, ..., K\}$ and $z \in \{0, 1\}$ can be obtained as a solution to a linear program. Since the linear system of interest here is generally small, it is also possible to obtain an analytic solution for $\underline{p}^{r}_{d,d'}$ (and therefore for $\underline{\gamma}^{z,r}_{d,d'}$) via Fourier-Motzkin elimination.

4.2. **Step 2: Sharp bounds on the treatment effects.** Equations (4.2) and (4.3) express the observed conditional distribution of earnings $F_{Y|D,Z}(y|d, z)$ as a finite mixture of potential outcome distributions conditional on response types, $F_{Y_{z,d}|T}(y|l, l')$:

$$f_{Y|D=d,Z=1}(y) = \sum_{d'=0}^{K} \gamma^{1}_{d',d} \times f_{Y_{1,d}|T}(y|d', d) \tag{4.6}$$

$$f_{Y|D=d,Z=0}(y) = \sum_{d'=0}^{K} \gamma^{0}_{d,d'} \times f_{Y_{0,d}|T}(y|d, d') \tag{4.7}$$

The unknowns are the weights $\gamma^{z}_{d,d'}$ for $z \in \{0, 1\}$ and $d, d' \in \{0, ..., T\}$. In Lee (2009), Assumption 2 implies that the weights of interest are point identified, reducing identification to recovering the mean average of the mixture components. However, in our setting, point identification of $\gamma^{z}_{d,d'}$ fails under Assumption 2 and even the stronger assumptions considered in Remark 2, primarily because multiple layers generate many more response types than in Lee (2009). Nevertheless, we can still derive informative bounds on these weights, as described above. As discussed in Remark 3, parametric selection models such as multinomial logit or probit can restore point identification of these weights by selecting a unique element of $\Theta_I(\mathcal{R}_T)$. In the applications below, however, the resulting treatment effect bounds are often very close to those obtained under simpler nonparametric restrictions such as the "Strong Monotonicity" condition considered in Remark 2, which are strictly weaker than the restrictions imposed by multinomial logit or probit.

Horowitz and Manski (1995) derived sharp bounds on mixture components with unknown weights in a single-equation mixture model with two components. Cross and Manski (2002) extended these results to single-equation models with many components. Our setting differs in two ways: it involves a system of mixture equations indexed by $d \in \{1, \dots, K\}$, each with up to $(K+1)^2$ components, and the weights are shared across equations, introducing a cross-equation dependence absent in both papers. We derive the identified set for the weights and extend their approaches to obtain closed-form bounds on our parameters of interest. A key step in doing so draws on an insight from Lee (2009): for continuous outcomes, the Horowitz-Manski bounds admit a tractable closed-form representation as the mean of a truncated distribution. Our setting requires generalizing this in two directions— accommodating cross-equation dependence in the weights and allowing for discrete or mixed outcome distributions— which we address through a generalized truncated mean representation that holds regardless of the outcome distribution.

For any d and z, define: $\underline{\mathbb{E}}_{F^{-1}_{Y|D,Z}}(\gamma; d, z) \equiv \mathbb{E}[F^{-1}_{Y|D=d,Z=z}(U)|U \leq \gamma]$, and $\overline{\mathbb{E}}_{F^{-1}_{Y|D,Z}}(\gamma; d, z) \equiv \mathbb{E}[F^{-1}_{Y|D=d,Z=z}(U)|U \geq 1-\gamma]$, where $U \sim \text{Uniform}(0,1)$. Let y_L and y_U denote the lower and upper bounds of the support of Y, respectively.[15] When the outcome is continuously distributed $\underline{\mathbb{E}}_{F^{-1}_{Y|D,Z}}(\gamma; d, z)$ coincides with the truncated mean in Lee (2009) i.e.,

$$\mathbb{E}[F^{-1}_{Y|D=d,Z=z}(U)|U \leq \gamma] = \mathbb{E}[Y|D=d, Z=z, Y \leq F^{-1}_{Y|D=d,Z=z}(\gamma)].$$

For binary outcomes,

$$\mathbb{E}[F^{-1}_{Y|D=d,Z=z}(U)|U \leq \gamma] = \max\left\{0, 1 - \frac{1}{\gamma} P[Y=0|D=d, Z=z]\right\}.$$

This formulation thus nests the continuous case and extends naturally to discrete and mixed outcomes. The next theorem presents our sharp bounds on the within-firm effects.

Theorem 1. *Suppose model* (2.3)-(2.4) *and Assumption 1 hold. For response-type restrictions $\mathcal{R}_T$, with $\Theta_I(\mathcal{R}_T) \neq \emptyset$, the following statements hold under $\mathcal{R}_T$:*

[15] Note that these bounds need not be finite.

(i) *For each* $d \in \{1, \ldots, K\}$*, the following bounds are pointwise sharp*

$$\underline{\mathbb{E}}_{F^{-1}_{Y|D,Z}}(\underline{\gamma}^{1,r}_{d,d}; d, 1) - \overline{\mathbb{E}}_{F^{-1}_{Y|D,Z}}(\underline{\gamma}^{0,r}_{d,d}; d, 0) \leq \mathbb{E}[Y_{1,d} - Y_{0,d} \mid T = (d, d)] \leq \overline{\mathbb{E}}_{F^{-1}_{Y|D,Z}}(\underline{\gamma}^{1,r}_{d,d}; d, 1) - \underline{\mathbb{E}}_{F^{-1}_{Y|D,Z}}(\underline{\gamma}^{0,r}_{d,d}; d, 0) \ .$$

(ii) *For* $d, d' \in \{1, \ldots, K\}$*,* $d \neq d'$*, the following bounds are pointwise sharp*

$$\underline{\mathbb{E}}_{F^{-1}_{Y|D,Z}}(\underline{\gamma}^{1,r}_{d',d}; d, 1) - y_U \leq \mathbb{E}[Y_{1,d} - Y_{0,d} \mid T = (d', d)] \leq \overline{\mathbb{E}}_{F^{-1}_{Y|D,Z}}(\underline{\gamma}^{1,r}_{d',d}; d, 1) - y_L$$

and $y_L - \overline{\mathbb{E}}_{F^{-1}_{Y|D,Z}}(\underline{\gamma}^{0,r}_{d,d'}; d, 0) \leq \mathbb{E}[Y_{1,d} - Y_{0,d} \mid T = (d, d')] \leq y_U - \underline{\mathbb{E}}_{F^{-1}_{Y|D,Z}}(\underline{\gamma}^{0,r}_{d,d'}; d, 0)$.

(iii) *Let* $\boldsymbol{p} \equiv (p_{d,d'} : d, d' \in \{0, ..., K\})$*. For any* $\mathcal{D} \subseteq \{1, \ldots, K\}$ *and nonnegative weight functions* $w_d : \boldsymbol{p} \mapsto w_d(\boldsymbol{p}) \in \mathbb{R}_{\geq 0}$*, the following bounds are sharp*

$$\inf_{\boldsymbol{p} \in \Theta_I(\mathcal{R}_T)} \sum_{d \in \mathcal{D}} w_d(\boldsymbol{p}) \left(\underline{\mathbb{E}}_{F^{-1}_{Y|D,Z}}\left(\frac{p_{d,d}}{\mathbb{P}(D = d \mid Z = 1)}; d, 1 \right) - \overline{\mathbb{E}}_{F^{-1}_{Y|D,Z}}\left(\frac{p_{d,d}}{\mathbb{P}(D = d \mid Z = 0)}; d, 0 \right) \right)$$
$$\leq \sum_{d \in \mathcal{D}} w_d(\boldsymbol{p})\, \mathbb{E}[Y_{1,d} - Y_{0,d} \mid T = (d, d)]$$
$$\leq \sup_{\boldsymbol{p} \in \Theta_I(\mathcal{R}_T)} \sum_{d \in \mathcal{D}} w_d(\boldsymbol{p}) \left(\overline{\mathbb{E}}_{F^{-1}_{Y|D,Z}}\left(\frac{p_{d,d}}{\mathbb{P}(D = d \mid Z = 1)}; d, 1 \right) - \underline{\mathbb{E}}_{F^{-1}_{Y|D,Z}}\left(\frac{p_{d,d}}{\mathbb{P}(D = d \mid Z = 0)}; d, 0 \right) \right) .$$

The derivation of the bounds in Theorem 1 comes from extending the Horowitz and Manski (1995) bounding approach summarized in Lemma B.1 in Appendix B. However, demonstrating sharpness is considerably more complex. It requires showing that solving equations (4.2) to (4.3) for all $y \in \mathcal{Y}$ and $d \in \{1, ..., K\}$ subject to the restrictions defined in $\mathcal{R}_T$ yields exactly the closed-form bounds in Theorem 1. The absence of the IV exclusion restriction is what makes this result possible.

Theorem 1(i) indicates that, without additional assumptions on the potential outcome distributions, the derived bounds can potentially determine the direction (sign) of the within-firm effect at layer d solely for individuals who remain with firm d under any treatment assignment Z, i.e., $\mathbb{E}[Y_{1,d} - Y_{0,d}|T = (d, d)]$. This finding is somewhat intuitive given that these "stayers" are equivalent to the so-called "always-employed" in Lee's model, where firm heterogeneity is not taken into account. On the other hand, Theorem 1(ii) suggests that the bounds for those who switch firms due to treatment ("switchers"), such as $\mathbb{E}[Y_{1,d} - Y_{0,d}|T = (d, d')]$ and $\mathbb{E}[Y_{1,d} - Y_{0,d}|T = (d', d)]$ for $d \neq d'$, always include 0. This is because the observed data (Y, D, Z) do not reveal any information on the unobserved counterfactuals $\mathbb{E}[Y_{0,d}|T = (d', d)]$ and $\mathbb{E}[Y_{1,d}|T = (d, d')]$.

Theorem 1(iii) presents the closed form bounds that correspond to weighted averages of $\mathbb{E}[Y_{1,d} - Y_{0,d}|T = (d, d)]$, $\sum_{d\in\mathcal{D}} w_d(\boldsymbol{p})\, \mathbb{E}[Y_{1,d} - Y_{0,d} \mid T = (d, d)]$. These bounds are sharp and take into account the interdependence between equations (4.6) and (4.7). A leading special case is the case where the w_d are proportional weights, i.e. $w_d(\boldsymbol{p}) = \frac{p_{d,d}}{\sum_{d'\in\mathcal{D}} p_{d',d'}}$, in which case the parameter in Theorem 1(iii) is equivalent to $\mathbb{E}\left[Y_{1,D} - Y_{0,D} \,|\, T \in \{(d, d) : d \in \mathcal{D}\}\right]$.

Remark 4. *To derive bounds on the aggregate quantity* $\sum_{d\in\mathcal{D}} w_d(\boldsymbol{p})\, \mathbb{E}[Y_{1,d} - Y_{0,d} \mid T = (d, d)]$ *one might be tempted to adopt a naïve approach by taking a weighted average of the pointwise sharp bounds derived in Theorem 1 (i). However, this approach not only fails to provide sharp bounds on the aggregate quantity but may also yield* ***invalid*** *bounds. We discuss this in more detail in Supplemental Appendix A. The difference between the sharp bounds and naïve "bounds" is illustrated using our empirical applications in Section 5 (see Figures 2 and 4 below), and using simulations in Supplemental Appendix C (see Figure 3 there).*

The Supplemental Appendix specializes the framework to two firms (as in our empirical applications), derives closed-form bounds on response-type probabilities and, by substitution into Theorem 1(i)-(ii), analytic bounds on the target parameters. A numerical illustration demonstrates that our bounds can distinguish a positive within-firm effect from a zero effect even when Lee bounds are strictly positive.

4.3. **Bounds on Welfare.** In this section, we demonstrate that the within-firm effects are informative about the welfare impact of the treatment (i.e., job training). More precisely, we obtain sharp bounds on the welfare gain and show that the within-firm effects are "sufficient statistics" in the spirit of Chetty (2009), Kline and Walters (2016) and Hendren and Sprung-Keyser (2020).[16]

[16]Kleven (2021) reviews the sufficient statistics approach to policy evaluation and notes that it relies on strong assumptions; in particular, it requires that the policy change in small. Our approach is valid for small or large policy changes.

Assumption 3 (Common Choice Set)**.** *There exists a random vector* $\boldsymbol{\epsilon}$ *and a function* $U : \{0,1\} \times \{0,\ldots,K\} \times \operatorname{supp}(\boldsymbol{\epsilon}) \to \mathbb{R}$ *such that, for each* $z \in \{0,1\}$*,*

$$D_z \in \underset{\{0,\ldots,K\}}{\operatorname{argmax}}\, U(z,d,\boldsymbol{\epsilon}) \ \ \textit{almost surely}$$

and D_z *is* $\sigma(\boldsymbol{\epsilon})$*-measurable.*

Assumption 3 embeds workers' firm choice into a discrete choice framework and assumes the existence of a "latent type" structural variable $\boldsymbol{\epsilon}$, of unrestricted (potentally infinite) dimension, that captures all worker-specific factors relevant for a worker's choice of firm under each training status. Assumption 3 does not, however, require that the utility maximizing firm is unique (i.e., the selection model is allowed to be incomplete), instead only requiring that D_z is chosen from among the set of utility maximizing firms in a way that is measurable as a function of $\boldsymbol{\epsilon}$, i.e., that any tie-breaking rule relevant for choice of D_z is included in $\boldsymbol{\epsilon}$.

Under Assumption 3, we define individual welfare given $Z = z$, denoted $W(z)$, as

$$W(z) \equiv U(z, D_z, \boldsymbol{\epsilon}) \ .$$

To link welfare to wages, we impose an additive separability condition: utility consists of expected wages at firm d, which may depend on treatment status z, and a non-pecuniary firm-specific component (e.g., amenities) which does not vary with z.

Assumption 4 (Quasi-linear Utility)**.** *For each* $d \in \{1,\ldots,K\}$*, there exists a function* $h_d : \operatorname{supp}(\boldsymbol{\epsilon}) \to \mathbb{R}$ *such that, for each* $z \in \{0,1\}$ *and almost-every* $\boldsymbol{e} \in \operatorname{supp}(\boldsymbol{\epsilon})$*,*

$$U(z,d,\boldsymbol{e}) = \mathbb{E}\left[Y_{z,d} \,|\, \boldsymbol{\epsilon} = \boldsymbol{e}\,\right] + h_d(\boldsymbol{e}) \,.$$

Further, $U(1,0,\boldsymbol{\epsilon}) = U(0,0,\boldsymbol{\epsilon})$*.*

Assumption 4 is the key restriction linking welfare to the wage objects studied above. It introduces an arbitrary and layer-specific non-pecuniary term, $h_d(\boldsymbol{\epsilon})$, which may vary freely across individuals and layers, but not with treatment status alone. This implies that treatment affects utility *at a given layer* d only through changes in the expected wage component $\mathbb{E}\left[Y_{z,d} \,|\, \boldsymbol{\epsilon}\right]$ and not through treatment-specific non-pecuniary factors.[17]

Theorem 2 (Bounds on Welfare Effects)**.** *Under model* (2.3)-(2.4)*, Assumptions 1, 3 and 4, the following statements hold.*

(i) *Given* $\left(\left(Y_{z,d} : d \in \{1, \ldots, K\}, z \in \{0,1\}\right), D_0, D_1, Z\right)$ *satisfying the above assumptions and (arbitrary) response-type restriction* $\mathcal{R}_T$*, the following bounds on welfare are sharp*

$$\sum_{d=1}^{K} \mathbb{P}\left(D_0 = d\right) \mathbb{E}\left[Y_{1,d} - Y_{0,d} \mid D_0 = d\right] \leq \mathbb{E}\left[W\left(1\right) - W\left(0\right)\right] \leq \sum_{d=1}^{K} \mathbb{P}\left(D_1 = d\right) \mathbb{E}\left[Y_{1,d} - Y_{0,d} \mid D_1 = d\right].$$

(ii) *For the "stayers", we have the following identity*

$$\mathbb{E}\left[W\left(1\right) - W\left(0\right) \,|\, D_1 = D_0 > 0\right] = \sum_{d=1}^{K} \frac{\mathbb{P}\left(T = (d,d)\right)}{\sum_{d'=1}^{K} \mathbb{P}\left(T = (d',d')\right)} \mathbb{E}\left[Y_{1,d} - Y_{0,d} \mid T = (d,d)\right] .$$

The first part of Theorem 2 gives sharp revealed-preference bounds on welfare effects given the latent joint distribution of all potential outcomes and choices. It shows that even when this joint distribution is known, the sharp bounds on welfare depend only on the within-firm wage effects. The second part shows that, for stayers, the welfare gain is exactly pinned down by the aggregate of within-layer wage effects. These treatment effects therefore serve as "sufficient statistics" for welfare analysis: recovering them is not only of intrinsic interest for understanding the mechanism through which a treatment operates, but also necessary and sufficient for bounding the welfare effects of the treatment. In particular, even though utility depends on job amenities, these are not relevant for welfare analysis under the stated assumptions.

[17] The quasi-linear structure imposed by Assumption 4 is standard in empirical labor market models; see, among others, Card et al. (2018a), Dube et al. (2020), Lamadon et al. (2022), Azar et al. (2022), Chan et al. (2024), and Kroft et al. (2025). It is also a central feature of generalized Roy models; see D'Haultfoeuille and Maurel (2013) and Eisenhauer et al. (2015).

This connection provides an additional justification for focusing on within-firm effects as the primary target parameters of the analysis.

Remark 5. *Note that the bounds in Theorem 2(i) are not, in general, identified from the observed data: although $P(D_z = d)$ is identified, $\mathbb{E}[Y_{1,d} - Y_{0,d} \mid D_z = d]$ involve fixed-layer counterfactuals for individuals who may choose a different layer under the other treatment state (i.e., "switchers"). For the "stayer" welfare parameter in Theorem 2(ii), however, the welfare effect is exactly an aggregate of stayer within-layer wage effects, for which bounds are provided in Theorem 1(iii); Corollary 1 below shows that these bounds remain sharp under the additional conditions assumed here.*

The following result shows that Assumptions 3 and 4, while essential for the welfare interpretation suggested by Theorem 2, do not provide any additional information on the parameters considered in Theorem 1. Therefore, Theorem 2 implies that we can obtain sharp bounds on the welfare effects for stayers using Theorem 1.

Corollary 1. *The bounds in Theorem 1(i) and Theorem 1(iii) remain sharp under the conditions of Theorem 2. In particular, under the conditions of Theorem 2, the bounds in Theorem 1(iii), with $l := 1, l' := K$ and $w_d := \frac{\mathbb{P}(T=dd)}{\sum_{d'=1}^{K} \mathbb{P}(T=d'd')}$ for each $d \in \{1, \ldots, K\}$, are the sharp bounds on $\mathbb{E}[W(1) - W(0) \mid D_1 = D_0 > 0]$.*

4.4. **Inference.** The bounds in Theorem 1(i)-(ii) are known functions of conditional truncated means of Y given (D, Z), evaluated at truncation levels

$$\gamma_{d,d'}^{z,r} = \frac{p_{d,d'}^{r}}{\Pr(D = d \mid Z = z)}.$$

Therefore, subject to the regularity conditions in Lee (2009), Semenova (2020), or Olma (2021), inference for these bounds can be based on their estimators of truncated conditional expectations, after plugging in a suitable estimator for $\gamma_{d,d'}^{z,r}$.[18] These methods also allow conditioning on, and aggregating over, covariates X, which can substantially tighten the bounds relative to unconditional estimation.

[18]However, $\underline{\gamma}_{d,d'}^{z,r}$ can often be only directionally differentiable with respect to the propensity score vector, as is evident in the closed form expressions for the 2 firms type case provided in the Supplemental Appendix. In such cases, inference procedures that depend on the bootstrap will need to be adapted to remain valid (see Fang and Santos (2019) for details). Alternatively, we can consider a smoothed version of the bounds as entertained in Heiler et al. (2024).

Inference for the aggregate bounds in Theorem 1(iii) is more involved because the relevant truncation levels are chosen by an optimization problem involving the outcome distributions. We leave this case to future work.

5. Empirical Applications: Job Corps Study and WorkAdvance RCT

5.1. **Application #1: Job Corps Study.** This section implements our multilayered bounds for the Job Corps Study.[19] We use publicly available data from the National Job Corps Study (Schochet et al. 2003) and impose two sample restrictions. First, following Lee (2009), we drop individuals with missing earnings or hours in any post-assignment week, leaving 9,145 individuals (3,599 control, 5,546 treated). Second, we drop individuals with missing health insurance values in weeks 90, 135, 180, and 208. This yields a final sample of 6,403 (2,540 control, 3,863 treated).

The three key variables of interest are employment, hourly wage, and provision of health insurance for employed individuals. Following Lee (2009), employment is defined by positive weekly earnings and hourly wage as weekly earnings divided by weekly hours worked. We use the provision of health insurance to classify firm type, with H denoting firms that offer health insurance and L denoting firms that do not.[20]

The Supplementary Appendix presents summary statistics for our final sample. Firms offering health insurance pay higher wages than non-offering firms, and Job Corps assignment increases sorting to amenity-offering firms. This combined evidence suggests that sample selection is multilayered and motivates the implementation of our sharp bounds to these data.

[19]Job Corps is the largest residential career training program in the United States. The Job Corps study randomized access to first time applicants to the program between November 1994 and December 1995. For more details on Job Corps, the Job Corps Study and summary statistics, see the Supplemental Appendix. For impact evaluations, see Schochet et al. (2001), Schochet et al. (2008), Lee (2009), and Blanco et al. (2013).

[20]This classification is motivated by evidence that highlights a cross-sectional correlation between wages and amenities, see, e.g., Pierce (2001), Dey and Flinn (2005), Lamadon et al. (2022), and Maestas et al. (2023). Results for classifying firms based on other job amenities, including the provision of pension/retirement benefits and paid vacation, are available upon request. For recent methods that suggest approaches toward better classifications in related settings, see Heiler and Knaus (2023) and Yoshikawa and Kawano (2026).

5.1.1. *Multilayered Bounds for Job Corps Study.* As a first step, we replicate the bounds reported in Lee (2009). For week 90, we estimate bounds of $[0.0468, 0.0484]$.[21] Next, we estimate our multilayered bounds, starting with Assumptions 1 and 2 and then adding the restrictions highlighted in Remark 2 and Remark 3. Under Assumption 2, the always-employed (AE) definition used in Lee (2009) combines four different response types: $\{D_0 > 0, D_1 > 0\} = \{(L, L), (H, H), (L, H), (H, L)\}$. We focus on the bounds for stayers, defined as the response types (H, H) and (L, L).

Table 1 presents the estimated propensity scores from the National Job Corps Study.[22] As expected, $\mathbb{P}[D > 0|Z = 1] > \mathbb{P}[D > 0|Z = 0]$ showing that treated individuals are more likely to be employed. The table also shows that $\mathbb{P}[D = H|D > 0, Z = 1] > \mathbb{P}[D = H|D > 0, Z = 0]$ implying that individuals who receive Job Corps training are more likely to be employed in firms that offer health insurance than individuals who do not receive Job Corps training, conditional on employment.

Table 1. Job Corps: Propensity scores, by wk. Health insurance amenity.

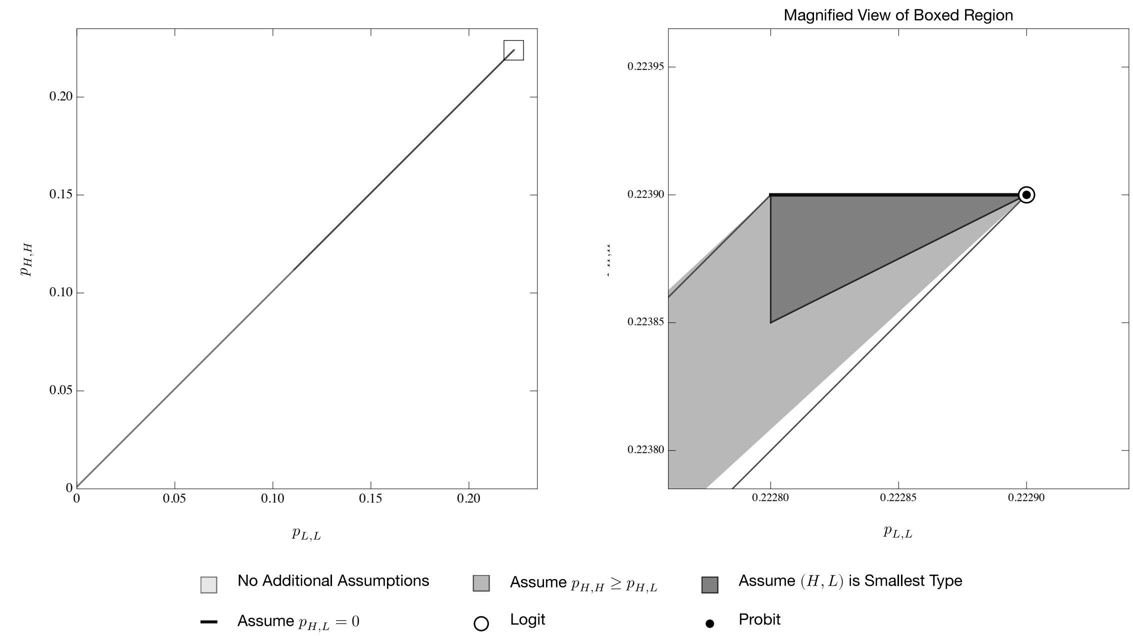
	$\mathbb{P}[D = H\|Z = 0]$	$\mathbb{P}[D = H\|Z = 1]$	$\mathbb{P}[D = L\|Z = 0]$	$\mathbb{P}[D = L\|Z = 1]$
Week 90	0.2239	0.2372	0.2361	0.2229
Week 135	0.2758	0.3037	0.2415	0.2414
Week 180	0.2941	0.3313	0.2462	0.2512
Week 208	0.3142	0.3559	0.2513	0.2509

Using the week 90 propensity scores from Table 1, Figure 1 presents the identified set for $(p_{L,L}, p_{H,H})$. Naturally, incorporating additional restrictions on the response types leads to a tightening of the identified sets.

Figure 2 presents our multilayered bounds for $\mathbb{E}[Y_{1,H} - Y_{0,H}|T = (H, H)]$ and $\mathbb{E}[Y_{1,L} - Y_{0,L}|T = (L, L)]$, along with our aggregate bounds, for weeks 90, 135, 180 and

[21]As detailed in the Supplementary Appendix, these are Lee's bounds when treating ln(hourly wage) as a continuous variable, as we do throughout. Lee (2009) uses vingtiles of ln(hourly wage) that produce bounds $[0.0423, 0.0428]$ in week 90.

[22]Our sample restriction to keep observations with non-missing health insurance provision drops only employed individuals, mechanically reduces the propensity scores. To ensure comparability with Lee (2009), we rescale our estimated propensity scores so that the probabilities of employment by treatment status are the same as those reported in Lee (2009).

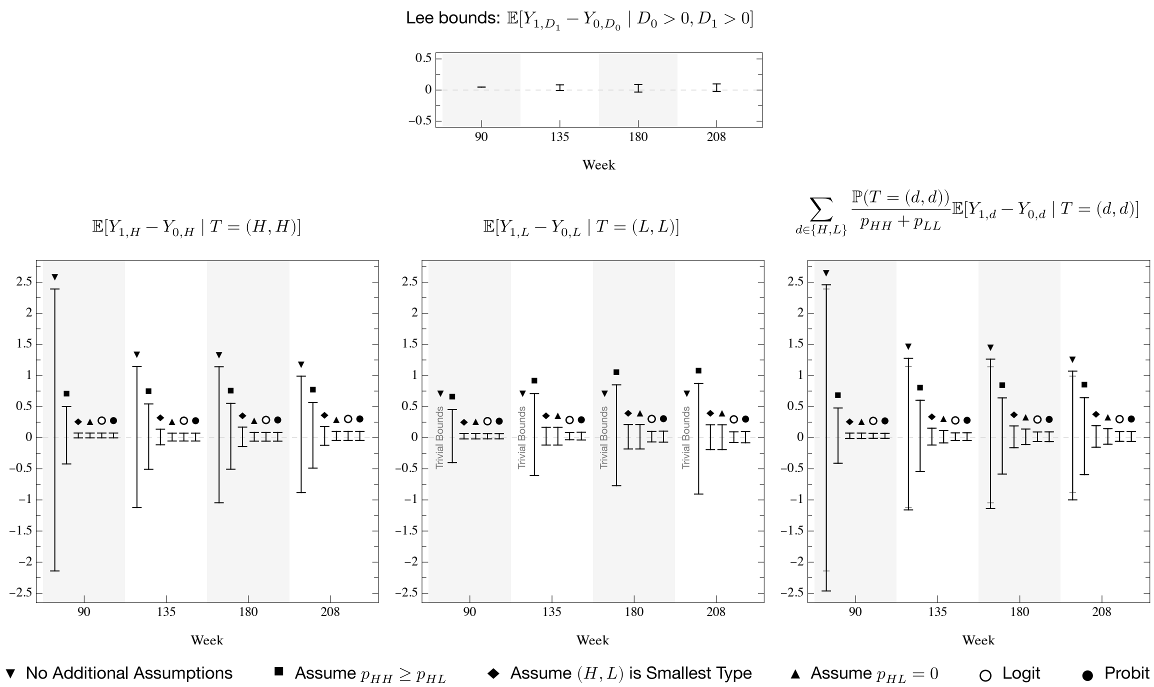


Notes: This figure plots the identified set for $(p_{L,L}, p_{H,H})$ under various assumptions, as indicated, for Week 90 of the Job Corps application. The right panel is a magnified view of the boxed region in the left panel. Lighter regions correspond to weaker restrictions and contain the darker regions. Under the Logit and Probit assumptions, $(p_{L,L}, p_{H,H})$ is point-identified (in this application, the identified point for $(p_{L,L}, p_{H,H})$ under Logit and Probit coincide).

FIGURE 1. Job Corps, Wk.90, Health Insurance Amenity: Id. set for $(p_{L,L}, p_{H,H})$.

208. For type H firms, our baseline estimates indicate $\mathbb{E}[Y_{1,H} - Y_{0,H} | T = (H, H)] \in [-2.1415, 2.3907]$. Adding restrictions naturally sharpens the bounds: requiring more stayers than downward switchers yields $[-0.4214, 0.5020]$; assuming (H, L) is the smallest response type narrows it to $[-0.0023, 0.0754]$; and imposing strong monotonicity results in $[-0.0018, 0.0750]$. Relative to strong monotonicity, parametric logit and probit specifications do not further tighten the bounds. We find a similar pattern for the L-type bounds with the main difference being that the logit and probit assumptions induce tightening relative to strong monotonicity. In summary, our bounds for the within-firm effects include zero suggesting that Lee bounds may capture a pure

sorting response to job training rather than a direct wage effect. Further, by Theorem 2(ii), the aggregate bounds imply that the bounds on the welfare gain of Job Corps for stayers similarly include zero.



Notes: Outcome is ln(hourly wage); hourly wage calculated as weekly earnings divided by weekly hours for the employed. The first two panels provide the bounds for each firm-level effect and the last panel provides bounds on the weighted average of firm-level effects, by week. In all panels, the solid black lines indicate sharp bounds; in the final panel, the gray lines indicate the (generally invalid) result of the "naïve approach" of taking the weighted average of firm-level bounds (see Remark 4). Under the conditions of Theorem 2, these bounds can be interpreted as sharp bounds on welfare effects and, in particular, the final panel provides sharp bounds on $\mathbb{E}\left[W(1) - W(0) \mid D_0 = D_1 > 0\right]$. Tables containing the numerical intervals are available in the Supplementary Appendix.

FIGURE 2. Job Corps, Health Insurance Amenity: Multilayered Bounds

5.2. **Application #2: WorkAdvance RCT.** Our second empirical application evaluates MDRC's WorkAdvance sectoral employment program, which trains disadvantaged adults to match them with high-quality jobs in sectors with strong labor demand. While the broader WorkAdvance demonstration spanned four community-based providers across three locations (New York City, Tulsa, and Northeast Ohio), we focus on the Madison Strategies RCT in Tulsa, Oklahoma, and the Towards Employment RCT in Northeast Ohio.[23] Both evaluations shared an enrollment period running from June 2011 to June 2013. The Madison Strategies RCT targeted high-quality jobs in transportation and manufacturing, randomizing 697 individuals: 353 to the treatment group (program enrollment) and 344 to control. The Towards Employment RCT targeted positions in health care and manufacturing, enrolling 698 individuals evenly split between the treatment and control groups (349 each).

The key variables for our analysis are quarterly wages along with employment status and sector. We define quarterly wages as quarterly earnings subject to UI, and quarterly employment based on whether an individual has positive earnings in a quarter. We set $d = H$ if the firm of employment is in the target sector and $d = L$ if the firm is not. All our results focus on outcomes 8 quarters post-random assignment.

The Supplementary Appendix presents summary statistics for both sites. They demonstrate that target-sector firms pay higher wages than non-target firms, and WorkAdvance induces sorting to these firms.

5.2.1. *Multilayered Bounds for WorkAdvance RCTs.* Table 2 presents propensity score estimates for WorkAdvance. For Madison Strategies, overall employment rates are similar between the treatment (0.67) and control (0.66) groups but, among the employed, target-sector employment differs sharply: 44 percent for the treated group versus 31 percent for the control group. Towards Employment increased both overall

[23]Our primary dataset is state-level administrative Unemployment Insurance (UI) data, obtained via a confidential data use agreement with MDRC. The administrative UI data for Oklahoma (Madison Strategies RCT) and Ohio (Towards Employment RCT) provide the two-digit North American Industry Classification System (NAICS) code for the industry in which the participant worked, while the New York data (Per Scholas and St. Nicks Alliance RCTs) do not. Therefore, we focus on the Oklahoma and Ohio evaluations. For more details on the WorkAdvance program and summary statistics, see Supplemental Appendix E. For an impact evaluation, see Katz et al. (2022); for Madison Strategies, the reported impact is 12.4 percent on earnings two years after the program. For Towards Employment, the impact is 14 percent.

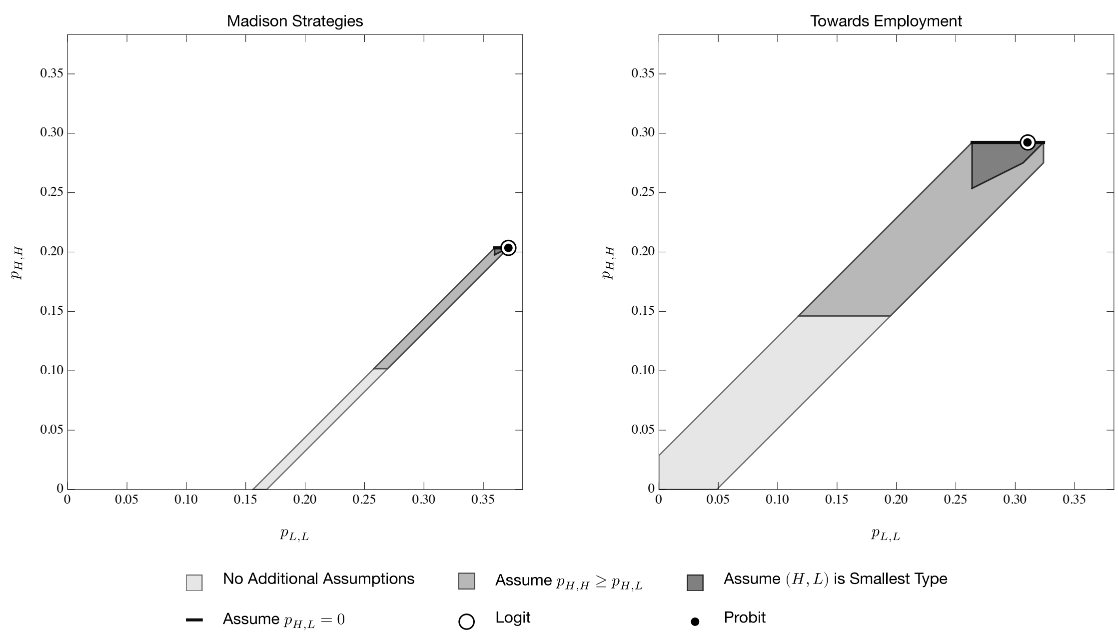
employment (0.62 to 0.69) and target-sector employment among the employed from 47 percent to 51 percent.

TABLE 2. Propensity scores for WorkAdvance RCTs

	$\mathbb{P}[D=H \mid Z=0]$	$\mathbb{P}[D=H \mid Z=1]$	$\mathbb{P}[D=L \mid Z=0]$	$\mathbb{P}[D=L \mid Z=1]$
Madison Strats.	0.2035	0.2975	0.4535	0.3711
Towards Emp.	0.2923	0.3524	0.3238	0.3410

Figure 3 shows the identified sets for $(p_{L,L}, p_{H,H})$ for both WorkAdvance RCTs. Although the Madison Strategies and Job Corps studies differ in the degree of sorting and the share of never-employed workers, $p^*_{H,H}$ is quite similar.[24] Recall that

[24]Note that $p_{0,0}$ is point identified by $P(D=0|Z=1)$. In Job Corps, $p_{0,0}=0.54$ whereas $p_{0,0}=0.33$ in Madison Strategies.

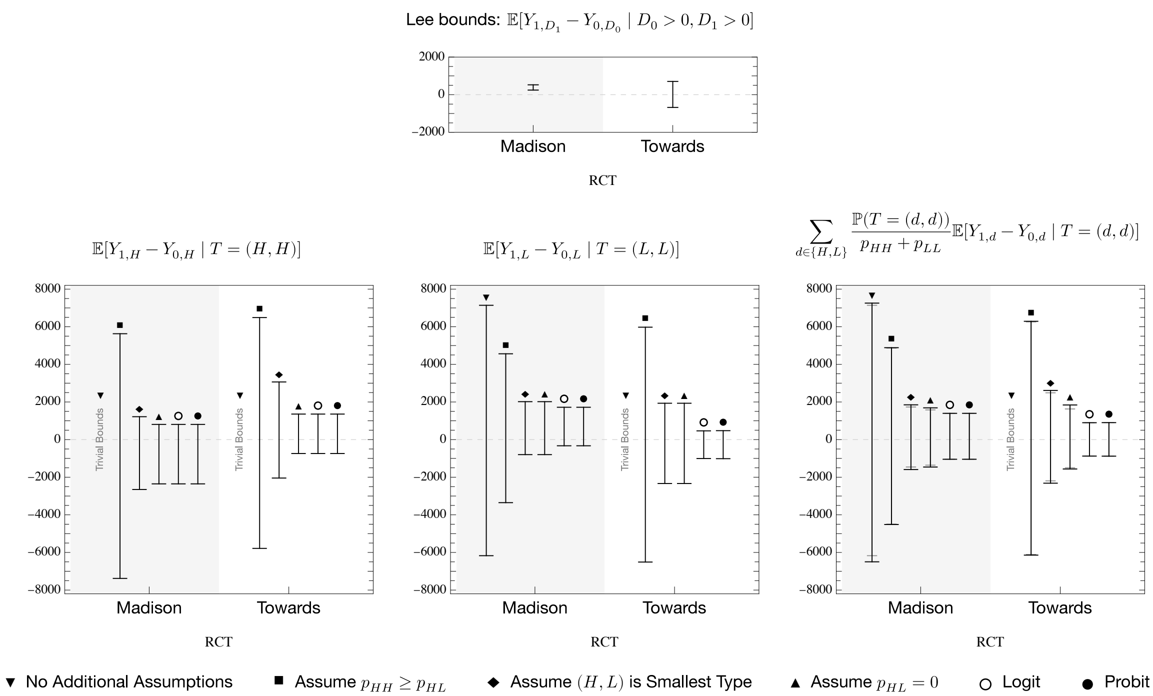


Notes: This figure plots the identified set for $(p_{L,L}, p_{H,H})$ under various assumptions, as indicated, for each of the WorkAdvance RCTs. Lighter regions correspond to weaker restrictions and contain the darker regions. Under the Logit and Probit assumptions, $(p_{L,L}, p_{H,H})$ is point-identified (in the Madison Strategies application, the identified point for $(p_{L,L}, p_{H,H})$ under Logit and Probit coincide, but they differ in the Towards Employment application).

FIGURE 3. WorkAdvance RCTs: Identified set for $(p_{L,L}, p_{H,H})$.

$p^*_{H,H}$ captures the mass of workers who always sort into high-type firms regardless of treatment. This is pinned down by $P(D = H \mid Z = 0)$, the share of the control group employed at high-type firms, and is comparable across studies. Under strong monotonicity, this mapping becomes exact as $P(D = H \mid Z = 0) = p^*_{H,H}$.

Figure 4 presents Lee bounds, our multilayered bounds for $\mathbb{E}[Y_{1,H} - Y_{0,H}|T = (H, H)]$ and $\mathbb{E}[Y_{1,L} - Y_{0,L}|T = (L, L)]$, along with our aggregate bounds, for both WorkAdvance RCTs. Our estimated Lee bounds for Madison Strategies are $[244.19, 524.98]$



Notes: Top panel illustrates Lee (2009) bounds, under Assumptions 1 and 2. Remaining panels illustrate the multilayered bounds under various sets of additional assumptions, for each WorkAdvance RCT: the first two panels provide the bounds for each firm-level effect, and the final panel provides bounds on the weighted average of firm-level effects. The solid black intervals indicate sharp bounds; in the final panel, the gray intervals indicate the (generally invalid) result of the "naïve approach" of taking the weighted average of firm-level bounds (see Remark 4). Under the conditions of Theorem 2, these bounds can be interpreted as sharp bounds on welfare effects and, in particular, the final panel provides sharp bounds on $\mathbb{E}\left[W(1) - W(0) \,|\, D_0 = D_1 > 0\right]$. Tables containing the numerical intervals are available in the Supplementary Appendix.

FIGURE 4. Bounds for the WorkAdvance RCTs

and our bounds for Towards Employment are $[-676.87, 705.66]$. Our within-firm bounds include 0 for both sites under all assumptions. This suggests that, for Madison Strategies, Lee bounds may be capturing sorting into the high-wage target sector. As in the case of Job Corps, we cannot rule out zero welfare gains among stayers.

6. Conclusion

This paper develops a new methodology to partially identify the causal effect of job training on wages in the presence of multilayered sample selection. We define new treatment effects that operate within and between firms and provide a new identification approach that extends Horowitz and Manski (1995) and Lee (2009) bounds. As a proof of concept, we show how to empirically implement these bounds by considering applications to the Job Corps Study and the WorkAdvance randomized experiments. Although we consider our approach in the context of job training with firms as layers, it applies to any setting with multilayered sample selection.

Throughout the paper, we have assumed that the layer is known to the econometrician and measured without error. In some empirical settings, however, the layer may be latent or mismeasured; extending our framework to such settings is deferred to future work.

Appendix A. Proofs of Main Results

We work with a complete non-atomic probability space, $(\Omega, \Sigma, \mathbb{P})$, on which all random variables in this paper are defined; following the usual convention, we identify the sample space Ω with the population of individuals. We assume that the probability space $(\Omega, \Sigma, \mathbb{P})$ is sufficiently rich for our purposes; this assumption can always be satisfied after suitably enriching $(\Omega, \Sigma, \mathbb{P})$, if needed, by taking product spaces.[25]

We first provide a self-contained proof of Lemma 3, and use this result to prove Lemma 4. The remaining proofs in this section rely on both Lemmas 3 and 4 and the auxiliary results in Appendix B, the proofs of which rely only on Lemmas 3 and 4.

[25] See, e.g., Halmos (1976), p. 157.

A.1. **Proof of Lemma 3.**

Proof. By the definition of T and law of total probability, the response-type probabilities satisfy (4.4)-(4.5). It suffices to show that given $\left(\boldsymbol{p}_{(d,d')} : (d',d) \in \{0,\ldots,K\}^2\right) \geq 0$ such that $\sum_{d=0,d'=0}^{K,K} \boldsymbol{p}_{(d,d')} = 1$ and, for every $d \in \{0,\ldots,K\}$,

$$\mathbb{P}(D=d|Z=1) = \sum_{d'=0}^{K} \boldsymbol{p}_{(d',d)} \text{ and } \mathbb{P}(D=d|Z=0) = \sum_{d'=0}^{K} \boldsymbol{p}_{(d,d')} \ , \tag{A.1}$$

there exists a distribution Q of $\left(\left((Y_{0,d},Y_{1,d}) : d \in \{0,\ldots,K\}\right), D_0, D_1, Z\right)$ such that

$$\left(\mathbb{P}_Q\left[T=(d,d')\right] : (d',d) \in \{0,\ldots,K\}^2\right) = \left(\boldsymbol{p}_{(d,d')} : (d',d) \in \{0,\ldots,K\}^2\right) \ ,$$

and Q induces a distribution of (Y,D,Z) under model (2.3)-(2.4) and Assumption 1, that is consistent with the observed data. Since Y is not observed when $D=0$, set $Y|D=0,Z=1$ and $Y|D=0,Z=0$ to arbitrary distributions, so that we can treat (Y,D,Z) as observed. We will now construct a Q that induces this distribution.

Define $\mathcal{Y}_{z,d} \equiv \operatorname{supp}(Y|D=d,Z=z)$ and, for each $d \in \{0,\ldots,K\}$, $z \in \{0,1\}$, and $(d',d'') \in \{0,\ldots,K\}^2$ define the CDF $F_{(d',d'')}^{(z,d)}$ as

$$F_{(d',d'')}^{(z,d)}(y) = P\left(Y \leq y | D=d, Z=z\right) \ ,$$

and note that it does not depend on (d',d''). Next, for every $(d',d'') \in \{0,\ldots,K\}^2$, let $C_{(d',d'')}$ be an arbitrary copula of dimension $|\{0,1\} \times \{0,\ldots,K\}|$. Define Q as

$$Q\left(y,t,z\right) \equiv C_t\left(\left(F_t^{(z,d)}\left(y_{(z,d)}\right) : (z,d) \in \{0,1\} \times \{0,\ldots,K\}\right)\right) \times \boldsymbol{p}_{(t)} \times P\left(Z=z\right) \ ,$$

for $y \in \prod_{(z,d)\in\{0,1\}\times\{0,\ldots,K\}} \mathcal{Y}_{z,d}$, $t \in \{0,\ldots,K\}^2$, and $z \in \{0,1\}$, where $Q\left(y,t,z\right)$ is shorthand for

$$Q\left(y,t,z\right) = Q\left(\left(Y_{z,d} : (z,d) \in \{0,1\} \times \{0,\ldots,K\}\right) \leq y, (D_0,D_1) = t, Z = z\right).$$

By construction, Q satisfies Assumption 1 and $\left(\mathbb{P}_Q\left[T=(d,d')\right] : (d',d) \in \{0,\ldots,K\}^2\right) = \left(\boldsymbol{p}_{(d,d')} : (d',d) \in \{0,\ldots,K\}^2\right)$. The result now follows immediately by noting that Q induces the observed data distribution under model (2.3)-(2.4) since, for any

$y \in \mathcal{Y}_{z,d}, d \in \{0, \ldots, K\}$ and $z \in \{0, 1\}$, we have that

$$\begin{aligned} Q\left(Y_{z,d} \leq y, D_z = d, Z = z\right) &= Q\left(Y_{z,d} \leq y | D_z = d, Z = z\right) Q\left(D_z = d | Z = z\right) Q\left(Z = z\right) \\ &= Q\left(Y_{z,d} \leq y | D_z = d\right) Q\left(D_z = d\right) P\left(Z = z\right) \\ &= P\left(Z = z\right) P\left(Y \leq y | D = d, Z = z\right) \sum_{d'=0}^{K} \left((1-z)\boldsymbol{p}_{(d,d')} + z\boldsymbol{p}_{(d',d)}\right) \\ &= P\left(Z = z\right) P\left(Y \leq y | D = d, Z = z\right) P\left(D = d | Z = z\right) \\ &= P\left(Y \leq y, D = d, Z = z\right) \end{aligned}$$

where the second equality follows from Q satisfying Assumption 1, and the penultimate equality follows from $\boldsymbol{p}$ satisfying (A.1). □

A.2. **Proof of Lemma 4.**

Proof. Given Lemma 3 above, this result follows immediately from Theorem 3 in Vayalinkal (2024). Since our proof of Lemma 3 is constructive, however, we can also argue directly, as follows. Both parts below proceed by showing the contrapositive.

($\Longleftarrow$) If Assumption 1 and $\mathcal{R}_T$ are consistent with the data, then there exists a joint distribution Q of $\left(\left(Y_{0,d} : d \in \{0, \ldots, K\}\right), \left(Y_{1,d} : d \in \{0, \ldots, K\}\right), D_0, D_1, Z\right)$ that is consistent with the observed data distribution such that the response-type probabilities induced by Q is in $\Theta_I(\mathcal{R}_T)$ and so $\Theta_I(\mathcal{R}_T) \neq \emptyset$.

($\Longrightarrow$) Suppose that $\Theta_I(\mathcal{R}_T) \neq \emptyset$, then there exists $\boldsymbol{p} = \left(\boldsymbol{p}_{(d,d')} : (d', d) \in \{0, \ldots, K\}^2\right) \in \Theta_I(\mathcal{R}_T)$. Now, our proof of Lemma 3 shows that we can construct a joint distribution Q of $\left(\left(Y_{0,d} : d \in \{0, \ldots, K\}\right), \left(Y_{1,d} : d \in \{0, \ldots, K\}\right), D_0, D_1, Z\right)$ such that Q satisfies Assumption 1 and induces the observed data distribution under (**??**, **??**), and

$$\left(\mathbb{P}_Q\left[T = (d, d')\right] : (d', d) \in \{0, \ldots, K\}^2\right) = \left(\boldsymbol{p}_{(d,d')} : (d', d) \in \{0, \ldots, K\}^2\right),$$

which implies that Q also satisfies $\mathcal{R}_T$, as required, since $\boldsymbol{p} \in \Theta_I(\mathcal{R}_T)$. □

The remaining proofs in this section rely on Appendix B.

A.3. **Proof of Lemma 1.** We first provide the universal expressions for $\underline{\theta}^{\ell}$ and $\overline{\theta}^{\ell}$:

$$\underline{\theta}^{\ell} \equiv \mathbb{E}\left[F^{-1}_{Y|D>0,Z=1}(U)\,\middle|\,U \leq p\right] - \mathbb{E}[Y|D>0,Z=0], \tag{A.2}$$

$$\overline{\theta}^{\ell} \equiv \mathbb{E}\left[F^{-1}_{Y|D>0,Z=1}(U)\,\middle|\,U \geq 1-p\right] - \mathbb{E}[Y|D>0,Z=0]. \tag{A.3}$$

Proof of Lemma 1. Define $S_z \equiv \mathbf{1}\{D_z > 0\}$ and $\tilde{Y}_z \equiv Y_{z,D_z}$. Now, by Assumption 1, we have that $\left(\tilde{Y}_0, \tilde{Y}_1, D_0, D_1\right) \perp Z$ and, by Assumption 2, we have that $\mathbb{P}\left(S_1 \geq S_0\right) = 1$. Note that $p \equiv \frac{\mathbb{P}(D>0|Z=0)}{\mathbb{P}(D>0|Z=1)} = \frac{\mathbb{P}(S_0=1)}{\mathbb{P}(S_1=1)} = \mathbb{P}\left(S_0 = 1 \,|\, S_1 = 1\right)$, by Assumption 2.

Note that the observed outcome distribution given $Z = 1$ can be represented as

$$F_{Y|D>0,Z=1}(y) = pF_{\tilde{Y}_1|S_0=S_1=1}(y) + (1-p)F_{\tilde{Y}_1|S_0=0,S_1=1}(y) \ ,$$

and the bounds follow immediately from Lemma B.1 by noting that $\mathbb{E}\left[\tilde{Y}_0\,\middle|\,S_0 = S_1 = 1\right] = \mathbb{E}\left[\tilde{Y}_0\,\middle|\,S_0 = 1\right] = \mathbb{E}\left[Y \mid D > 0, Z = 0\right]$. Sharpness now follows from Lemma B.2 by the argument in the proof of Theorem 1 below. Now it suffices to show the special cases. The continuous case follows by noting that, for any continuous random variable W, $F_W(W) \sim$ Uniform$[0,1]$, and so, $\mathbb{E}\left[F_W^{-1}(U)\,\middle|\,U \leq p\right] = \mathbb{E}\left[W\,\middle|\,W \leq F_W^{-1}(p)\right]$, and analogously for the upper bound. The binary case follows by noting that

$$\mathbb{E}\left[F^{-1}_{Y|D>0,Z=1}(U)\,\middle|\,U \leq p\right] = \frac{1}{p}\int_0^p F^{-1}_{Y|D>0,Z=1}(u)\,du \ ,$$

$$\mathbb{E}\left[F^{-1}_{Y|D>0,Z=1}(U)\,\middle|\,U \geq 1-p\right] = \frac{1}{p}\int_{1-p}^1 F^{-1}_{Y|D>0,Z=1}(u)\,du \ ,$$

plugging in $F^{-1}_{Y|D>0,Z=1}(u) = \mathbf{1}\left\{u > \mathbb{P}\left(Y = 0 \,|\, D > 0, Z = 1\right)\right\}$, and simplifying. □

A.4. **Proof of Lemma 2.**

Proof. First, notice that

$$\begin{aligned}\mathbb{E}\left[Y_{1,D_1} - Y_{0,D_0}|D_0 > 0, D_1 > 0\right] = {} & \mathbb{E}\left[Y_{1,D_1} - Y_{0,D_1}|D_0 > 0, D_1 > 0\right] \\ & + \mathbb{E}\left[Y_{0,D_1} - Y_{0,D_0}|D_0 > 0, D_1 > 0\right] \ .\end{aligned} \tag{A.4}$$

Next, we have that

$$\begin{aligned}
&\mathbb{E}\left[Y_{1,D_1} - Y_{0,D_1} | D_0 > 0, D_1 > 0\right] \\
= &\sum_{d=1,d'=1}^{K,K} \mathbb{E}\left[Y_{1,D_1} - Y_{0,D_1} | D_0 = d', D_1 = d\right] \mathbb{P}\left(D_0 = d', D_1 = d | D_0 > 0, D_1 > 0\right) \\
= &\sum_{d=1,d'=1}^{K,K} \mathbb{E}\left[Y_{1,d} - Y_{0,d} | D_0 = d', D_1 = d\right] \mathbb{P}\left(D_0 = d', D_1 = d | D_0 > 0, D_1 > 0\right) \ ,
\end{aligned}$$

and similarly,

$$\begin{aligned}
&\mathbb{E}\left[Y_{0,D_1} - Y_{0,D_0} | D_0 > 0, D_1 > 0\right] \\
= &\sum_{d=1,d'=1}^{K,K} \mathbb{E}\left[Y_{0,d} - Y_{0,d'} | D_0 = d', D_1 = d\right] \mathbb{P}\left(D_0 = d', D_1 = d | D_0 > 0, D_1 > 0\right) \\
= &\sum_{d=1,d'=1,d\neq d'}^{K,K} \mathbb{E}\left[Y_{0,d} - Y_{0,d'} | D_0 = d', D_1 = d\right] \mathbb{P}\left(D_0 = d', D_1 = d | D_0 > 0, D_1 > 0\right) \ ,
\end{aligned}$$

and plugging these values back into (A.4) immediately implies the result. □

A.5. **Proof of Theorem 1.**

Proof. Since the two mixtures given by (4.2) and (4.3) do not share any components, Lemma B.2 reduces the problem of finding bounds on the conditional expectation of $Y_{z,d}$ given T to the problem of finding bounds on expectations of mixture components. Therefore, we now complete the proof using the results given in Lemma B.1, as follows.

Proof of (i)-(ii): For any type (d', d) and any $z \in \{0, 1\}$, the weighted conditional density $\mathbb{P}[T = (d', d)] \times f_{Y_{z,T(z)}|T}(y|d', d)$ only appears in at most one of (4.2) and (4.3). Therefore, sharp bounds on the expectation of any such component can be obtained as the bounds of Horowitz and Manski (1995) (HM), which are the bounds provided in Lemma B.1(i) evaluated at the smallest feasible value of γ_k. This immediately implies the validity and sharpness of (ii). This also implies that the bounds (i) are valid, as follows: (I) $\mathbb{E}\left[Y_{1,d} - Y_{0,d} \,|\, D_0 = D_1 = d\right]$ is the difference in expectation of two such components, and (II) the lower (upper) bound is given by the HM lower (upper) bound of the first component minus the HM upper (lower) bound of the second component. For sharpness, first note that $\mathbb{E}\left[Y_{1,d} - Y_{0,d} \,|\, D_0 = D_1 = d\right]$ is the difference in expectation of two components, each of which belongs to a *different* mixture of the two defined by (4.2) and (4.3). Since these two mixtures do not share any components (only weights), the two HM bounds can be attained jointly whenever the weights $\underline{\gamma}_{d,d}^{1,r}$ and $\underline{\gamma}_{d,d}^{0,r}$ are jointly feasible. The result now follows by noting that $\underline{\gamma}_{d,d}^{1,r}$ and $\underline{\gamma}_{d,d}^{0,r}$ are jointly feasible if and only if $\mathbb{P}(D = d|Z = 1)\underline{\gamma}_{d,d}^{1,r} = \mathbb{P}(D = d|Z = 0)\underline{\gamma}_{d,d}^{0,r} = \underline{p}_{d,d}^{r}$ belongs to the identified set for $p_{d,d}$ which is true by definition of $\underline{p}_{d,d}^{r}$.

Proof of (iii): Finally, for (iii), note that

$$\sum_{d=l}^{l'} w_d(p_{1,1},\ldots,p_{K,K})\mathbb{E}[Y_{1,d} - Y_{0,d}|T = (d,d)]$$
$$= \sum_{d=l}^{l'} w_d(p_{1,1},\ldots,p_{K,K})\mathbb{E}[Y_{1,d}|T = (d,d)] - \sum_{d=l}^{l'} w_d(p_{1,1},\ldots,p_{K,K})\mathbb{E}[Y_{0,d}|T = (d,d)]\ .$$

First, fix a $p \in \Theta_I(\mathcal{R}_T)$. Now, the weights $w_d(p)$ are fixed. Part (i) gives, for each $d \in \{l,\ldots,l'\}$, the sharp upper (lower) bounds for $\mathbb{E}[Y_{1,d} - Y_{0,d} \mid T = (d,d)]$ as the treated lower-tail (upper-tail) component bound minus the control upper-tail (lower-tail) component bound, evaluated at $p_{d,d}$. Since, as in the case above, these components enter distinct observed mixture equations across (d,z), Lemma B.2 implies that the component-wise bounds are jointly attainable for fixed p. Therefore, for fixed p, the sharp lower and upper bounds for the weighted average are the corresponding weighted averages of these component-wise bounds. Optimizing these fixed-p bounds over $p \in \Theta_I(R_T)$ gives the bounds, and sharpness follows from Lemma B.2. □

A.6. **Proofs of Theorem 2 and Corollary 1.**

Proof of Theorem 2. First, by Assumption 3, we have that

$$U(1, D_0, \boldsymbol{\epsilon}) \leq U(1, D_1, \boldsymbol{\epsilon}) \quad \text{and} \quad U(0, D_1, \boldsymbol{\epsilon}) \leq U(0, D_0, \boldsymbol{\epsilon})\ ,$$

which implies that, almost surely,

$$U(1, D_0, \boldsymbol{\epsilon}) - U(0, D_0, \boldsymbol{\epsilon}) \leq W(1) - W(0) \leq U(1, D_1, \boldsymbol{\epsilon}) - U(0, D_1, \boldsymbol{\epsilon})\ .$$

Now, by Assumption 4, whenever $D_z = d \neq 0$,

$$U(1, D_z, \boldsymbol{\epsilon}) - U(0, D_z, \boldsymbol{\epsilon}) = \mathbb{E}[Y_{1,d} - Y_{0,d} \mid \boldsymbol{\epsilon}]\ ,$$

and when $d=0$, $U(1,D_z,\boldsymbol{\epsilon})-U(0,D_z,\boldsymbol{\epsilon})=0$. Therefore,

$$U(1,D_z,\boldsymbol{\epsilon})-U(0,D_z,\boldsymbol{\epsilon})=\sum_{d=1}^{K}\mathbf{1}\{D_z=d\}\,\mathbb{E}\left[Y_{1,d}-Y_{0,d}\,|\,\boldsymbol{\epsilon}\right]\ .$$

First, by taking expectations on all sides, this implies

$$\sum_{d=1}^{K}\mathbb{E}\left[\mathbf{1}\{D_0=d\}\,\mathbb{E}\left[Y_{1,d}-Y_{0,d}\,|\,\boldsymbol{\epsilon}\right]\right]\le\mathbb{E}\left[W(1)-W(0)\right]\le\sum_{d=1}^{K}\mathbb{E}\left[\mathbf{1}\{D_1=d\}\,\mathbb{E}\left[Y_{1,d}-Y_{0,d}\,|\,\boldsymbol{\epsilon}\right]\right]\ ,$$

and the validity of the bounds in (i) now follows immediately by noting that

$$\mathbb{E}\left[\mathbf{1}\{D_z=d\}\,\mathbb{E}\left[Y_{1,d}-Y_{0,d}\,|\,\boldsymbol{\epsilon}\right]\right]=\mathbb{P}(D_z=d)\,\mathbb{E}\left[\mathbb{E}\left[Y_{1,d}-Y_{0,d}\,|\,\boldsymbol{\epsilon}\right]|\,D_z=d\right]$$

and $\mathbb{E}\left[\mathbb{E}\left[Y_{1,d}-Y_{0,d}\,|\,\boldsymbol{\epsilon}\right]|\,D_z=d\right]=\mathbb{E}\left[Y_{1,d}-Y_{0,d}\,|\,D_z=d\right]$ since $\sigma(D_z)\subseteq\sigma(\boldsymbol{\epsilon})$.

Second, by multiplying by $\mathbf{1}\{D_0=D_1>0\}$ before taking expectations, we have

$$\begin{aligned}\sum_{d=1}^{K}&\mathbb{E}\left[\mathbf{1}\{D_0=D_1=d\}\,\mathbb{E}\left[Y_{1,d}-Y_{0,d}\,|\,\boldsymbol{\epsilon}\right]\right]\\&\le\mathbb{P}(D_1=D_0>0)\,\mathbb{E}\left[W(1)-W(0)\,|\,D_1=D_0>0\right]\\&\le\sum_{d=1}^{K}\mathbb{E}\left[\mathbf{1}\{D_0=D_1=d\}\,\mathbb{E}\left[Y_{1,d}-Y_{0,d}\,|\,\boldsymbol{\epsilon}\right]\right]\ ,\end{aligned}$$

which reduces to an equality since both bounds coincide. Whenever $\mathbb{P}(D_0=D_1>0)>0$ we can argue as above via $\sigma((D_0,D_1))\subseteq\sigma(\boldsymbol{\epsilon})$ to immediately get (ii).

It now suffices to show that, given any $((Y_{z,d}:d\in\{1,\ldots,K\},z\in\{0,1\}),D_0,D_1,Z)$ satisfying our assumptions, the bounds in (i) can be attained by h_ds that are consistent with this joint distribution and the assumptions above. To see this, first note that a given function h_d rationalizes observed choices iff, for each $z\in\{0,1\}$,

$$h_{D_z}(\boldsymbol{\epsilon})\ge\mathbb{E}\left[Y_{z,d}\,|\,\boldsymbol{\epsilon}\right]-\mathbb{E}\left[Y_{z,D_z}\,|\,\boldsymbol{\epsilon}\right]+h_d(\boldsymbol{\epsilon})\quad\text{almost surely,}$$

for all $d\in\{0,\ldots,K\}$, where (here and in the remainder of the proof) we WLOG take $\mathbb{E}\left[Y_{z,0}\,|\,\boldsymbol{\epsilon}\right]=0$ for all $z\in\{0,1\}$. Next, since D_z is $\sigma(\boldsymbol{\epsilon})$-measurable, there exists (by Doob-Dynkin lemma) a measurable function $g_z:\operatorname{supp}(\boldsymbol{\epsilon})\to\{0,\ldots,K\}$ such that $D_z=g_z(\boldsymbol{\epsilon})$ almost surely; for simplicity I will simply write $D_z=D_z(\epsilon)$ in the remainder of the proof. We will now construct a suitable h_d. Fix some $\boldsymbol{e}\in\operatorname{supp}(\boldsymbol{\epsilon})$. For all $d\notin\{D_0(\epsilon),D_1(\epsilon)\}$, let $h_d(\boldsymbol{e})\equiv-\max_{z\in\{0,1\},d\in\{0,\ldots,K\}}\{\mathbb{E}\left[Y_{z,d}\,|\,\boldsymbol{\epsilon}=\boldsymbol{e}\right]\}-$

$M(\boldsymbol{e})$, with $M(\boldsymbol{e})$ as specified below. First, note that fixing this value cannot affect the h_ds' ability to rationalize choices at values of $\boldsymbol{\epsilon} \neq \boldsymbol{e}$. Next, note that we can always choose $M(\boldsymbol{e})$ large enough such that choices are rationalized (given $\boldsymbol{\epsilon} = \boldsymbol{e}$) iff

$$\mathbb{E}\left[Y_{1,D_0(\boldsymbol{e})} \,\middle|\, \boldsymbol{\epsilon} = \boldsymbol{e}\right] - \mathbb{E}\left[Y_{1,D_1(\boldsymbol{e})} \,\middle|\, \boldsymbol{\epsilon} = \boldsymbol{e}\right] \\ \leq h_{D_1(\boldsymbol{e})}(\boldsymbol{e}) - h_{D_0(\boldsymbol{e})}(\boldsymbol{e}) \\ \leq \mathbb{E}\left[Y_{0,D_0(\boldsymbol{e})} \,\middle|\, \boldsymbol{\epsilon} = \boldsymbol{e}\right] - \mathbb{E}\left[Y_{0,D_1(\boldsymbol{e})} \,\middle|\, \boldsymbol{\epsilon} = \boldsymbol{e}\right] \; ,$$

where Assumption 4 ensures that bounds do not cross. Therefore, we are free to choose any $h_{D_1(\boldsymbol{e})}(\boldsymbol{e})$ and ${}_{D_0(\boldsymbol{e})}(\boldsymbol{e})$ such that $h_{D_1(\boldsymbol{e})}(\boldsymbol{e}) - h_{D_0(\boldsymbol{e})}(\boldsymbol{e})$ equals some value between these two intervals, and since $\boldsymbol{e}$ was arbitrary, we can do this for all $\boldsymbol{e} \in \operatorname{supp}(\boldsymbol{\epsilon})$. Finally, sharpness of the bounds in (i) follows immediately by noting that any choice of h_ds such that the lower (upper) bound above is attained for each $\boldsymbol{e} \in \operatorname{supp}(\boldsymbol{\epsilon})$ will result in $\mathbb{E}[W(1) - W(0)]$ attaining the lower (upper) bound in (i). □

A.6.1. *Proof of Corollary 1.*

Proof. Note that Theorem 2(ii) immediately implies validity of the bounds in Theorem 1(iii), with weights and layers as specified in this claim, for $\mathbb{E}[W(1)-W(0) \mid D_0 = D_1 > 0]$. Therefore, it suffices to verify that any value attainable under the baseline conditions of Theorem 1(iii) remains attainable after imposing Assumptions 3 and 4.

To this end, fix some point x in between the bounds in Theorem 1(iii), and let a joint distribution Q such that $\left((Y_{z,d} : d \in \{0, \dots, K\}, z \in \{0,1\}), T, Z\right) \sim Q$ and $\sum_{d \in \mathcal{D}} w_d(\boldsymbol{p}) \, \mathbb{E}_Q[Y_{1,d} - Y_{0,d} \mid T = (d,d)] = x$ be given. We will take $\epsilon \equiv T$, and construct a new law $\tilde{Q}$ by modifying Q as follows. For each stayer type $T = (d,d)$, set $h_d(T) = 0$ and set $h_{d'}(T)$ sufficiently low for all $d' \neq d$ (e.g., using the construction in the proof of Theorem 2 above), so that d maximizes utility under both $z = 0$ and $z = 1$. For each switcher type $T = (d_0, d_1)$, leave the observed potential outcomes Y_{0,d_0} and Y_{1,d_1} unchanged, set the unobserved potential outcomes Y_{1,d_0} and Y_{0,d_1} to be sufficiently low constants, and then choose $h_{d_1}(T) - h_{d_0}(T)$ in the nonempty interval

$$\left[\mathbb{E}_{\widetilde{Q}}(Y_{1,d_0} \mid T) - \mathbb{E}_{\widetilde{Q}}(Y_{1,d_1} \mid T), \, \mathbb{E}_{\widetilde{Q}}(Y_{0,d_0} \mid T) - \mathbb{E}_{\widetilde{Q}}(Y_{0,d_1} \mid T)\right],$$

with $h_d(T)$ again taken sufficiently low whenever $d \notin \{d_0, d_1\}$. This construction makes D_z $\sigma(\epsilon)$-measurable and utility-maximizing under each z, while preserving the observed distribution, Assumption 1, $\mathcal{R}_T$, and the value x of the aggregate stayer parameter. Therefore, every point in the Theorem 1(iii) identified interval remains attainable under Assumptions 3 and 4, as required. □

Appendix B. Auxiliaries

B.1. Sharp Bounds on Mixture Components in the Single Equation Case.

The following result follows from Horowitz and Manski (1995); Cross and Manski (2002) and Molinari and Peski (2006); we provide a self-contained argument here.

Lemma B.1. *Let $W, W_1, ..., W_K \in \mathcal{W} \subseteq \mathbb{R}$ be random variables such that*

$$\exists \{\gamma_k\}_{i=1}^K \subseteq \mathbb{R}_{\geq 0} \text{ such that } F_W(w) = \sum_{k=0}^{K} \gamma_k F_{W_k}(w) \ \forall w \in \mathcal{W}. \tag{B.1}$$

Then, the following statements hold, with $U \sim$ Uniform$[0,1]$.

(i) *For any $k \in \{1, ..., K\}$, the following bounds are sharp:*

$$\mathbb{E}[F_W^{-1}(U)|U \leq \gamma_k] \leq \mathbb{E}[W_k] \leq \mathbb{E}[F_W^{-1}(U)|U \geq 1-\gamma_k]. \tag{B.2}$$

(ii) *For any $1 \leq l \leq l' \leq K$, the following bounds are sharp:*

$$\mathbb{E}\left[F_W^{-1}(U)|U \leq \sum_{k=l}^{l'} \gamma_k\right] \leq \sum_{k=l}^{l'} \frac{\gamma_k}{\sum_{k=l}^{l'} \gamma_k} \mathbb{E}[W_k] \leq \mathbb{E}\left[F_W^{-1}(U)|U \geq 1 - \sum_{k=l}^{l'} \gamma_k\right]. \tag{B.3}$$

Proof. Since (i) is just a special case of (ii), it suffices to show (ii). We can WLOG take W and $\{W_k\}_{k=0}^K$ to be integrable and have densities with respect to a common dominating measure μ on $\mathcal{W}$. Denote the μ-density of W by f_W, and of W_k by f_{W_k}.

First, we show validity. Define $\bar{\gamma} \equiv \sum_{k=l}^{l'} \gamma_k$ and let $U \sim$ Uniform$[0,1]$. Now, suppose that $\mathbb{E}\left[F_W^{-1}(U)|U \leq \bar{\gamma}\right] > \sum_{k=l}^{l'} \frac{\gamma_k}{\bar{\gamma}} \mathbb{E}[W_k]$, then there must exist w such that

$$\mathbb{P}\left(F_W^{-1}(U) \leq w | U \leq \bar{\gamma}\right) < \sum_{k=l}^{l'} \frac{\gamma_k}{\bar{\gamma}} \int_{(-\infty,w]} f_{W_k}(x) \ d\mu(x) = \frac{1}{\bar{\gamma}} \int_{(-\infty,w]} \sum_{k=l}^{l'} \gamma_k f_{W_k}(x) \ d\mu(x) \ .$$

Now, note that, for such a w,

$$\mathbb{P}\left(F_W^{-1}(U) \leq w | U \leq \bar{\gamma}\right) = \mathbb{P}\left(F_W^{-1}(\bar{\gamma}U) \leq w\right) = \mathbb{P}\left(\bar{\gamma}U \leq F_W(w)\right) = \frac{\int_{(-\infty,w]} f_W(x) \ d\mu(x)}{\bar{\gamma}} \ ,$$

but since $\int_{(-\infty,w]} f_W(x)\ \ d\mu(x) = \int_{(-\infty,w]} \sum_{k=0}^{K} \gamma_k f_{W_k}(x)\ \ d\mu(x)$ this implies that $\int_{(-\infty,w]} f_W(x)\ \ d\mu(x) < \int_{(-\infty,w]} \sum_{k=l}^{l'} \gamma_k f_{W_k}(x)\ \ d\mu(x)$, which, in turn, implies that $\int_{(-\infty,w]} \sum_{k=0}^{l-1} \gamma_k f_{W_k}(x) + \sum_{k=l'+1}^{K} \gamma_k f_{W_k}(x)\ \ d\mu(x) < 0$, a contradiction. Therefore, the lower bound is valid. The validity of the upper bound follows analogously.

Define $\underline{\gamma} := \bar{\gamma} + \sum_{k'=0}^{l-1} \gamma_{k'}$. The lower bound is sharp since we can pick F_{W_k} so that

$$F_{W_k}(w) \equiv \begin{cases} \mathbb{P}\left(F_W^{-1}(U) \leq w | U \in \left(\sum_{k'=l}^{k-1} \gamma_{k'}, \sum_{k'=l}^{k} \gamma_{k'}\right)\right) & \text{if } k \in \{l, \ldots, l'\} \\ \mathbb{P}\left(F_W^{-1}(U) \leq w | U \in \left(\bar{\gamma} + \sum_{k'=0}^{k-1} \gamma_{k'}, \bar{\gamma} + \sum_{k=0}^{k} \gamma_{k'}\right)\right) & \text{if } k \in \{0, \ldots, l-1\} \\ \mathbb{P}\left(F_W^{-1}(U) \leq w | U \in \left(\underline{\gamma} + \sum_{k'=l'+1}^{k-1} \gamma_{k'}, \underline{\gamma} + \sum_{k'=l'+1}^{k} \gamma_{k'}\right)\right) & \text{all other } k \end{cases} .$$

Sharpness of the upper bound follows from the analogous construction. □

B.2. **Identified Set of Type-Weighted Potential Outcome Densities.** In the remainder, let $\mathcal{D} \equiv \{0, \ldots, K\}, \mathcal{D}_+ \equiv \mathcal{D} \setminus \{0\}$ and let $\mathcal{Y}$ be the support of Y, with $y_L \equiv \inf \mathcal{Y}$ and $y_U \equiv \sup \mathcal{Y}$.

Define $\mathbf{f}_{(z,d)|d,d'}(y) \equiv \mathbb{P}[T = (d', d)] f_{Y_{z,d}|T=(d',d)}(y|d', d)$, and also define the following shorthand notation $\mathbf{f}_{z|d,d'}(y) = \mathbf{f}_{z|t}(y)$, $\mathbf{f}_{z|d,d'}(y) \equiv \mathbb{P}[T = (d', d)] f_{Y_{z,T(z)}|T=(d',d)}(y|d', d) = \mathbb{P}[T = t] f_{Y_{z,t(z)}|T=t}(y|t)$. Define $\mathbf{f}(y) \equiv \left(\mathbf{f}_{(z,d)|t}(y) : d \in \mathcal{D}_+, t \in \mathcal{D}^2, z \in \{0, 1\}\right)$

$$(\mathbf{p}, \mathbf{f}) \equiv ((p_{d,d'} : d, d' \in \{0, \ldots, K\}), \mathbf{f}(\cdot)) = \left(\left(p_t : t \in \{0, \ldots, K\}^2\right), \mathbf{f}(\cdot)\right) .$$

Let $\mathfrak{D}$ denote $(\mathbf{p}, \mathbf{f})$ space, i.e. the space of tuples $\left(\tilde{\mathbf{p}}, \tilde{\mathbf{f}}\right)$ such that $\tilde{\mathbf{p}}$ belongs to the $\mathcal{D}^2$ probability simplex and $\tilde{\mathbf{f}}$ is a stacked vector function of the same dimension as $\mathbf{f}$, with each component being a μ-integrable real-valued function $\mathcal{Y} \to \mathbb{R}$.

Lemma B.2. *Under model* (2.3)-(2.4)*, Assumption 1 and (arbitrary) response-type restriction* $\mathcal{R}_T$*, the identified set for* $(\mathbf{p}, \mathbf{f})$ *is given by*

$$\left\{ \left(\tilde{\mathbf{p}}, \tilde{\mathbf{f}}\right) \in \mathfrak{D} \;\middle|\; \begin{array}{l} \tilde{\mathbf{p}} \in \Theta_I(\mathcal{R}_T) \ , \int_{y_L}^{y_U} \tilde{\mathbf{f}}_{z|t}(y)\, d\mu(y) = \tilde{\mathbf{p}}_t \ \forall z \in \{0, 1\}, t \in \mathcal{D}^2 \ , \\ \sum_{t \in \mathcal{D}^2 : t(z) = d} \tilde{\mathbf{f}}_{z|t}(y) = f_{Y, D=d|Z=z}(y) \ \forall (y, d, z) \in \mathcal{Y} \times \mathcal{D} \times \{0, 1\} \\ \tilde{\mathbf{f}}_{z|t}(y) \geq 0 \ \forall (y, t, z) \in \mathcal{Y} \times \mathcal{D}^2 \times \{0, 1\} \ . \end{array} \right\}$$

Proof. Note that for any type t and $z \in \{0, 1\}$, $f_{Y_{z,d''}|T}(y|t)$ is independent of the data whenever $d'' \neq t(z)$, and so, is only constrained to be a density that has support in $\mathcal{Y}$; this immediately implies that the sharp identification region for the expectation of any such component is $\mathcal{Y}$. Given Lemma 3 above, sharpness now follows from Theorem 3.2 in Vayalinkal (2024). We summarize the argument here, as follows.

First, note that the observed data depends only on the (i) the distribution of Z (F_Z), (ii) the marginal distribution of D_z for each $z \in \{0,1\}$, and (iii) the conditional marginal distribution of $Y_{z,d}$ given $D_z = d, Z = z$ for all $d \in \{1,\dots,K\}$ and $z \in \{0,1\}$. For any joint distribution $((Y_{z,d} : d \in \{0,\dots,K\}, z \in \{0,1\}), T, Z) \sim Q$, let $\mathbf{f}_Q$ be the vector of weighted response-type conditional densities implied by Q. Given $\mathbf{f}$ satisfying the conditions above, we construct a Q with $\mathbf{f}_Q = \mathbf{f}$ as follows: define $Q_Z = F_Z$, $Q(T = t) = \int_{\mathcal{Y}} \mathbf{f}_{1|t}(y)d\mu(y)$, and define

$$Q(Y_{0,0} \le y_{0,0}, \dots, Y_{0,K} \le y_{0,K}, Y_{1,0} \le y_{1,0}, \dots, Y_{1,K} \le y_{1,K}, T = t, Z = z)$$
$$= \left(\prod_{k=0}^{K} \int_{y_L}^{y_{0,k}} f_{(0,k)|t}(y)d\mu(y) \right) \left(\prod_{k=0}^{K} \int_{y_L}^{y_{1,k}} f_{(1,k)|t}(y)d\mu(y) \right) Q(T = t) Q(Z = z) \ ,$$

where $f_{(z,k)|t}(y) \equiv \frac{1}{Q(T=t)} \mathbf{f}_{(z,k)|t}$ if $\int_{\mathcal{Y}} \mathbf{f}_{(z,k)|t}(y)d\mu(y) \neq 0$, else $f_{(z,k)|t} \equiv 0$.

The above construction assumes that the potential outcome distributions are independent given T, but any dependence structure (copula) can be used, after conditioning on a value of T. Suppose we are given a Q such that $\mathbf{f}_Q$ satisfies the conditions above. By construction, $Q_Z = F_Z$, $Q(D_z = d) = \sum_{t:t(z)=d} Q(T = t) = \sum_{t:t(z)=d} \int_{y_L}^{y_U} \mathbf{f}_{z|t}(y)d\mu(y) = P(D = d|Z = z)$ for all $d \in \{1,\dots,K\}$ and $z \in \{0,1\}$. This also implies $Q(D_z = 0) = P(D = 0|Z = z)$ by the definition of Θ_I and, finally,

$$Q(Y_{z,d} \le y|D_z = d, Z = z) = \sum_{t:t(z)=d} \int_{y_L}^{y} \mathbf{f}_{z|t}(y)d\mu(y) = \int_{y_L}^{y} \sum_{t:t(z)=d} \mathbf{f}_{z|t}(y)d\mu(y)$$
$$= \int_{y_L}^{y} f_{Y,D=d|Z=z}(y)d\mu(y) = P(Y \le y|D = d, Z = z) \ ,$$

for any $z \in \{0,1\}$ and $d \in \{1,\dots,K\}$, as required. □

References

Abowd, John M., Francis Kramarz, and David N. Margolis (1999): "High Wage Workers and High Wage Firms," *Econometrica : journal of the Econometric Society*, 67, 251–333.

Andersson, Fredrik, Harry J. Holzer, Julia I. Lane, David Rosenblum, and Jeffrey Smith (2022): "Does Federally-Funded Job Training Work? Nonexperimental Estimates of WIA Training Impacts Using Longitudinal Data on Workers and Firms," *Journal of Human Resources*, 0816–8185R1.

Andrews, Donald W. K. and Gustavo Soares (2010): "Inference for Parameters Defined by Moment Inequalities Using Generalized Moment Selection," *Econometrica*, 78, 119–157.

Azar, José A., Steven T. Berry, and Ioana Marinescu (2022): "Estimating Labor Market Power," NBER Working Paper 30365, National Bureau of Economic Research.

Blanco, German, Carlos A Flores, and Alfonso Flores-Lagunes (2013): "The Effects of Job Corps Training on Wages of Adolescents and Young Adults," *American Economic Review*, 103, 418–422.

Canay, Ivan, Andres Santos, and Azeem M Shaikh (2013): "On the Testability of Identification in Some Nonparametric Models With Endogeneity," *Econometrica : journal of the Econometric Society*, 81, 2535–2559.

Card, David, Ana Rute Cardoso, Joerg Heining, and Patrick Kline (2018a): "Firms and Labor Market Inequality: Evidence and Some Theory," *Journal of Labor Economics*, 36, S13–S70.

Card, David, Ana Rute Cardoso, and Patrick Kline (2016): "Bargaining, Sorting, and the Gender Wage Gap: Quantifying the Impact of Firms on the Relative Pay of Women *," *The Quarterly Journal of Economics*, 131, 633–686.

Card, David, Jörg Heining, and Patrick Kline (2013): "Workplace Heterogeneity and the Rise of West German Wage Inequality*," *The Quarterly Journal of Economics*, 128, 967–1015.

Card, David, Jochen Kluve, and Andrea Weber (2010): "Active Labour Market Policy Evaluations: A Meta-Analysis," *The Economic Journal*, 120, F452–F477.

——— (2018b): "What Works? A Meta Analysis of Recent Active Labor Market Program Evaluations," *Journal of the European Economic Association*, 16, 894–931.

Chan, Mons, Kory Kroft, Elena Mattana, and Ismael Mourifié (2024): "An Empirical Framework For Matching With Imperfect Competition," Tech. Rep. w32493, National Bureau of Economic Research, Cambridge, MA.

Chetty, Raj (2009): "Sufficient Statistics for Welfare Analysis: A Bridge Between Structural and Reduced-Form Methods," *Annual Reviews*, 1, 451–488.

Cross, Philip J. and Charles F. Manski (2002): "Regressions, Short and Long," *Econometrica : journal of the Econometric Society*, 70, 357–368.

Dey, Matthew S. and Christopher J. Flinn (2005): "An Equilibrium Model of Health Insurance Provision and Wage Determination," *Econometrica : journal of the Econometric Society*, 73, 571–627.

D'Haultfoeuille, Xavier (2011): "On the Completeness Condition in Nonparametric Instrumental Problems," *Econometric Theory*, 27, 460–471.

D'Haultfoeuille, Xavier and Arnaud Maurel (2013): "Inference on an Extended Roy Model, with an Application to Schooling Decisions in France," *Journal of Econometrics*, 174, 95–106.

Dube, Arindrajit, Jeff Jacobs, Suresh Naidu, and Siddharth Suri (2020): "Monopsony in Online Labor Markets," *American Economic Review: Insights*, 2, 33–46.

Eisenhauer, Philipp, James J. Heckman, and Edward J. Vytlacil (2015): "The Generalized Roy Model and the Cost-Benefit Analysis of Social Programs," *Journal of Political Economy*, 123, 413–443.

Fang, Zheng and Andres Santos (2019): "Inference on Directionally Differentiable Functions," *The Review of Economic Studies*, 86, 377–412.

Fang, Zheng, Andres Santos, Azeem M. Shaikh, and Alexander Torgovitsky (2023): "Inference for Large-Scale Linear Systems With Known Coefficients," *Econometrica : journal of the Econometric Society*, 91, 299–327.

Gerard, François, Lorenzo Lagos, Edson Severnini, and David Card (2021): "Assortative Matching or Exclusionary Hiring? The Impact of Employment and Pay Policies on Racial Wage Differences in Brazil," *American Economic Review*, 111, 3418–3457.

Halmos, Paul R. (1976): *Measure Theory*, Springer New York.

Heckman, James J. (1979): "Sample Selection Bias as a Specification Error," *Econometrica : journal of the Econometric Society*, 47, 153.

Heckman, James J., Robert J. Lalonde, and Jeffrey A. Smith (1999): "The Economics and Econometrics of Active Labor Market Programs," in *Handbook of Labor Economics*, Elsevier, vol. 3, 1865–2097.

Heckman, James J. and Rodrigo Pinto (2018): "Unordered Monotonicity," *Econometrica : journal of the Econometric Society*, 86, 1–35.

Heiler, Philipp, Katrin Kaufmann, and Elvin Veliyev (2024): "Treatment Evaluation at the Intensive and Extensive Margins," *arXiv preprint*.

Heiler, Phillip and Michael C. Knaus (2023): "Effect or Treatment Heterogeneity? Policy Evaluation with Aggregated and Disaggregated Treatments," .

Hendren, Nathaniel and Ben Sprung-Keyser (2020): "A Unified Welfare Analysis of Government Policies*," *The Quarterly Journal of Economics*, 135, 1209–1318.

Honoré, Bo E. and Luojia Hu (2020): "Selection Without Exclusion," *Econometrica : journal of the Econometric Society*, 88, 1007–1029.

Horowitz, Joel L. and Charles F. Manski (1995): "Identification and Robustness with Contaminated and Corrupted Data," *Econometrica : journal of the Econometric Society*, 63, 281.

Huber, Martin, Lukas Laffers, and Giovanni Mellace (2017): "Sharp IV Bounds on Average Treatment Effects on the Treated and Other Populations Under Endogeneity and Noncompliance," *Journal of Applied Econometrics*, 32, 56–79.

Katz, Lawrence F., Jonathan Roth, Richard Hendra, and Kelsey Schaberg (2022): "Why Do Sectoral Employment Programs Work? Lessons from WorkAdvance," *Journal of Labor Economics*, 40, S249–S291.

Kitagawa, Toru (2015): "A Test for Instrument Validity," *Econometrica : journal of the Econometric Society*, 83, 2043–2063.

——— (2021): "The Identification Region of the Potential Outcome Distributions under Instrument Independence," *Journal of Econometrics*, 225, 231–253.

Kleven, Henrik (2021): "Sufficient Statistics Revisited," *Annual Reviews*, 13, 515–538.

Kline, Patrick and Christopher Walters (2016): "Evaluating Public Programs with Close Substitutes: The Case of Head Start," *The Quarterly Journal of Economics*, 131, 1795–1848.

Kroft, Kory, Yao Luo, Magne Mogstad, and Bradley Setzler (2025): "Imperfect Competition and Rents in Labor and Product Markets: The Case of the Construction Industry," *American Economic Review*, 115, 2926–2969.

Kwon, Soonwoo and Jonathan Roth (2024): "Testing Mechanisms," *arXiv preprint arXiv:2404.11739*.

Lachowska, Marta, Alexandre Mas, and Stephen A. Woodbury (2020): "Sources of Displaced Workers' Long-Term Earnings Losses," *American Economic Review*, 110, 3231–3266.

Lamadon, Thibaut, Magne Mogstad, and Bradley Setzler (2022): "Imperfect Competition, Compensating Differentials, and Rent Sharing in the US Labor Market," *American Economic Review*, 112, 169–212.

Lee, David S. (2009): "Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects," *Review of Economic Studies*, 76, 1071–1102.

Maestas, Nicole, Kathleen J. Mullen, David Powell, Till Von Wachter, and Jeffrey B. Wenger (2023): "The Value of Working Conditions in the United States and Implications for the Structure of Wages," *American Economic Review*, 113, 2007–2047.

Molinari, Francesca and Marcin Peski (2006): "GENERALIZATION OF A RESULT ON "REGRESSIONS, SHORT AND LONG"," *Econometric Theory*, 22.

Mourifié, Ismael and Yuanyuan Wan (2017): "Testing Local Average Treatment Effect Assumptions," *The Review of Economics and Statistics*, 99, 305–313.

Olma, Tomasz (2021): "Nonparametric Estimation of Truncated Conditional Expectation Functions," *arXiv preprint*.

Pearl, Judea (2001): "Direct and Indirect Effects," Tech. rep.

Pierce, B. (2001): "Compensation Inequality," *The Quarterly Journal of Economics*, 116, 1493–1525.

Robins, James M. and Sander Greenland (1992): "Identifiability and Exchangeability for Direct and Indirect Effects:," *Epidemiology (Cambridge, Mass.)*, 3, 143–155.

Schmieder, Johannes F., Till Von Wachter, and Jörg Heining (2023): "The Costs of Job Displacement over the Business Cycle and Its Sources: Evidence from Germany," *American Economic Review*, 113, 1208–1254.

Schochet, Peter, Jeanne Bellotti, Ruo-Jiao Cao, Steven Glazerman, April Grady, Mark Gritz, Sheena McConnell, Terry Johnson, and John Burghardt (2003): "National Job Corps Study: Data Documentation and Public Use Files: Volume I," .

Schochet, Peter, John Burghardt, and Steven Glazerman (2001): "National Job Corps Study: The Impacts of Job Corps on Participants' Employment and Related Outcomes," Tech. rep., Mathematica Policy Research, Princeton, NJ.

Schochet, Peter Z, John Burghardt, and Sheena McConnell (2008): "Does Job Corps Work? Impact Findings from the National Job Corps Study," *American Economic Review*, 98, 1864–1886.

Semenova, Vira (2020): "Generalized Lee Bounds," *arXiv preprint*.

Słoczyński, Tymon (2020): "When Should We (Not) Interpret Linear IV Estimands as LATE?" *arXiv preprint*.

Vayalinkal, Atom (2024): "Sharp Identification Regions in General Selection Models with (Un)Ordered Treatments and Discrete Instruments," *Unpublished Manuscript, University of Toronto*.

Yoshikawa, Kohei and Shuichi Kawano (2026): "Identification and Estimation under Multiple Versions of Treatment: Mixture-of-Experts Approach," .

Zuo, Shuozhi, Debashis Ghosh, Peng Ding, and Fan Yang (2022): "Mediation Analysis with the Mediator and Outcome Missing Not at Random," *arXiv preprint*.

SUPPLEMENT TO "HOROWITZ-MANSKI-LEE BOUNDS WITH MULTILAYERED SAMPLE SELECTION"

KORY KROFT*, ISMAEL MOURIFIÉ †, AND ATOM VAYALINKAL‡

Appendix A. Additional Results

A.1. **Bounds on Aggregate Effect for "Stayers".** To derive bounds on the aggregate LCDE, defined as $\sum_{d=l}^{l'} \frac{p_{d,d}}{\sum_{d'=l}^{l'} p_{d',d'}} \mathbb{E}[Y_{1,d} - Y_{0,d} \mid T = (d,d)]$, one might be tempted to adopt a naive approach by taking a weighted average of the pointwise sharp bounds derived in Theorem 1(i). However, this approach not only fails to provide sharp bounds on the aggregate quantity but may also yield invalid bounds. The primary issue lies in the fact that both the weights and the LCDE quantities (i.e. the $\mathbb{E}[Y_{1,d} - Y_{0,d} \mid T = (d,d)]$ terms) are only set-identified. Consequently, the naive aggregation does not preserve the sharpness or validity of the bounds. To better illustrate this problem, consider the following.

The sharp bounds on the aggregate LCDE can be tighter than the weighted average of the marginal bounds on $\mathbb{E}[Y_{1,d} - Y_{0,d} \mid T = (d,d)]$ for $d = l, \ldots, l'$ when these marginal bounds are achieved at values of $p_{d,d}$ for $d = l, \ldots, l'$ that are not *jointly* attainable. Specifically, this occurs whenever $\left(\underline{p}_{d,d} : d = l, \ldots, l'\right)$ does not lie within the identified set for $(p_{d,d} : d = l, \ldots, l')$. As a result, the sharp bounds, which are obtained by optimizing over the identified set for the *vector* $(p_{d,d} : d = l, \ldots, l')$, may be tighter.

The sharp bounds on the LCDE can be wider than the weighted average of the marginal LCDE bounds when there exist $d, d' \in \{l, \ldots, l'\}$ such that the lower bound for $\mathbb{E}[Y_{1,d} - Y_{0,d} \mid T = (d,d)]$ is significantly smaller than the lower bound for $\mathbb{E}[Y_{1,d'} - Y_{0,d'} \mid T = (d',d')]$, and $\underline{p}_{d,d}$ is also much smaller than $\underline{p}_{d',d'}$.

In such cases, the choice of $p_{d,d}$ directly affects the relative weight of $\mathbb{E}[Y_{1,d} - Y_{0,d} \mid T = (d,d)]$ in the aggregate LCDE. Consequently, the value of the objective function in the optimization problem defining the lower bound in Theorem 1(iii) may actually decrease as $p_{d,d}$ increases.

To illustrate this more clearly, consider a simple example with only two groups, a and b. For each $d \in \{a, b\}$, let $L_d(p)$ denote the sharp lower bound for $\mathbb{E}[Y_{1,d} - Y_{0,d} \mid T = (d,d)]$ under the assumption that $p_{d,d} = p$. Suppose $\underline{p}_{b,b} > \underline{p}_{a,a} = 0$, so that the lower bound for $\mathbb{E}[Y_{1,a} - Y_{0,a} \mid T = (a,a)]$, given by $L_a(\underline{p}_{a,a}) = L_a(0) = y_L$, is the trivial bound. Additionally, assume the lower bound for $\mathbb{E}[Y_{1,b} - Y_{0,b} \mid T = (b,b)]$ is non-trivial, i.e., $L_b(\underline{p}_{b,b}) > y_L$.

In this scenario, the weighted average of these lower bounds simplifies to:

$$\frac{\underline{p}_{a,a} L_a(\underline{p}_{a,a}) + \underline{p}_{b,b} L_b(\underline{p}_{b,b})}{\underline{p}_{a,a} + \underline{p}_{b,b}} = L_b(\underline{p}_{b,b}),$$

since $\underline{p}_{a,a} = 0$.

Now, consider the case where there exists a point $p > 0$ such that $(p, \underline{p}_{b,b})$ lies within the identified set for $(p_{a,a}, p_{b,b})$, and $L_a(p) < L_b(\underline{p}_{b,b})$. In this case, we have:

$$\frac{p L_a(p) + \underline{p}_{b,b} L_b(\underline{p}_{b,b})}{p + \underline{p}_{b,b}} < L_b(\underline{p}_{b,b}).$$

By definition, the sharp lower bound given in Theorem 1(iii) will be at least as small as the left-hand side above, resulting in a smaller lower bound for the aggregate LCDE. Similarly, it is also possible for the sharp upper bound to be larger than the weighted average of the marginal upper bounds.

Appendix B. Relation to the Mediation Analysis Literature

B.1. **Direct and Indirect Effects in Presence of Sample Selection.** This section establishes a connection between our model and the literature on mediation analysis (Pearl, 2001). In mediation analysis, where Z represents the randomized treatment, D is the "mediator", and Y is the outcome, the treatment (job training) can influence the outcome through two distinct channels: a direct channel and an indirect channel that passes through the mediator. The vector W of latent unobserved variables (often called "confounding variables") simultaneously affect D and Y, making D an endogenous variable. In our setting, the mediator corresponds to the firm where the individual would be employed if they were externally assigned to job training. The graphical representation of the outcome equation in our model takes the form:[1]

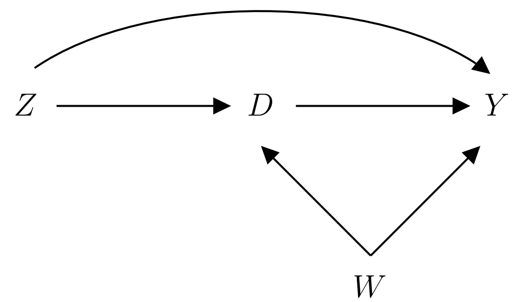


FIGURE 1. DAG of causal relationships between variables in our model.

The mediation literature distinguishes two causal estimands: direct and indirect effects. The *control direct effect* (CDE) is defined as:

$$\text{CDE}(d) \equiv \mathbb{E}[Y_{1,d} - Y_{0,d}]. \tag{B.1}$$

The CDE captures the causal effect of job training on earnings when firm type d is held fixed —equivalently, the within-firm wage effect for layer d. In Lee's (2009) terminology, the CDE corresponds to the causal impact of job training on the wage rate illustrated by the curved arrow in Figure 1. This parameter is the primary focus of Lee (2009).

CDEs are useful when the policymaker is primarily interested in the impact of job training on wages at a specific firm. Policymakers may also be interested in understanding the overall impact of training on wages at firms that workers naturally choose when they receive training. The second type of direct effect – the "natural direct effect" (NDE) – captures this notion:

$$\text{NDE} \equiv \mathbb{E}[Y_{1,D_1} - Y_{0,D_1}] = \sum_{d=0}^{K} \mathbb{E}[Y_{1,d} - Y_{0,d}|D_1 = d] \times \mathbb{P}[D_1 = d] \tag{B.2}$$

[1]For the sake of clarity, this graph simplifies the discussion by omitting sample selection.

where $Y_{z,D_{z'}} \equiv \sum_{d=0}^{K} Y_{z,d} 1\{D_{z'} = d\}$ for $z, z' \in \{0,1\}$, and $d \in \{0, ..., K\}$. The expression $Y_{1,D_1} - Y_{0,D_1}$ is the wage impact of job training at the firm a worker would select under treatment assignment. The NDE averages these individual effects.

Turning to indirect effects, the "natural indirect effect" (NIE) is defined as:

$$\begin{aligned} \text{NIE} &\equiv \mathbb{E}[Y_{0,D_1} - Y_{0,D_0}] \\ &= \sum_{d=0}^{K} \sum_{d'=0:d' \neq d}^{K} \mathbb{E}[Y_{0,d} - Y_{0,d'} | D_0 = d', D_1 = d] \times \mathbb{P}[D_0 = d', D_1 = d]. \end{aligned} \quad \text{(B.3)}$$

The term $Y_{0,d} - Y_{0,d'}$ represents the wage gap between firms d and d' absent job training. Evaluated at the natural representative firms D_1 and D_0, this becomes $Y_{0,D_1} - Y_{0,D_0}$. The indirect effect isolates the impact of job training operating purely through firm change. As highlighted by Pearl (2009), this estimand is empirically controversial since suppressing the direct effect of Z on Y while preserving the indirect channel is not realistic, though it remains standard in the mediation literature.

We now introduce two key parameters, the Local Controlled Direct Effect (LCDE) and the Local Controlled Indirect Effect (LCIE):

$$\text{LCDE}(d|t) = \mathbb{E}[Y_{1,d} - Y_{0,d} | T = t], d \in \{1, ..., K\}, \text{ and } t \in \mathcal{T} \quad \text{(B.4)}$$

$$\text{LCIE}(z, d, d'|t) = \mathbb{E}[Y_{z,d} - Y_{z,d'} | T = t], d \in \{1, ..., K\}, \text{ and } t \in \mathcal{T} \quad \text{(B.5)}$$

The LCDE allows the CDE to vary across response types t, accommodating individual-level heterogeneity in firm-specific treatment effects. The LCDE can be more policy-relevant than the CDE in specific contexts, mirroring the ATE vs. LATE debate in the IV literature. The same applies to the LCIE.

The "sample selection" analogues of the CDE, NDE, and NIE defined conditionally on always-employed workers $\{D_0 > 0, D_1 > 0\}$ are weighted average of LCDE$(d|t)$ or

LCIE$(z, d, d'|t)$.

$$\mathbb{E}[Y_{1,d} - Y_{0,d}|D_0 > 0, D_1 > 0] = \sum_{l=1}^{K}\sum_{l'=1}^{K} \text{LCDE}(d|l, l') \times \mathbb{P}[T = (l, l')|D_0 > 0, D_1 > 0],$$

$$\mathbb{E}[Y_{1,D_1} - Y_{0,D_1}|D_0 > 0, D_1 > 0] = \sum_{d=1}^{K}\sum_{d'=1}^{K} \text{LCDE}(d|d', d) \times \mathbb{P}[T = (d', d)|D_0 > 0, D_1 > 0],$$

$$\mathbb{E}[Y_{0,D_1} - Y_{0,D_0}|D_0 > 0, D_1 > 0] = \sum_{d=1}^{K}\sum_{d'=1:d\neq d'}^{K} \text{LCIE}(0, d, d'|d', d) \times \mathbb{P}[T = (d', d)|D_0 > 0, D_1 > 0].$$

Appendix C. 2 Firm Types Case: A numerical illustration

This section considers our model two firm types: high type (H) and low type (L). Under Assumption 2, the always-employed (AE) consist of four response types:

$$\{D_0 > 0, D_1 > 0\} = \{(L, L), (H, H), (L, H), (H, L)\} \equiv AE.$$

To establish bounds on our causal effects of interest, we first characterize identification of the response-type probabilities: $\{p_t : t \in \mathcal{T}\}$. Using information on (D, Z) only, $\{p_t : t \in \mathcal{T}\}$ has to satisfy equations (4.4, 4.5) from step 1 above. In addition, if we impose Assumption 2, i.e., $\mathcal{R}_T = \{\text{Assumption 2}\}$, the identified set for the response types in this simple case is characterized by non-negative solutions to the following set of (in)equalities:

$$p_{0,0} = 1 - \mathbb{P}(D = H \mid Z = 1) - \mathbb{P}(D = L \mid Z = 1), \tag{C.1}$$

$$p_{0,L} = \mathbb{P}(D = L \mid Z = 1) - \mathbb{P}(D = H \mid Z = 0) + p_{H,H} - p_{L,L}, \tag{C.2}$$

$$p_{0,H} = \mathbb{P}(D = H \mid Z = 1) - \mathbb{P}(D = L \mid Z = 0) + p_{L,L} - p_{H,H}, \tag{C.3}$$

$$p_{L,H} = \mathbb{P}(D = L \mid Z = 0) - p_{L,L}, \tag{C.4}$$

$$p_{H,L} = P(D = H \mid Z = 0) - p_{H,H}, \tag{C.5}$$

$$\max\{0, \mathbb{P}(D = H|Z = 0) - \mathbb{P}(D = L|Z = 1)\} \leq p_{H,H} \leq \min\{\mathbb{P}(D = H|Z = 0), \mathbb{P}(D = H|Z = 1)\} \tag{C.6}$$

$$\max\{0, \mathbb{P}(D = L|Z = 0) - \mathbb{P}(D = H|Z = 1)\} \leq p_{L,L} \leq \min\{\mathbb{P}(D = L|Z = 0), \mathbb{P}(D = L|Z = 1)\}. \tag{C.7}$$

More precisely, we can show that

$$\Theta_I(\mathcal{R}_T) = \{(p_t : t \in \mathcal{T}) \geq 0 : \text{ such that eqs } (C.1) \text{ to } (C.7) \text{ are satisfied}\}.$$

In this case, it is also straightforward to show that $\Theta_I(\mathcal{R}_T) \neq \emptyset$ if and only if the propensity scores satisfy

$$\mathbb{P}(D = 0|Z = 1) \leq \mathbb{P}(D = 0|Z = 0). \quad \text{(C.8)}$$

In this particular case, we have:

$$\underline{\gamma}^z_{H,H} = \frac{\max\{0, \mathbb{P}(D = H|Z = 0) - \mathbb{P}(D = L|Z = 1)\}}{\mathbb{P}(D = H|Z = z)}, \text{ for } z \in \{0, 1\},$$
$$\underline{\gamma}^z_{L,L} = \frac{\max\{0, \mathbb{P}(D = L|Z = 0) - \mathbb{P}(D = H|Z = 1)\}}{\mathbb{P}(D = L|Z = z)}, \text{ for } z \in \{0, 1\},$$

In the numerical illustration below, we describe $\Theta_I(\mathcal{R}_T)$ and demonstrate how imposing further assumptions on response types can significantly refine $\Theta_I(\mathcal{R}_T)$. In this case, as shown by (C.1)-(C.5), the response-type probability p_t, for each $t \in \{(0, 0), (0, L), (0, H), (L, H), (H, L)\}$, can be represented as a linear function of $p_{H,H}$ and $p_{L,L}$, which are themselves only set identified; in other words, $\Theta_I(\mathcal{R}_T)$ is parameterized by $(p_{H,H}, p_{L,L})$. Hence, our forthcoming discussion will focus primarily on illustrating the projection of $\Theta_I(\mathcal{R}_T)$ with respect to the coordinates of $(p_{H,H}, p_{L,L})$.

C.0.1. *Data Generating Process.* We first generate propensity scores to be consistent with the data in Lee (2009). This is reported in Table 1.

Table 1. Propensity scores for numerical illustrations

Job Training	$P(D = L\|Z = z)$	$P(D = H\|Z = z)$
$Z = 1$	0.302886	0.408114
$Z = 0$	0.313959	0.373041

The data are generated such that the true values for $p_{H,H}$, and $p_{L,L}$ are:

$$p_{H,H} = \mathbb{P}(D = H \mid Z = 0) = 0.373041,$$
$$p_{L,L} = 0.278886.$$

Next, we randomly generate the outcomes.[2] For each type $t \in \mathcal{T}$, let D_z denote employment status when externally assigned $Z = z$. The conditional distributions of $\exp(Y_{z,D_z}) \mid T = t$ for each $t \in \mathcal{T}$ are assumed to follow Lognormal$(\mu_{z|t}, \sigma_{z|t})$ distributions. Here, $\sigma_{z|t} = 1$ for all combinations of z and t, indicating that variability only arises through $\mu_{z|t}$ between types. We present two distinct potential earnings distribution models as described in Table 2.

Table 2. Potential Outcome distributions for the simulated DGPs

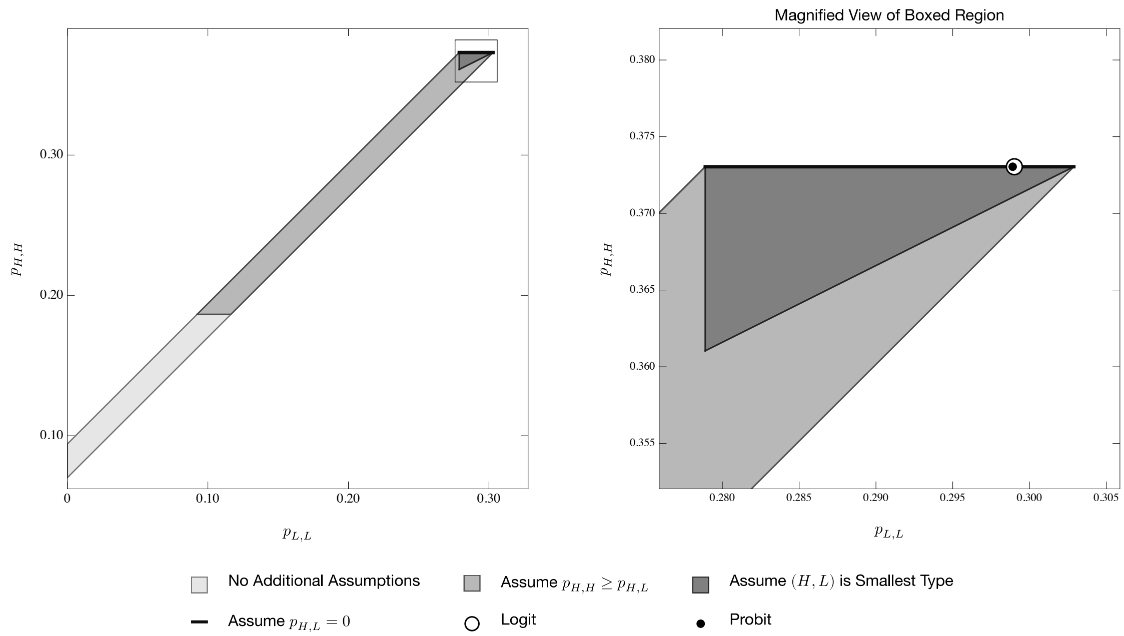
	Design 1		Design 2	
t	$\exp(Y_{1,D_1}) \mid T = t$	$\exp(Y_{0,D_0}) \mid T = t$	$\exp(Y_{1,D_1}) \mid T = t$	$\exp(Y_{0,D_0}) \mid T = t$
$(0, L)$	Lognormal$(9.5, 1)$	Lognormal$(9.5, 1)$	Lognormal$(10.5, 1)$	Lognormal$(9.5, 1)$
$(0, H)$	Lognormal$(11.5, 1)$	Lognormal$(9.5, 1)$	Lognormal$(12.5, 1)$	Lognormal$(9.5, 1)$
(L, H)	Lognormal$(16.5, 1)$	Lognormal$(9.5, 1)$	Lognormal$(14.5, 1)$	Lognormal$(9.5, 1)$
(H, L)	Lognormal$(9.75, 1)$	Lognormal$(9.6, 1)$	Lognormal$(10.5, 1)$	Lognormal$(10.5, 1)$
(L, L)	Lognormal$(9.5, 1)$	Lognormal$(9.5, 1)$	Lognormal$(10.5, 1)$	Lognormal$(9.5, 1)$
(H, H)	Lognormal$(14.5, 1)$	Lognormal$(14.5, 1)$	Lognormal$(14, 1)$	Lognormal$(12, 1)$

C.0.2. *Simulations Results.* We begin by exploring the geometry of $\Theta_I(\mathcal{R}_T)$, and demonstrate how incorporating further assumptions regarding response types can significantly shrink its shape.

Figure 3, along with Tables 3 and 4, present the results for both designs. Initially, we compute Lee bounds, which set identifies the total effect $\mathbb{E}(Y_{1,D_1} - Y_{0,D_0} | D_0 > 0, D_1 > 0)$, as in Lemma 2. Within the framework of design 1, these bounds lie entirely within the positive quadrant, excluding 0, suggesting that job training increases wage rates for the always-employed AE. However, for this DGP, the within-firm effect is 0. As illustrated below, whenever $\mathbb{E}(Y_{1,H} - Y_{0,H} | T = (H, H)) = \mathbb{E}(Y_{1,L} - Y_{0,L} | T = (L, L)) = 0$, Lee bounds primarily capture the sorting effect:

$$\begin{aligned}\mathbb{E}(Y_{1,D_1} - Y_{0,D_0} | D_0 > 0, D_1 > 0) = \frac{p_{L,H}}{\mathbb{P}(AE)} \mathbb{E}[Y_{1,H} - Y_{0,L} | T = (L, H)] \\ + \frac{p_{H,L}}{\mathbb{P}(AE)} \mathbb{E}[Y_{1,L} - Y_{0,H} | T = (H, L)].\end{aligned}$$

[2]By construction, the outcomes (wages) simulated here are independent of the dataset used in Lee (2009).



This figure plots the identified set for $(p_{L,L}, p_{H,H})$ under various assumptions, as indicated, for our simulation designs (the plots are identical for both designs). The right panel is a magnified view of the boxed region in the left panel. Lighter regions correspond to weaker restrictions and contain the darker regions. Under the Logit and Probit assumptions, $(p_{L,L}, p_{H,H})$ is point-identified.

FIGURE 2. Identified set for $(p_{L,L}, p_{H,H})$ in both simulated DGPs.

TABLE 3. Multilayered for Simulation Designs.

			$\mathbb{E}(Y_{1,H} - Y_{0,H} \mid T = (H,H))$		$\mathbb{E}(Y_{1,L} - Y_{0,L} \mid T = (L,L))$	
Design 1	$p^*_{H,H}$	$p^*_{L,L}$	lower	upper	lower	upper
Baseline	0.0702	0.0000	-2.8749	3.4001	Trivial Bounds	
$p_{H,H} \geq p_{H,L}$	0.1865	0.0924	-1.5956	1.9588	-2.3438	2.2672
(H, L) is smallest type	0.3610	0.2789	-0.1698	0.4873	-0.3958	0.3458
$p_{H,L} = 0$	0.3730	0.2789	-0.0382	0.3591	-0.3958	0.3458
Logit	0.3730	0.2990	-0.0382	0.3591	-0.1628	0.1095
Probit	0.3730	0.2989	-0.0382	0.3591	-0.1643	0.1110
Design 2						
Baseline	0.0702	0.0000	-0.8954	4.9799	Trivial Bounds	
$p_{H,H} \geq p_{H,L}$	0.1865	0.0924	0.3699	3.7242	-1.3438	3.2672
(H, L) is smallest type	0.3610	0.2789	1.7503	2.3440	0.6042	1.3458
$p_{H,L} = 0$	0.3730	0.2789	1.8743	2.2196	0.6042	1.3458
Logit	0.3730	0.2990	1.8743	2.2196	0.8372	1.1095
Probit	0.3730	0.2989	1.8743	2.2196	0.8357	1.1110

Notes: Outcome is simulated log earnings. p^*_t is the minimum value of p_t over the identified set for response-types under the given assumption.

Table 4. Aggregate multilayered bounds for numerical illustrations.

	p^*_{HH}	p^*_{LL}	p^*_{HH}	p^*_{LL}	$\sum_{d\in\{L,H\}} \frac{p_{d,d}}{p_{HH}+p_{LL}} \mathbb{E}[Y_{1,d}-Y_{0,d}\|T=(d,d)]$	
Design 1	(for lower bound)		(for upper bound)		lower	upper
Baseline	0.0758	0.0056	0.0723	0.0021	-2.9349	3.4206
$p_{H,H} \geq p_{H,L}$	0.1865	0.0924	0.1865	0.0924	-1.8434	2.0609
(H, L) is smallest type	0.3610	0.2789	0.3610	0.2789	-0.2683	0.4256
$p_{H,L} = 0$	0.3730	0.2789	0.3730	0.2789	-0.1912	0.3534
Logit	0.3730	0.2990	0.3730	0.2990	-0.0937	0.2481
Probit	0.3730	0.2989	0.3730	0.2989	-0.0943	0.2487
Design 2						
Baseline	0.0823	0.0122	0.0715	0.0013	-1.0519	4.9918
$p_{H,H} \geq p_{H,L}$	0.1865	0.0924	0.1865	0.0924	-0.1977	3.5728
(H, L) is smallest type	0.3610	0.2789	0.3610	0.2789	1.2509	1.9090
$p_{H,L} = 0$	0.3730	0.2789	0.3730	0.2789	1.3310	1.8458
Logit	0.3730	0.2990	0.3730	0.2990	1.4129	1.7257
Probit	0.3730	0.2989	0.3730	0.2989	1.4123	1.7264

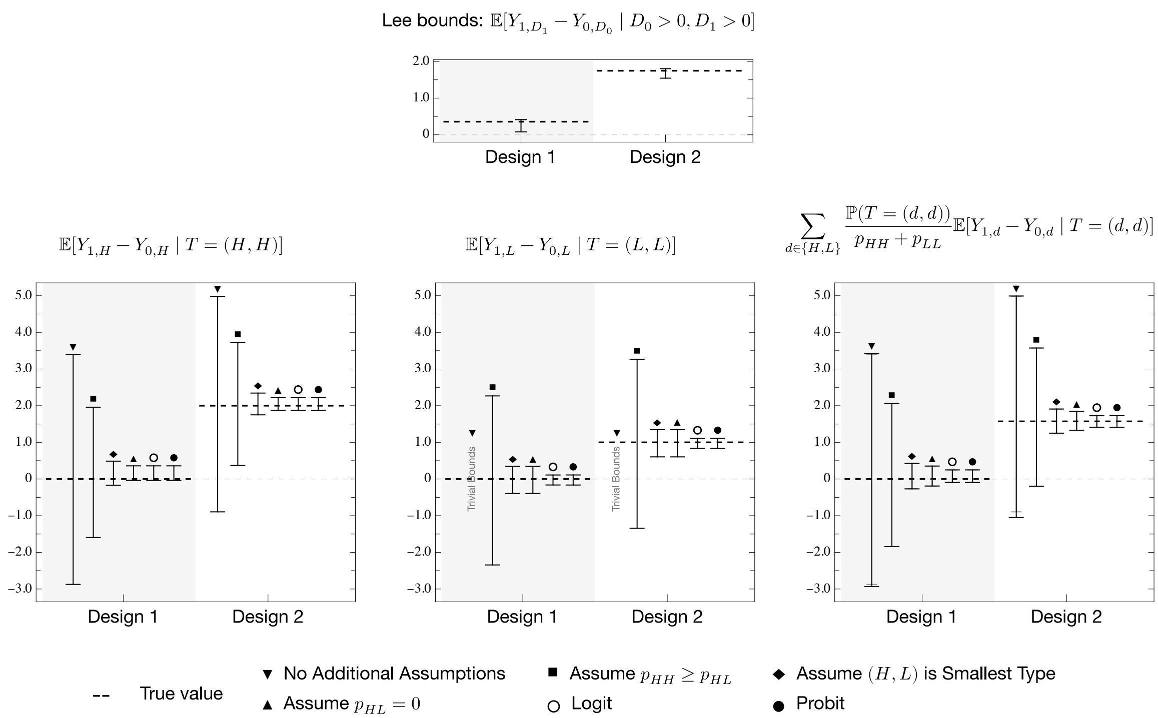
Notes: Outcome is simulated log earnings. p^*_t is the optimal value of p_t over the identified set for response-types under the given assumption; p^*_t may differ for lower and upper bounds and is therefore presented separately.

Our bounds on the within-firm effects $\mathbb{E}(Y_{1,H}-Y_{0,H}|T=(H,H))$ and $\mathbb{E}(Y_{1,L}-Y_{0,L}|T=(L,L))$ always include 0 across the different scenarios (which correspond to different assumptions on response types).

In design 2, Lee bounds also lie entirely within the positive quadrant, excluding 0. However, in this case, the DGP is consistent with the true within-firm effects being strictly positive. Interestingly, our bounds for $\mathbb{E}(Y_{1,H}-Y_{0,H}|T=(H,H))$ and $\mathbb{E}(Y_{1,L}-Y_{0,L}|T=(L,L))$ lie entirely within the positive quadrant, excluding 0, when restricting the response types, showing that they are informative enough to reveal the effect of job training on wages.

Appendix D. Job Corps Study: Additional Empirical Tables and Figures for Job Corps RCT

This section provides institutional background and summary statistics for the Job Corps program, alongside the empirical evidence of treatment-induced sorting into high-type (amenity-providing) firms referenced in the main text. Additionally, it details our replication of Lee's bounds and then presents the identified sets for response



Notes: Dashed black line indicates the true value of the parameter. In all panels, the solid black lines indicate sharp bounds; in the final panel, the gray lines indicate the (generally invalid) result of the "naïve approach" of taking the weighted average of firm-level bounds (see Remark 4).

FIGURE 3. Lee (2009) Bounds and Multilayered Bounds in Simulations.

types for weeks 135, 180 and 208 of the Job Corps study. Finally, it provides tabular counterparts to the bounds figures displayed in the main text.

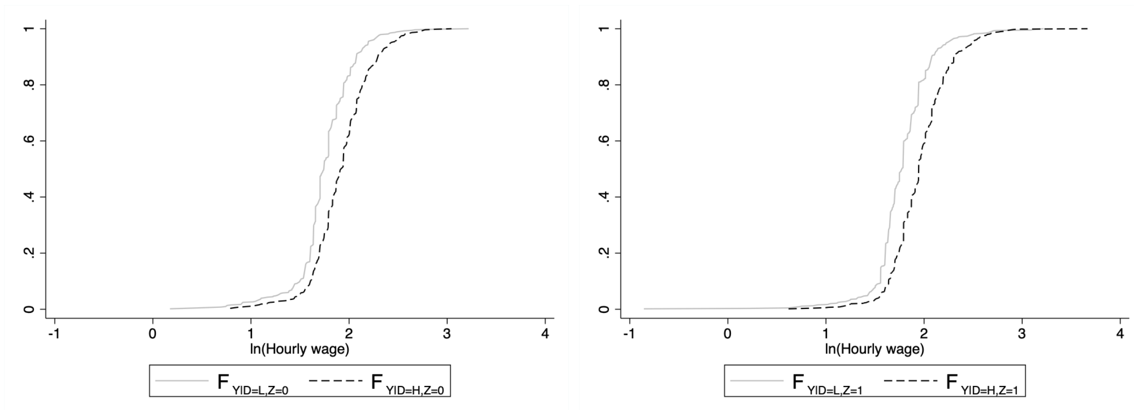
D.1. **Job Corps program.** Job Corps is the largest residential career training program in the U.S. It is free for participants and targets disadvantaged people aged 16 to 24 with the aim of helping these people become more responsible, employable, and productive citizens (Johnson et al. 1999). Most participants live at a local Job Corps center and complete 440 hours of academic instruction and 700 hours of vocational training. Job Corps also provides job search assistance upon participant completion of the program. The typical participant completes the program over a span of 30 weeks. Job Corps has trained more than two million individuals since its inception under the Economic Opportunity Act of 1964. The program trains over 60,000 enrollees per year, at roughly 130 Job Corps centers nationwide, with an estimated cost of 34,301 USD per enrollee and 57,312 USD per graduate (Liu et al. 2020).[3]

D.2. **Job Corps Study.** During the mid- to late 1990s, the U.S. Department of Labor funded a randomized evaluation of Job Corps, implemented by Mathematica Policy Research, Inc. Existing evaluations of the Job Corps Study include Schochet et al. (2001), Schochet et al. (2008), Lee (2009), and Blanco et al. (2013). The Job Corps Study randomized 80,883 eligible individuals who applied to Job Corps for the first time between November 1994 and December 1995 into two groups: (i) 5,977 individuals into the control group (embargoed from participating in Job Corps for three years) and (ii) 74,906 individuals into the treatment group. Of the 74,906 individuals assigned to treatment, 9,409 were randomly selected for data collection. All control individuals were selected for data collection. The final sample consists of 15,386 participants who were interviewed at the time of random assignment and then subsequently 12, 30, and 48 months after random assignment.

D.3. **Summary Statistics.** Table 5 presents summary statistics for the sample of individuals who have non-missing values for weekly earnings and hours for every week following random assignment. Means and standard deviations for a number of baseline and post-randomization variables are reported separately by treatment status. Consistent with successful randomization in the National Job Corps Study, the table shows that there are no statistical differences in the means of demographic, education, background and baseline employment/income variables across treatment and control groups. Table 5 also shows economically and statistically significant differences in employment and earnings outcomes by treatment status, post randomization. After 208 weeks, the hours and earnings of the treatment group are approximately 8% and 14% higher, respectively, than those of the control group.

[3]Included in these costs are direct transfers to Job Corps participants. The average participant in the 1995 Job Corps randomized evaluation (detailed below) received 2,361 USD in direct transfers, which consisted of 1,427 USD in pay, 704 USD of food and 230 USD of clothing; these transfers accounted for approximately 14% of the cost per participant at the time (McConnell and Glazerman 2001).

D.4. **Joint Distribution of Wages and Amenities.** Figure 4 presents the empirical cumulative distribution functions of log wage by firm type, classifying firms according to the provision of health insurance, for treatment and control groups at week 90.[4,5] For both treated and control units, the distribution of log wages for firms that provide health insurance stochastically dominates the distribution of log wages for firms that do not.[6] This suggests that firms providing amenities pay higher wages than firms that do not.[7]



(a) control units (b) treated units

Notes: Hourly wage calculated as weekly earnings divided by weekly hours for the employed.

FIGURE 4. Cumulative distribution function by firm type (provision of health insurance) at week 90. RCT = Job Corps Study.

[4]At the time of the National Job Corps study, there were no legal requirements for firms to provide health insurance and, conditional on firm provision, federal law generally prohibited discriminatory provision across workers (United States Equal Employment Opportunity Commission 2009). The relevant federal laws at the time of the National Job Corps Study included the following. Title VII of the Civil Rights Act of 1964; the Age Discrimination in Employment Act of 1967; Title I and Title V of the Americans with Disabilities Act of 1990 (United States Equal Employment Opportunity Commission 2009).

[5]We present results for week 90, following preferred specification of Lee (2009); results for alternative weeks are available upon request.

[6]At week 90, mean wages at firms that offered amenities were approximately 15% higher compared to firms that did not.

[7]Of course, it is possible that firms pay compensating differentials which causally reduce wages. The evidence presented here shows that the cross-sectional variation across firms dominates the variation within firms. This is consistent with evidence in Lamadon et al. (2022) who show that high-amenities firms are also more productive firms.

D.5. **Differential Sorting of Treatment and Control Workers.** Table 6 presents the probability of working in a firm (conditional on employment) that provides observable amenities, at week 90, according to the status of treatment. The evidence shows that treated individuals are more likely to work at firms with job amenities in all but one case (and we also find that this trend persists across all weeks). This is consistent with the evidence presented in Schochet et al. (2008).

D.6. **Replication of Lee (2009) Bounds.** Table 7 reports our replication of the bounds reported in Lee (2009) for weeks 90, 135, 180 and 208. We do not report bounds for week 45 since we discovered that the monotonicity assumption is violated. Table 7 reports Lee's bounds when treating ln(hourly wage) as a continuous variable (as we do throughout the paper). All quantities are very close to the estimates in Lee (2009), though a small difference arises in the bounds due to Lee's use of vingtiles of ln(hourly wage). Table 8 shows that when we use vingtiles of ln(hourly wage), the bounds are identical to the ones reported in Lee (2009).

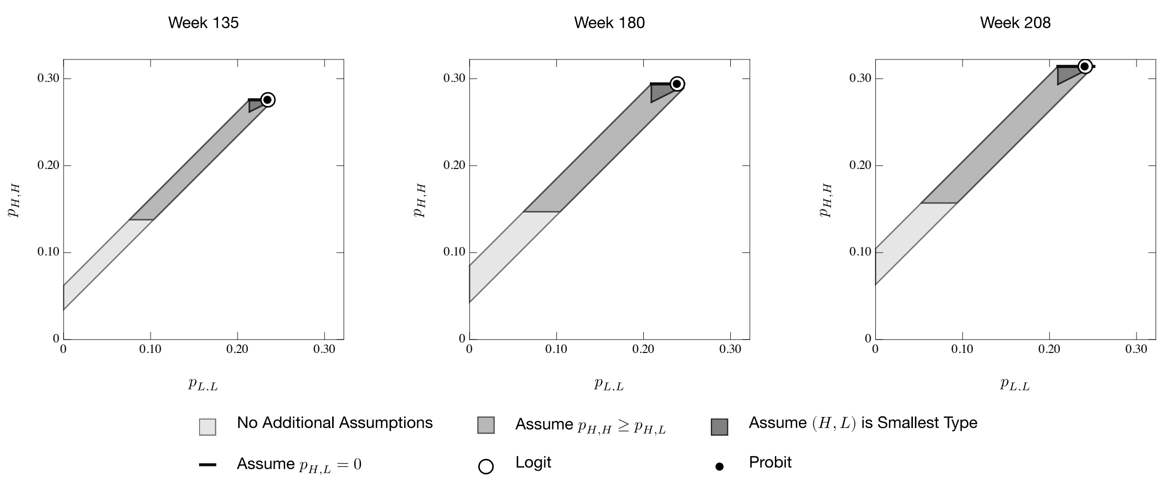


FIGURE 5. Identified set for $(p_{L,L}, p_{H,H})$ at weeks 135, 180 and 208. RCT = Job Corps Study.

Appendix E. Description of WorkAdvance RCTs and Additional Empirical Tables and Figures

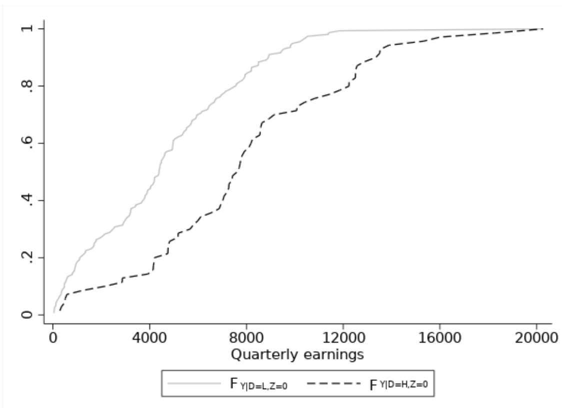
This section expands on the WorkAdvance RCTs by providing additional information on program curricula and summary statistics, alongside the empirical evidence of treatment-induced sorting into high-type (target-sector) firms referenced in the main text. Finally, it presents tabular counterparts to the bounds figures displayed in the main text.

E.1. **WorkAdvance Program and RCTs.** As detailed in the main text, our analysis focuses on the Madison Strategies and Towards Employment WorkAdvance RCTs. Although they targeted different in-demand sectors, the two programs shared a highly similar core structure. Both ranged from 4 to 32 weeks in duration, with curricula centered on pre-employment career readiness and sector-specific skills training. Each provider also integrated sector-specific job placement assistance with post-employment retention and advancement services. Upon completion of both programs, participants received the appropriate certification. While both evaluations targeted low-income adults, their screening requirements differed slightly: Madison Strategies required applicants to test at an eighth-grade level, pass a behavioral assessment, pass mechanical aptitude and manual dexterity exams, and hold a valid driver's license; by contrast, Towards Employment required testing at a sixth- to tenth-grade level alongside passing background and drug screenings.

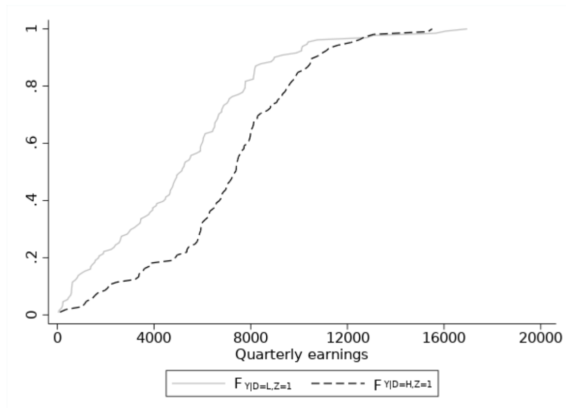
E.2. **Summary Statistics.** Tables 11 and 12 present summary statistics for the Madison Strategies and Towards Employment RCTs, respectively. Means and standard deviations for a number of baseline and post-randomization variables are reported separately by treatment status. Consistent with successful randomization, there are no statistical differences in the means of demographic, education, and baseline employment/income variables across treatment and control groups.

E.3. **Joint Distribution of Wages and Amenities.** Figure 6 presents the empirical cumulative distribution functions of quarterly earnings for the treatment and control groups, with firms classified by whether they operate within the target sector. Complementing our earlier evidence for amenity-providing firms in the Job Corps study, the log wage distribution for target-sector firms stochastically dominates that of non-target-sector firms across both treatment and control groups, in both WorkAdvance RCTs.

RCT = Madison Strategies

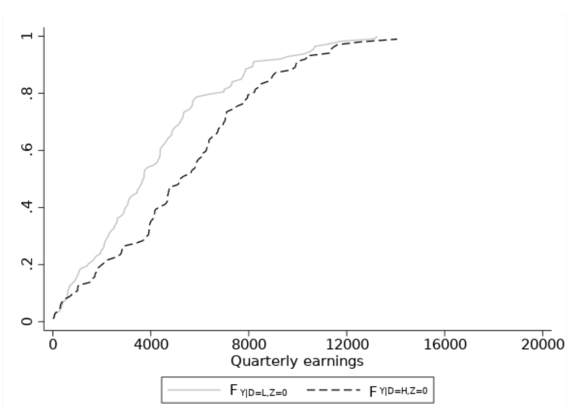


(a) control units

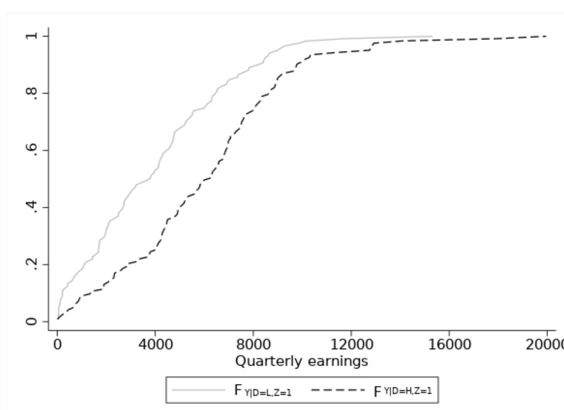


(b) treated units

RCT = Towards Employment



(c) control units



(d) treated units

Notes: Quarterly earnings are for the employed (i.e., does not include 0s for the unemployed).

Figure 6. Cumulative distribution function by firm sector. RCT = WorkAdvance.

	Control		Treated		Difference	
	Mean	S.D.	Mean	S.D.	Difference	S.E.
Female	0.46	0.50	0.45	0.50	-0.01	0.01
Age at baseline	18.35	2.10	18.44	2.16	0.09	0.05
White, non-Hispanic	0.26	0.44	0.27	0.44	0.00	0.01
Black, non-Hispanic	0.49	0.50	0.49	0.50	0.00	0.01
Hispanic	0.17	0.38	0.17	0.37	-0.00	0.01
Other race/ethnicity	0.07	0.26	0.07	0.26	-0.00	0.01
Never married	0.92	0.28	0.92	0.28	0.00	0.01
Married	0.02	0.15	0.02	0.14	-0.00	0.00
Living together	0.04	0.20	0.04	0.19	-0.00	0.00
Separated	0.02	0.14	0.02	0.15	0.00	0.00
Has child	0.19	0.39	0.19	0.39	-0.00	0.01
Number of children	0.27	0.64	0.27	0.65	0.00	0.01
Education	10.11	1.54	10.11	1.56	0.01	0.03
Mother's education	11.46	2.59	11.48	2.56	0.02	0.06
Father's education	11.54	2.79	11.39	2.85	-0.15	0.08
Ever arrested	0.25	0.43	0.25	0.43	-0.00	0.01
Household income						
<3,000	0.25	0.43	0.25	0.44	0.00	0.01
3,000-6,000	0.21	0.41	0.21	0.40	-0.00	0.01
6,000-9,000	0.11	0.32	0.12	0.32	0.00	0.01
9,000-18,000	0.24	0.43	0.24	0.43	-0.00	0.01
>18,000	0.18	0.39	0.18	0.38	-0.00	0.01
Personal income						
<3,000	0.79	0.41	0.79	0.41	-0.00	0.01
3,000-6,000	0.13	0.34	0.13	0.33	-0.00	0.01
6,000-9,000	0.05	0.21	0.05	0.22	0.01	0.00
>9,000	0.03	0.18	0.03	0.17	-0.00	0.00
At baseline						
Have job	0.19	0.39	0.20	0.40	0.01	0.01
Mths. empl. prev. yr.	3.53	4.24	3.60	4.25	0.07	0.09
Had job, prev. yr.	0.63	0.48	0.63	0.48	0.01	0.01
Earnings, prev. yr.	2810.48	4435.62	2906.45	6401.33	95.97	117.10
Usual hours/week	20.91	20.70	21.82	21.05	0.91	0.45
Usual weekly earn.	102.89	116.46	110.99	350.61	8.10	5.09
Post randomization						
Week 52 hours	17.78	23.39	15.30	22.68	-2.49	0.49
Week 104 hours	21.98	26.08	22.64	26.25	0.67	0.56
Week 156 hours	23.88	26.15	25.88	26.57	2.00	0.56
Week 208 hours	25.83	26.25	27.79	25.74	1.95	0.56
Week 52 earn.	103.80	159.89	91.55	149.28	-12.25	3.33
Week 104 earn.	150.41	210.24	157.42	200.27	7.02	4.42
Week 156 earn.	180.88	224.43	203.71	239.80	22.84	4.94
Week 208 earn.	200.50	230.66	227.91	250.22	27.41	5.11
Total 4 yr. earn.	30006.69	26893.60	30800.41	26437.39	793.72	571.83
Sample size	3599		5546		9145	

Notes: Weekly earnings calculated as the sum of total earnings in a given week and are not conditional on employment (i.e., includes 0s for the unemployed).

TABLE 5. Summary statistics by treatment status. RCT = Job Corps Study.

	Control		Treated		Difference	
	Mean	S.D.	Mean	S.D.	Difference	S.E.
Health insurance	0.4860	0.5000	0.5072	0.5001	0.0211	0.0186
Paid sick leave	0.3922	0.4884	0.4251	0.4945	0.0329	0.0183
Paid vacation	0.5407	0.4985	0.5800	0.4937	0.0393	0.0180
Childcare assistance	0.1311	0.3376	0.1422	0.3494	0.0111	0.0127
Flexible hours	0.5330	0.4991	0.5568	0.4969	0.0237	0.0183
Employer-provided transportation	0.1973	0.3981	0.1876	0.3905	-0.0097	0.0147
Pension or retirement benefits	0.3670	0.4822	0.3938	0.4887	0.0268	0.0180
Dental plan	0.3863	0.4871	0.4288	0.4950	0.0425	0.0182
Tuition reimbursement	0.2212	0.4152	0.2602	0.4389	0.0390	0.0158
Employment	0.4368	0.4961	0.4391	0.4963	0.0022	0.0111
Sample size	3288		5091		8379	

Notes: Control and treatment probabilities have interpretation as $P[D = H|D > 0, Z = z]$ for $z \in \{0, 1\}$, respectively, when classifying firms as type H if they provide a given amenity.

TABLE 6. Probability of working at amenity-providing firm at week 90 (conditional on employment). RCT = Job Corps Study.

				$\mathbb{E}[Y_{1,D_1} - Y_{0,D_0}\|D_0 > 0, D_1 > 0]$	
	$\mathbb{P}[D > 0\|Z = 0]$	$\mathbb{P}[D > 0\|Z = 1]$	Trimming Proportion	lower	upper
Week 90	0.4600	0.4601	0.0003	0.0468	0.0484
Week 135	0.5173	0.5451	0.0509	-0.0072	0.0842
Week 180	0.5403	0.5825	0.0724	-0.0325	0.0901
Week 208	0.5655	0.6068	0.0680	-0.0217	0.0989

Notes: Treatment bounds are for ln(hourly wage), where hourly wage equals weekly earnings divided by weekly hours for the employed. Propensity scores and trimming proportions are numerically equivalent to Lee (2009). The trimming proportion, which Lee reports, is equal to $1 - p$ where $p \equiv \mathbb{P}(AE)$, e.g., the share of always-employed among employed individuals receiving job training. The slight numerical difference in bounds arises as Lee uses vingtiles of ln(hourly wage); these are presented in Table 8.

TABLE 7. Lee's bounds: continuous ln(hourly wage). RCT = Job Corps Study.

				$\mathbb{E}[Y_{1,D_1} - Y_{0,D_0}\|D_0 > 0, D_1 > 0]$	
	$\mathbb{P}[D > 0\|Z = 0]$	$\mathbb{P}[D > 0\|Z = 1]$	Trimming Proportion	lower	upper
Week 90	0.4600	0.4601	0.0003	0.0423	0.0428
Week 135	0.5173	0.5451	0.0509	-0.0159	0.0757
Week 180	0.5403	0.5825	0.0724	-0.0325	0.0868
Week 208	0.5655	0.6068	0.0680	-0.0194	0.0933

Notes: Treatment bounds are for vingtiles of ln(hourly wage), where hourly wage equals weekly earnings divided by weekly hours for the employed. Propensity scores and trimming proportions are numerically equivalent to Lee (2009). The trimming proportion, which Lee reports, is equal to $1 - p$ where $p \equiv \mathbb{P}(AE)$, e.g., the share of always-employed among employed individuals receiving job training.

TABLE 8. Lee's bounds: vingtiles of ln(hourly wage). RCT = Job Corps Study.

Week 90	$p^*_{H,H}$	$p^*_{L,L}$	$\mathbb{E}(Y_{1,H}-Y_{0,H}\mid T=(H,H))$ lower	upper	$\mathbb{E}(Y_{1,L}-Y_{0,L}\mid T=(L,L))$ lower	upper
Baseline	0.0010	0.0000	-2.1415	2.3907		
$p_{H,H} \geq p_{H,L}$	0.1120	0.1108	-0.4214	0.5020	-0.4002	0.4542
(H,L) is smallest type	0.2239	0.2228	-0.0023	0.0754	-0.0191	0.0673
$p_{H,L}=0$	0.2239	0.2228	-0.0018	0.0750	-0.0191	0.0673
Logit	0.2239	0.2229	-0.0018	0.0750	-0.0180	0.0656
Probit	0.2239	0.2229	-0.0018	0.0750	-0.0180	0.0656
Week 135						
Baseline	0.0344	0.0000	-1.1228	1.1454		
$p_{H,H} \geq p_{H,L}$	0.1379	0.0758	-0.5075	0.5433	-0.6064	0.7090
(H,L) is smallest type	0.2619	0.2137	-0.1135	0.1369	-0.1180	0.1672
$p_{H,L}=0$	0.2758	0.2137	-0.0529	0.0732	-0.1180	0.1672
Logit	0.2758	0.2349	-0.0529	0.0732	-0.0326	0.0847
Probit	0.2758	0.2343	-0.0529	0.0732	-0.0367	0.0881
Week 180						
Baseline	0.0429	0.0000	-1.0454	1.1410		
$p_{H,H} \geq p_{H,L}$	0.1471	0.0620	-0.5063	0.5525	-0.7710	0.8503
(H,L) is smallest type	0.2730	0.2090	-0.1421	0.1704	-0.1806	0.2110
$p_{H,L}=0$	0.2941	0.2090	-0.0552	0.0854	-0.1806	0.2110
Logit	0.2941	0.2389	-0.0552	0.0854	-0.0675	0.1020
Probit	0.2941	0.2382	-0.0552	0.0854	-0.0712	0.1056
Week 208						
Baseline	0.0633	0.0000	-0.8821	0.9895		
$p_{H,H} \geq p_{H,L}$	0.1571	0.0525	-0.4888	0.5670	-0.9059	0.8730
(H,L) is smallest type	0.2935	0.2096	-0.1217	0.1797	-0.1914	0.2082
$p_{H,L}=0$	0.3142	0.2096	-0.0430	0.1016	-0.1914	0.2082
Logit	0.3142	0.2411	-0.0430	0.1016	-0.0771	0.0952
Probit	0.3142	0.2403	-0.0430	0.1016	-0.0809	0.0995

Notes: Treatment bounds are for ln(hourly wage); hourly wage calculated as weekly earnings divided by weekly hours for the employed. p^*_t is the minimum value of p_t over $\Theta_I(\mathcal{R}_T)$, for the corresponding $\mathcal{R}_T$.

TABLE 9. Multilayered bounds based on health insurance amenity. RCT = Job Corps Study.

	p^*_{HH}	p^*_{LL}	p^*_{HH}	p^*_{LL}	$\sum_{d\in\{L,H\}} \frac{p_{d,d}}{p_{HH}+p_{LL}} \mathbb{E}[Y_{1,d}-Y_{0,d}\|T=(d,d)]$	
Week 90	(for lower bound)		(for upper bound)		lower	upper
Baseline	0.0014	0.0004	0.0014	0.0004	-2.4634	2.4588
$p_{H,H} \geq p_{H,L}$	0.1120	0.1109	0.1120	0.1109	-0.4106	0.4780
(H,L) is smallest type	0.2235	0.2224	0.2235	0.2224	-0.0136	0.0749
$p_{H,L} = 0$	0.2235	0.2224	0.2235	0.2224	-0.0136	0.0749
Logit	0.2239	0.2229	0.2239	0.2229	-0.0099	0.0703
Probit	0.2239	0.2229	0.2239	0.2229	-0.0099	0.0703
Week 135						
Baseline	0.0366	0.0022	0.0392	0.0048	-1.1602	1.2762
$p_{H,H} \geq p_{H,L}$	0.1380	0.0758	0.1380	0.0758	-0.5423	0.6017
(H,L) is smallest type	0.2614	0.2127	0.2614	0.2127	-0.1179	0.1532
$p_{H,L} = 0$	0.2754	0.2132	0.2754	0.2132	-0.0833	0.1165
Logit	0.2758	0.2349	0.2758	0.2349	-0.0436	0.0785
Probit	0.2758	0.2343	0.2758	0.2343	-0.0454	0.0800
Week 180						
Baseline	0.0475	0.0046	0.0485	0.0056	-1.1369	1.2631
$p_{H,H} \geq p_{H,L}$	0.1471	0.0620	0.1471	0.0620	-0.5846	0.6407
(H,L) is smallest type	0.2727	0.2081	0.2727	0.2081	-0.1607	0.1900
$p_{H,L} = 0$	0.2937	0.2086	0.2937	0.2086	-0.1101	0.1397
Logit	0.2941	0.2389	0.2941	0.2389	-0.0607	0.0929
Probit	0.2941	0.2382	0.2941	0.2382	-0.0624	0.0945
Week 208						
Baseline	0.0705	0.0072	0.0671	0.0038	-0.9987	1.0704
$p_{H,H} \geq p_{H,L}$	0.1571	0.0525	0.1571	0.0525	-0.5932	0.6436
(H,L) is smallest type	0.2928	0.2087	0.2928	0.2087	-0.1532	0.1941
$p_{H,L} = 0$	0.3138	0.2092	0.3138	0.2092	-0.1049	0.1461
Logit	0.3142	0.2411	0.3142	0.2411	-0.0578	0.0988
Probit	0.3142	0.2403	0.3142	0.2403	-0.0594	0.1007

Notes: Treatment bounds are for ln(hourly wage); hourly wage calculated as weekly earnings divided by weekly hours for the employed. p^*_t is the optimal value of p_t over the joint identified set for response-types under the given assumption; p^*_t may differ for lower and upper bounds and is therefore presented separately. Because p^*_t values are determined by a grid search, the values in columns 2-4 should be read only as the grid point used in the numerical approximation, not as an exact feasible response-type probability.

Table 10. Aggregate multilayered bounds based on health insurance amenity. RCT = Job Corps Study.

	Control		Treated		Difference	
	Mean	S.D.	Mean	S.D.	Difference	S.E.
Female	0.14	0.34	0.18	0.39	0.05	0.03
Adults $\leq$ 24	0.22	0.42	0.23	0.42	0.01	0.03
Black	0.35	0.48	0.39	0.49	0.03	0.04
HS/GED or less	0.40	0.49	0.43	0.50	0.03	0.04
At baseline:						
Have job	0.28	0.45	0.26	0.44	-0.02	0.03
Quarterly earnings	2257.07	2820.54	1944.51	2495.57	-312.56	201.91
8 quarters post randomization:						
Employment	0.66	0.48	0.67	0.47	0.01	0.04
Share employment in target sector	0.31	0.46	0.44	0.50	0.14	0.04
Quarterly earnings	3703.87	4169.10	4049.91	4003.27	346.04	309.72
Sample size	344		353		697	

Notes: Quarterly earnings are not conditional on employment (i.e., includes 0s for the unemployed).

Table 11. Summ. stats. by treatment status, RCT: Madison Strategies

	Control		Treated		Difference	
	Mean	S.D.	Mean	S.D.	Difference	S.E.
Female	0.59	0.49	0.58	0.49	-0.01	0.04
Adults $\leq$ 24	0.23	0.42	0.22	0.42	-0.01	0.03
Black	0.73	0.45	0.77	0.42	0.05	0.03
HS/GED or less	0.42	0.49	0.44	0.50	0.03	0.04
At baseline:						
Have job	0.27	0.45	0.26	0.44	-0.01	0.03
Quarterly earnings	1504.59	2201.72	1723.72	2452.60	219.12	176.42
8 quarters post randomization:						
Employment	0.62	0.49	0.69	0.46	0.08	0.04
Share employment in target sector	0.47	0.50	0.51	0.50	0.03	0.05
Quarterly earnings	3040.57	4024.21	3497.54	3684.01	456.97	292.04
Sample size	349		349		698	

Notes: Quarterly earnings are not conditional on employment (i.e., includes 0s for the unemployed).

Table 12. Summ. stats. by treatment status, RCT: Towards Employment

				$\mathbb{E}[Y_{1,D_1} - Y_{0,D_0} \| D_0 > 0, D_1 > 0]$	
	$\mathbb{P}[D > 0 \| Z = 0]$	$\mathbb{P}[D > 0 \| Z = 1]$	Trimming Proportion	lower	upper
Madison Strategies	0.6570	0.6686	0.0173	244.1928	524.9793
Towards Employment	0.6160	0.6934	0.1116	-676.8651	705.6605

Notes: Treatment bounds are for quarterly wages. The trimming proportion, which Lee (2009) reports, is equal to $1 - p$ where $p \equiv \mathbb{P}(AE)$, e.g., the share of always-employed among employed individuals receiving job training.

Table 13. Lee's bounds. RCTs = WorkAdvance.

			$\mathbb{E}(Y_{1,H} - Y_{0,H} \mid T = (H,H))$		$\mathbb{E}(Y_{1,L} - Y_{0,L} \mid T = (L,L))$	
Madison Strategies	$p^*_{H,H}$	$p^*_{L,L}$	lower	upper	lower	upper
Baseline	0.0000	0.1560			-6172.4634	7139.9806
$p_{H,H} \geq p_{H,L}$	0.1017	0.2578	-7380.1460	5628.8654	-3353.0004	4562.1007
(H,L) is smallest type	0.1977	0.3595	-2652.6847	1216.5151	-798.7005	2016.2079
$p_{H,L} = 0$	0.2035	0.3595	-2353.4366	806.8709	-798.7005	2016.2079
Logit	0.2035	0.3711	-2353.4366	806.8709	-328.9279	1718.8632
Probit	0.2035	0.3711	-2353.4366	806.8709	-328.9279	1718.8632
Towards Employment						
Baseline	0.0000	0.0000				
$p_{H,H} \geq p_{H,L}$	0.1461	0.1175	-5784.7451	6488.1176	-6510.9268	5975.5610
(H,L) is smallest type	0.2536	0.2636	-2044.9605	3064.8644	-2335.5870	1932.3913
$p_{H,L} = 0$	0.2923	0.2636	-740.4216	1356.9608	-2335.5870	1932.3913
Logit	0.2923	0.3106	-740.4216	1356.9608	-1006.6086	464.6827
Probit	0.2923	0.3103	-740.4216	1356.9608	-1014.3705	474.0062

Notes: Outcome is quarterly wages. p^*_t is the optimal value of p_t over the identified set for response-types under the given assumption.

TABLE 14. Multilayered bounds. RCTs = WorkAdvance.

	p^*_{HH}	p^*_{LL}	p^*_{HH}	p^*_{LL}	$\sum_{d \in \{L,H\}} \frac{p_{d,d}}{p_{HH} + p_{LL}} \mathbb{E}[Y_{1,d} - Y_{0,d} \mid T = (d,d)]$	
Madison Strategies	(for lower bound)		(for upper bound)		lower	upper
Baseline	0.0145	0.1710	0.0090	0.1655	-6498.0976	7255.9812
$p_{H,H} \geq p_{H,L}$	0.1010	0.2570	0.1010	0.2570	-4514.3517	4886.1844
(H,L) is smallest type	0.1960	0.3545	0.1960	0.3545	-1593.0171	1848.4113
$p_{H,L} = 0$	0.2010	0.3570	0.2010	0.3570	-1459.5360	1688.9594
Logit	0.2035	0.3711	0.2035	0.3711	-1045.8942	1395.8872
Probit	0.2035	0.3711	0.2035	0.3711	-1045.8942	1395.8872
Towards Employment						
Baseline	0.0000	0.0000	0.0000	0.0000		
$p_{H,H} \geq p_{H,L}$	0.1450	0.1170	0.1450	0.1170	-6142.9166	6288.9131
(H,L) is smallest type	0.2500	0.2590	0.2500	0.2590	-2318.5475	2613.7796
$p_{H,L} = 0$	0.2900	0.2620	0.2900	0.2620	-1562.2141	1840.9057
Logit	0.2923	0.3106	0.2923	0.3106	-877.5657	897.2439
Probit	0.2923	0.3103	0.2923	0.3103	-881.5065	902.2352

Notes: Outcome is quarterly wages. p^*_t is the optimal value of p_t over the joint identified set for response-types p^*_t under the given assumption; p^*_t may differ for lower and upper bounds and is therefore presented separately. Because p^*_t values are determined by a grid search, the values in columns 2-4 should be read only as the grid point used in the numerical approximation, not as an exact feasible response-type probability.

TABLE 15. Aggregate multilayered bounds. RCTs = WorkAdvance

Appendix F. 'Top 5' papers with multilayered sample selection

In the literature survey we conducted, referenced in the introduction, we counted 56 papers published in 'top 5' general interest economic journals that cited Lee (2009) and 42 that empirically implemented Lee bounds.[8] This section details 6 of these papers that feature multilayered selection, where researchers simplified the sample selection problem by collapsing it to a single dimension.

Daruich et al. (2023) Studies a 2001 Italian reform which lifted constraints on the employment of temporary contract workers but maintained employment protection laws for permanent contract employees. The outcome is individual earnings. Lee bounds are employed to address concern that the reform affected employment. Conditional on labor market entry, the reform can affect worker sorting across industries and/or firms; indeed the paper finds meaningful changes in the shares of temporary contract workers in certain industries.

Cullen and Pakzad-Hurson (2023) Studies state-level laws in the U.S. protecting the right of private sector workers to discuss salary information with co-workers. The main outcome is worker wages. They estimate Lee bounds to address sample selection into employment. The treatment may also affect worker sorting across firms; for example, knowledge of co-workers' salaries could cause workers to sort to firms with flatter pay hierarchies.

Bianchi and Giorcelli (2022) Studies the impact of the Training Within Industry program, a U.S. government training program intended to be provided to all firms involved in war production between 1940 and 1945. The main outcome is firm total factor productivity. They estimate Lee bounds to address the higher attrition rate of untrained firms: treated firms had 90% survival rate at least 10 years following treatment whereas control firms only had 64% survival rate. Conditional on firm survival, training may also affect the sorting of firms across industries or other important dimensions. This is particularly plausible in this paper's setting where: (i) trained firms undertook structural changes transforming them into larger and more complex

[8]The 'top 5' refer to the American Economic Review, Econometrica, the Journal of Political Economy, the Quarterly Journal of Economics and the Review of Economic Studies.

organizations; (ii) this paper estimates treatment bounds for outcomes observed post-Second World War, when many firms would have plausibly switched industries (i.e., left war production).

Fink et al. (2020) Evaluate an experiment offering subsidized loans to randomly selected villages in rural Zambia where farmers suffer from liquidity constraints in the months prior to harvest (i.e., the lean season). The outcomes are individual- and village-level earnings. They estimate Lee bounds to address that the likelihood of entering the labor market decreases with the loan treatment. Conditional on entry to the labor market, treatment has potential to affect the types of jobs individuals accept; this is particularly plausible in this paper's setting as labor sales occur within villages between better- and worse-off farmers at individually negotiated rates.

Giorcelli (2019) Studies long-run effects of U.S. Technical Assistance and Productivity Program (USTAPP) which provided management training and technologically advanced machines to Italian firms from 1952 to 1958. Outcomes include firm-level sales, number of employees and total factor productivity revenue. They estimate Lee bounds to address the treatment-control difference in firm survival probability. As in Bianchi and Giorcelli (2022), conditional on firm survival, management training likely affects the sorting of firms across industries.

Fisman et al. (2017) Estimate effect of cultural proximity on loan outcomes for lenders and borrowers using dyadic data on religion and caste for lending officers and borrowers from a state-owned Indian bank. Outcomes include amount of debt received, total number of borrowers and average loan size. They estimate Lee bounds as outcomes are only observed conditional on a group receiving credit. Conditional on a group receiving credit, "same group matches" also plausibly affect the type of loans a group receives. For example, "same group match" borrowers may receive favorable loan terms; this paper indeed notes the potential for these effects but is constrained by data limitations.

References

Bianchi, Nicola and Michela Giorcelli (2022): "The Dynamics and Spillovers of Management Interventions: Evidence from the Training within Industry Program," *Journal of Political Economy*, 130, 1630–1675.

Blanco, German, Carlos A Flores, and Alfonso Flores-Lagunes (2013): "The Effects of Job Corps Training on Wages of Adolescents and Young Adults," *American Economic Review*, 103, 418–422.

Cullen, Zoë B. and Bobak Pakzad-Hurson (2023): "Equilibrium Effects of Pay Transparency," *Econometrica : journal of the Econometric Society*, 91, 765–802.

Daruich, Diego, Sabrina Di Addario, and Raffaele Saggio (2023): "The Effects of Partial Employment Protection Reforms: Evidence from Italy," *Review of Economic Studies*, 90, 2880–2942.

Fink, Günther, B. Kelsey Jack, and Felix Masiye (2020): "Seasonal Liquidity, Rural Labor Markets, and Agricultural Production," *American Economic Review*, 110, 3351–3392.

Fisman, Raymond, Daniel Paravisini, and Vikrant Vig (2017): "Cultural Proximity and Loan Outcomes," *American Economic Review*, 107, 457–492.

Giorcelli, Michela (2019): "The Long-Term Effects of Management and Technology Transfers," *American Economic Review*, 109, 121–152.

Johnson, Terry, Mark Gritz, Russell Jackson, John Burghardt, Carol Boussy, Jan Leonard, and Carlyn Orians (1999): "National Job Corps Study: Report on the Process Analysis," Tech. rep., Mathematica Policy Research, Princeton, NJ.

Lamadon, Thibaut, Magne Mogstad, and Bradley Setzler (2022): "Imperfect Competition, Compensating Differentials, and Rent Sharing in the US Labor Market," *American Economic Review*, 112, 169–212.

Lee, David S. (2009): "Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects," *Review of Economic Studies*, 76, 1071–1102.

Liu, Albert, Clayton Perry, Robbie Skinner, and Scott Cody (2020): "Estimating Job Corps Cost per Enrollee and Cost per Graduate," Tech. rep., Insight Policy Research, Arlington, VA.

McConnell, Sheena and Steven Glazerman (2001): "National Job Corps Study: The Benefits and Costs of Job Corps." Tech. Rep. ED457357, Mathematica Policy Research, Washington, DC.; Battelle Memorial Inst., Seattle, WA.; Decision Information Resources, Inc., Houston, TX.

Pearl, Judea (2001): "Direct and Indirect Effects," Tech. rep.

——— (2009): "Causal Inference in Statistics: An Overview," *Statistics Surveys*, 3.

Schochet, Peter, John Burghardt, and Steven Glazerman (2001): "National Job Corps Study: The Impacts of Job Corps on Participants' Employment and Related Outcomes," Tech. rep., Mathematica Policy Research, Princeton, NJ.

Schochet, Peter Z, John Burghardt, and Sheena McConnell (2008): "Does Job Corps Work? Impact Findings from the National Job Corps Study," *American Economic Review*, 98, 1864–1886.

United States Equal Employment Opportunity Commission (2009): "Federal Laws Prohibiting Job Discrimination Questions and Answers," Tech. rep., United States Equal Employment Opportunity Commission, Washington, DC.