HOROWITZ-MANSKI-LEE BOUNDS WITH MULTILAYERED SAMPLE SELECTION

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

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.

Keywords: job training, sample selection, unordered treatments, sufficient statistics.

JEL subject classification: C12, C14, C21, and C26.

Date: 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.

* Department of Economics, University of Toronto, & NBER. 150 St. George Street, Toronto ON M5S 3G7, Canada, kory.kroft@utoronto.ca.

† Department of Economics, Washington University in St. Louis & NBER, ismaelm@wustl.edu

‡ Department of Economics, University of Toronto, 150 St. George Street, Toronto ON M5S 3G7, Canada, 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: $\mathbb{P}\left[\mathbf{1}\{D_1 > 0\} \geq \mathbf{1}\{D_0 > 0\} | \, X\right] = 1$ 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:*

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

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.,

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

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 $d \in \{1, \ldots, K\}$, 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*

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

(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 (potentially 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.

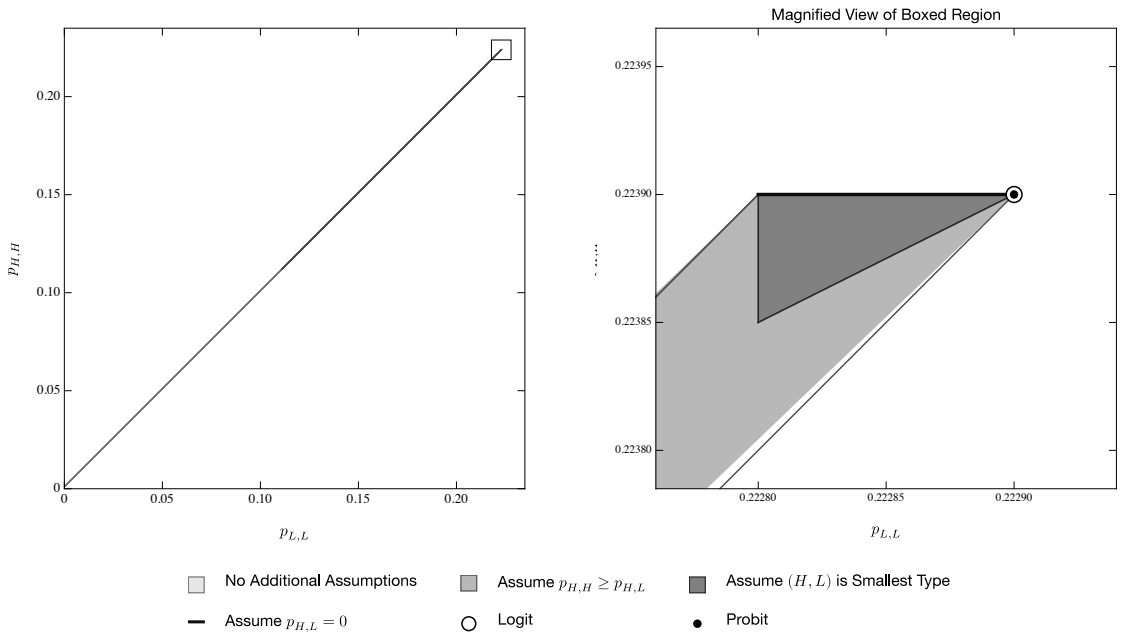
	$\mathbb{P}[D = H\|Z = 0]$	$\mathbb{P}[D = H\|Z = 1]$	$\mathbb{P}[D = L\|Z = 0]$	$\mathbb{P}[D = L\|Z = 1]$
Week 90	0.2239	0.2372	0.2361	0.2229
Week 135	0.2758	0.3037	0.2415	0.2414
Week 180	0.2941	0.3313	0.2462	0.2512
Week 208	0.3142	0.3559	0.2513	0.2509

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).

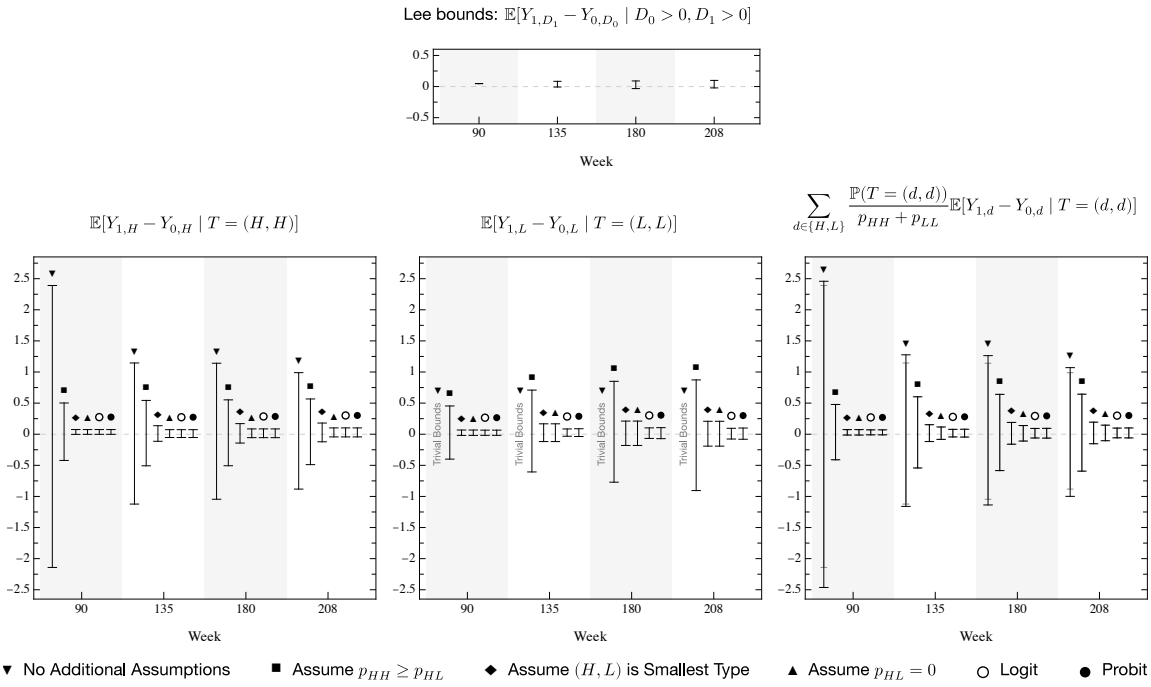


Notes: This figure plots the identified set for $(p_{L,L}, p_{H,H})$ under various assumptions, as indicated, for Week 90 of the Job Corps application. The right panel is a magnified view of the boxed region in the left panel. Lighter regions correspond to weaker restrictions and contain the darker regions. 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 $\mathbb{E}\left[W(1) - W(0) \,|\, D_0 = D_1 > 0\right]$. 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.

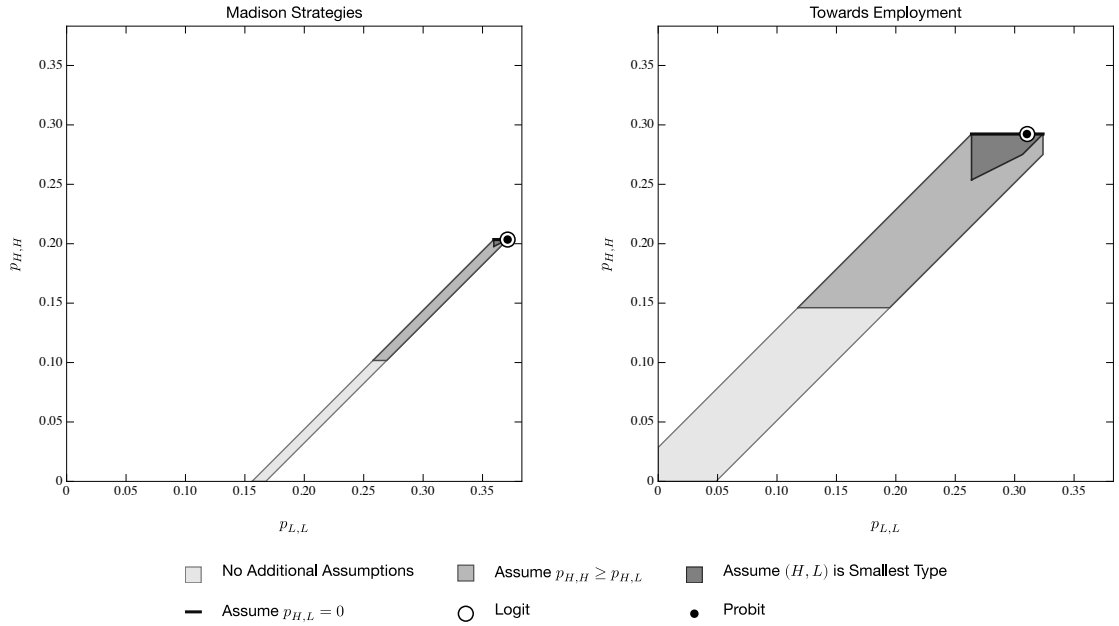
employment (0.62 to 0.69) and target-sector employment among the employed from 47 percent to 51 percent.

TABLE 2. Propensity scores for WorkAdvance RCTs

	$\mathbb{P}[D=H\|Z=0]$	$\mathbb{P}[D=H\|Z=1]$	$\mathbb{P}[D=L\|Z=0]$	$\mathbb{P}[D=L\|Z=1]$
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.

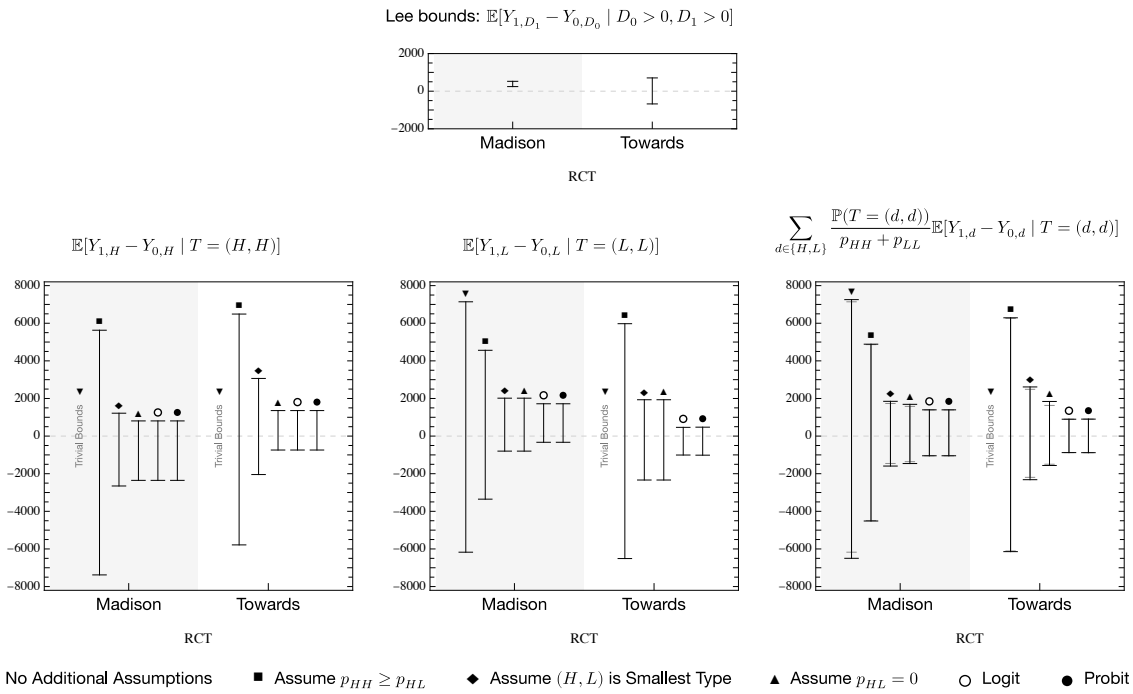


Notes: This figure plots the identified set for $(p_{L,L}, p_{H,H})$ under various assumptions, as indicated, for each of the WorkAdvance RCTs. Lighter regions correspond to weaker restrictions and contain the darker regions. Under the Logit and Probit assumptions, $(p_{L,L}, p_{H,H})$ is point-identified (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 $\mathbb{E}[W(1) - W(0) \,|\, D_0 = D_1 > 0]$. 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 $(((Y_{0,d}, Y_{1,d}) : d \in \{0,\ldots,K\}), D_0, D_1, Z)$ 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 $\mathcal{Y}_{z,d} \equiv \text{supp}\,(Y|D=d,Z=z)$ and, for each $d \in \{0,\ldots,K\}$, $z \in \{0,1\}$, and $(d',d'') \in \{0,\ldots,K\}^2$ define the CDF $F^{(z,d)}_{(d',d'')}$ as

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

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

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

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

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

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 $\mathbb{P}(S_1 \geq S_0) = 1$. Note that $p \equiv \frac{\mathbb{P}(D>0|Z=0)}{\mathbb{P}(D>0|Z=1)} = \frac{\mathbb{P}(S_0=1)}{\mathbb{P}(S_1=1)} = \mathbb{P}(S_0 = 1 \,|\, S_1 = 1)$, 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 $\mathbb{E}\left[\tilde{Y}_0 \,\middle|\, S_0 = S_1 = 1\right] = \mathbb{E}\left[\tilde{Y}_0 \,\middle|\, S_0 = 1\right] = \mathbb{E}[Y \mid D>0, Z=0]$. 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, $F_W(W) \sim \text{Uniform}[0,1]$, and so, $\mathbb{E}\left[F^{-1}_W(U) \,\middle|\, U \leq p\right] = \mathbb{E}\left[W \,\middle|\, W \leq F^{-1}_W(p)\right]$, 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 $F^{-1}_{Y|D>0,Z=1}(u) = \mathbf{1}\{u > \mathbb{P}(Y=0 \,|\, D>0, Z=1)\}$, and simplifying. □

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

Proof. First, notice that

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

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 $\underline{\gamma}^{1,r}_{d,d}$ and $\underline{\gamma}^{0,r}_{d,d}$ are jointly feasible. The result now follows by noting that $\underline{\gamma}^{1,r}_{d,d}$ and $\underline{\gamma}^{0,r}_{d,d}$ are jointly feasible if and only if $\mathbb{P}(D=d|Z=1)\underline{\gamma}^{1,r}_{d,d} = \mathbb{P}(D=d|Z=0)\underline{\gamma}^{0,r}_{d,d} = \underline{p}^{r}_{d,d}$ belongs to the identified set for $p_{d,d}$ which is true by definition of $\underline{p}^{r}_{d,d}$.

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

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

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 $D_z = d \neq 0$,

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

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

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

First, by taking expectations on all sides, this implies

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

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

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

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

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

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

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 $z \in \{0,1\}$,

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

for all $d \in \{0, \ldots, K\}$, where (here and in the remainder of the proof) we WLOG take $\mathbb{E}[Y_{z,0} \mid \boldsymbol{\epsilon}] = 0$ 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 $h_d(\boldsymbol{e}) \equiv -\max_{z \in \{0,1\}, d \in \{0,\ldots,K\}} \{\mathbb{E}[Y_{z,d} \mid \boldsymbol{\epsilon} = \boldsymbol{e}]\} -$

$M(\boldsymbol{e})$, 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

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

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 $((Y_{z,d} : d \in \{0, \ldots, K\}, z \in \{0,1\}), T, Z) \sim Q$ 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 $\mathbf{f}(y) \equiv \left(\mathbf{f}_{(z,d)|t}(y) : d \in \mathcal{D}_+, t \in \mathcal{D}^2, z \in \{0,1\}\right)$

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

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 $\left((Y_{z,d} : d \in \{0,\dots,K\}, z \in \{0,1\}), T, Z\right) \sim Q$, 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

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

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,

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

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*.