cemmap
centre for microdata methods and practice

Treatment effects with targeting instruments

Sokbae Lee
Bernard Salanié

The Institute for Fiscal Studies
Department of Economics, UCL

cemmap working paper CWP23/24

UKRI Economic and Social Research Council

Treatment Effects with Targeting Instruments*

Sokbae Lee† **Bernard Salanié**‡

December 5, 2024

Abstract

Multivalued treatments are commonplace in applications. We explore the use of discrete-valued instruments to control for selection bias in this setting. Our discussion revolves around the concept of targeting: which instruments target which treatments. It allows us to establish conditions under which counterfactual averages and treatment effects are point- or partially-identified for composite complier groups. We illustrate the usefulness of our framework by applying it to data from the Head Start Impact Study. Under a plausible positive selection assumption, we derive informative bounds that suggest less beneficial effects of Head Start expansions than the parametric estimates of Kline and Walters (2016).

KEYWORDS: Identification, selection, multivalued treatments, discrete instruments, monotonicity.

*This is a shorter, revised version of our "Filtered and Unfiltered Treatment Effects with Targeting Instruments" (first arXiv version July 2020). We thank Josh Angrist, Junlong Feng, Len Goff, and Jim Heckman for helpful comments. This work is in part supported by the European Research Council (ERC-2014-CoG-646917-ROMIA) and the UK Economic and Social Research Council for research grant (ES/P008909/1) to the CeMMAP.

†Department of Economics, Columbia University and Centre for Microdata Methods and Practice, Institute for Fiscal Studies, sl3841@columbia.edu.

‡Department of Economics, Columbia University and FGV EPGE, Rio de Janeiro, bsalanie@columbia.edu.

Introduction

Much of the literature on the evaluation of treatment effects has concentrated on the paradigmatic "binary/binary" example, in which both treatment and instrument only take two values. Multivalued treatments are common in actual policy implementations, however, as are multivalued instruments. Many different programs aim to help train job seekers for instance, and each of them has its own eligibility rules. Tax and benefit regimes distinguish many categories of taxpayers and eligible recipients. The choice of a college and major has many dimensions too, and responds to a variety of financial help programs and other incentives. Finally, more and more randomized experiments in economics resort to factorial designs[1].

Existing work on multivalued treatments under selection on observables includes Imbens (2000), Cattaneo (2010), and Ao, Calonico, and Lee (2021) among others. As the training, education choice, and tax-benefit examples illustrate, in non-experimental settings multivalued treatments are also subject to selection on unobservables. The use of instruments to evaluate the effects of multivalued treatments under selection on unobservables has received increasing attention in recent literature. In previous work (Lee and Salanié, 2018), we analyzed the case when enough continuous instruments are available. Identification is of course more difficult when instruments only take discrete values. We explore in this paper the use of such discrete-valued instruments in order to control for selection bias when evaluating discrete-valued treatments. Our goal is to find plausible conditions on treatment assignment and on the distribution of outcomes under which counterfactual averages and treatment effects are point- or partially identified for various (sometimes composite) complier groups. This distinguishes our paper from the recent contributions of Bai, Huang, Moon, Shaikh, and Vytlacil (2024), which focuses on population-wide average outcomes, and of Goff (2024b), which studies identification without any assumption on outcomes.

In the binary/binary model, the analyst can often take for granted that switching on the binary instrument makes treatment (weakly) more likely for all or no observations. This is satisfied under the local average treatment effect (LATE)-monotonicity assumption (e.g., Imbens and Angrist, 1994; Vytlacil, 2002; Heckman and Vytlacil, 2007a). With multiple instrument values and multiple treatments, there may be no natural ordering of instrument or treatment values that would give meaning to the word "monotonicity". Since Heckman and Pinto (2018) defined an "unordered monotonicity" property, various papers have proposed other definitions of (qualified) monotonicity[2].

Even when some sort of monotonicity holds, there exist several groups of compliers—

[1] Muralidharan, Romero, and Wüthrich (2023) review recent applications of factorial designs.

[2] See Navjeevan and Pinto (2022) for a detailed analysis of some of these proposals.

individuals whose treatment assignment changes with the value of the instrument. The multiplicity of treatments and instruments may give rise to a bewildering number of cases, as existing literature demonstrates. Angrist and Imbens (1995) analyzed two-stage least squares (TSLS) estimation when the treatment takes a finite number of ordered values. Closer to us, Heckman, Urzua, and Vytlacil (2006); Heckman and Vytlacil (2007b); Heckman, Urzua, and Vytlacil (2008) discussed the identification of treatment effects in the presence of discrete-valued instruments when assignment to treatment can be modeled as a discrete choice model. Several recent papers have studied the case of binary treatments with multiple instruments. Mogstad, Torgovitsky, and Walters (2021) and Goff (2024a) analyzed the identifying power of different monotonicity assumptions in this context[3]. Others have studied models with binary instruments and multivalued or continuous treatments. Torgovitsky (2015), D'Haultfoeuille and Février (2015), Huang, Khalil, and Yildiz (2019), Caetano and Escanciano (2021), and Feng (2024) developed identification results for different models.

In a wide-ranging contribution, Heckman and Pinto (2018) derived results on partial identification in discrete-instrument, discrete-treatment models; they also showed how additional identifying assumptions, such as unordered monotonicity, can be applied to shrink the identified set of treatment effects for various complier groups. While their results are very general, they are not as transparent as one would like. Our approach to this issue is different: we seek a parsimonious framework within which we can make constructive progress, and that can still be useful in many applications. In order to reduce the complexity of the problem, we start by imposing an additive random-utility model (ARUM) structure. Under ARUM, the selection into treatment depends on mean values and additive, observation-specific shocks. Some, but not all, ARUM models satisfy the unordered monotonicity property of Heckman and Pinto (2018), which was applied by Pinto (2021) to the Moving to Opportunity program. In many applications, some observations are not treated; in others, another treatment value is particularly salient. We call it the "control". Under ARUM, each treatment t generates a change in the mean value, relative to the control, that depends on the value z of the instrument. It is natural to speak of an instrument value z *targeting* a treatment value t when it maximizes this change in mean value. Most of our paper relies on the assumption of *strict targeting*, which obtains when each instrument only changes the mean values of the treatments it targets. Strict targeting holds for instance in models of imperfect compliance when the cost of non-compliance does not depend on its nature. Some of our results also require *one-to-one targeting*, where each non-zero instrument targets one treatment only,

[3]Mogstad, Torgovitsky, and Walters (2020) further apply their framework of monotonicity with multiple instruments to marginal treatment effects (e.g., Heckman and Vytlacil, 2001, 2005; Carneiro, Heckman, and Vytlacil, 2011).

and each treatment (apart from the control) is targeted by one instrument only.

Our use of "targeting" instruments is similar in spirit to Section 7.3 of Heckman and Vytlacil (2007b)[4]. We define it differently and we seek to identify a more general class of treatment effects. The term "targeting" is inspired by the time-honored Targeting Principle[5]. Some policies act directly on final outcomes, and others aim to modify choices. Our use of the term "targeting" refers to the latter. Take a Roy model in which workers choose among occupations on the basis of their net utilities; we observe the choice of occupation and the wage in that occupation. A safety regulation that reduces the disutility of labor for (say) construction workers is, in our terminology, an instrument that targets the choice to be a construction worker. Policymakers might also seek to increase average incomes by offering a college credit. While their final aim is to increase wages (an outcome), we would say that the college credit is an instrument that targets the choice to go to college—a treatment variable.

To illustrate, consider a typical randomized experiment with imperfect compliance: (i) individuals are randomly assigned to a treatment branch t (including a non-treatment option 0) based on the instrument value z that they draw; (ii) some individuals self-select into a treatment branch t' that they prefer, even though they did not draw the corresponding instrument value z'. In our terminology, z targets t and z' targets t'. Often this mapping is one-to-one; this is what our one-to-one targeting assumption states. Strict targeting of a treatment branch t is more restrictive, as its name indicates. One way to interpret it in a non-compliance context is that it is equally difficult for an average individual in the population to select into treatment t when she was not assigned to that branch, no matter what the experimenter's intended treatment branch was. To cite two examples, consider the interventions reported in Angrist, Lang, and Oreopoulos (2009) and in Attanasio, Fernández, Fitzsimons, Grantham-McGregor, Meghir, and Rubio-Codina (2014; 2020). These are 4-way factorial randomized experiments: each subject is randomly assigned to a control group, to receive treatment 1, to receive treatment 2, or to receive both treatments. By definition, this is one-to-one targeting. Compliance was very imperfect in Angrist, Lang, and Oreopoulos (2009), and it is described as "high" in the other two papers. If subjects self-selected into treatments on the basis of their expected benefits, then strict targeting is a natural assumption.

Combining ARUM and assumptions on targeting allows us to point-identify the size of some complier groups and the corresponding counterfactual averages and treatment effects on any function of the outcomes, and to partially identify others. We use two examples to

[4]See also the recent contribution by Buchinsky, Gertler, and Pinto (2023), which uses revealed preference arguments.

[5]Early references include Tinbergen (1952) and Bhagwati (1971).

demonstrate the identification power and implications of ARUM and targeting. Our first example is a $2 \times T$ model where a binary instrument targets only one of $T \geqslant 3$ treatment values, as in Kline and Walters (2016). In our second example, three unordered treatment values target three instrument values. This 3×3 model was also studied by Kirkeboen, Leuven, and Mogstad (2016)[6]. Unlike them, we do not assume that the data contains information on next-best alternatives. Whereas the $2 \times T$ model satisfies unordered monotonicity under our strongest targeting assumptions, the 3×3 model does not[7].

We obtain novel identification results for both examples; they lead to new estimands or bounds for average treatment effects on various groups. Additional identifying assumptions can refine these bounds. One example is what we call *positive selection*. This assumes that the average outcome for a given treatment t is larger for some response group than for another. Consider for instance the binary instrument case. It seems natural to assume that the always-takers of a treatment get more from it than compliers who only take it if they are incentivized to do so. Positive selection also obtains under weak assumptions in the generalized Roy model. More generally, let us return to our earlier illustration of a randomized experiment under imperfect compliance. Consider the response group of individuals who would end up in treatment t' both when drawing z and when drawing z'. We would expect this response group to have better outcomes under t', on average, than the response group that exhibits perfect compliance to z and z' draws—assuming that these two response groups end up in the same treatment branches for all other instrument values. This falls exactly under our positive selection assumption. It adds identifying power in both of our leading examples.

To illustrate the usefulness of our framework, we apply it to the Head Start Impact Study (HSIS), a randomized experiment that sought to evaluate the value added of Head Start preschools. Kline and Walters (2016) revisited the HSIS; they took into account the presence of a substitute treatment (alternative preschools in this case). They found that Head Start was only beneficial for children who would not have attended an other preschool program instead. We confirm the importance of taking into consideration alternative preschools when evaluating Head Start. Unlike Kline and Walters (2016), we do not rely on parametric selection models. Under a plausible positive selection assumption, our estimates suggest that the large difference between complier groups that they find can only be rationalized under *negative* selection into Head Start. As a by-product, we provide an upper bound on the welfare effect of expanding access to Head Start. Interestingly, the estimated upper

[6]See also more recent work by Bhuller and Sigstad (2024), Heinesen, Hvid, Kirkeboen, Leuven, and Mogstad (2022), and Nibbering, Oosterveen, and Silva (2022).

[7]It does satisfy the weaker generalized monotonicity assumption of Bai, Huang, Moon, Shaikh, and Vytlacil (2024), however.

bound turns out to be lower than the point estimate of Kline and Walters (2016); and it yields a lower marginal value for public funds used in expanding access to Head Start.

The paper is organized as follows. Section 1 defines our framework. In Section 2, we define and discuss the concepts of targeting, one-to-one targeting, and strict targeting. Section 3 derives their implications for the identification of population shares, counterfactual averages, and the effects of the treatments on various complier groups; it also defines and illustrates positive selection. Finally, we present estimation results for Head Start in Section 4. The Appendices contain the proofs of all propositions and lemmata, along with some additional material.

1 The Framework

In all of the paper, we denote observations as $i = 1, \ldots, n$. Each observation consists of covariates X_i, instruments Z_i, outcome variables Y_i, and treatments T_i. We assume that the covariates X_i are exogenous to treatment assignment and outcomes. Since they will not play any role in our identification strategy, we condition on the covariates throughout and we omit them from the notation. Our results should therefore be interpreted as conditional on X.

We assume that observations are independent and identically distributed. Random sampling rules out that the treatment status of one observation influences other observations. This further implies that the outcome for a specific observation does not impact the outcomes of other members within the population. In other words, we rely on the Stable Unit Treatment Value Assumption (SUTVA).

We focus in this paper on treatment variables that take discrete values, which we label $t \in \mathcal{T}$. For simplicity, we will call $T = t$ "treatment t". These values do not have to be ordered; e.g., when $t = 2$ is available, it does not necessarily indicate "more treatment" than $t = 1$. We assume that the only available instruments are discrete-valued, and we label their values as $z \in \mathcal{Z}$.

We will use the standard counterfactual notation: $T_i(z)$ and $Y_i(t, z)$ denote respectively potential treatments and outcomes. $\mathbf{1}(A)$ denotes the indicator of set A.

The validity of the instruments requires the usual exclusion and independence restrictions:

Assumption 1 (Valid Instruments)**.** *(i)* $Y_i(t, z) = Y_i(t)$ *for all* (t, z) *in* $\mathcal{T} \times \mathcal{Z}$.

(ii) $Y_i(t)$ *and* $T_i(z)$ *are independent of* Z_i *for all* (t, z) *in* $\mathcal{T} \times \mathcal{Z}$.

Under Assumption 1, we define $T_i := T_i(Z_i)$ and $Y_i := Y_i(T_i)$.

Throughout the paper, we assume that we observe (Y_i, T_i, Z_i) for each i.

1.1 Restricting Heterogeneity

As in most of this literature, we will need an assumption that restricts the heterogeneity in the counterfactual mappings $T_i(z)$. In the binary/binary model, this is most often done by imposing LATE-monotonicity. As is well-known, LATE-monotonicity imposes that (denoting instrument values as $z = 0, 1$) (i) or (ii) must hold:

(i) for each observation i, $T_i(1) \geqslant T_i(0)$;

(ii) for each observation i, $T_i(0) \geqslant T_i(1)$.

With more than two treatment values and/or more than two instrument values, there are many ways to restrict the heterogeneity in treatment assignment. Since treatments may not be ordered in any meaningful way, we cannot apply the results in Angrist and Imbens (1995) for instance. Mogstad, Torgovitsky, and Walters (2021) state several versions of monotonicity for a binary treatment model with $|\mathcal{Z}| > 2$. They propose a "partial monotonicity" assumption which applies binary LATE-monotonicity component by component. This requires that the instruments be interpretable as combinations of component instruments, which is not necessarily the case here.

To cut through this complexity, we assume from now on that assignment to treatment can be represented by an Additive Random-Utility Model (ARUM), that is by a discrete choice problem with additively separable errors:

$$T_i(z) = \arg\max_{t \in \mathcal{T}} (U_z(t) + u_{it})$$

for some real numbers $U_z(t)$ which are common across observations, and random vectors $(u_{it})_{t \in \mathcal{T}}$ that are distributed independently of Z_i. We do not restrict the codependence of the random variables u_{it}. The usual models of multinomial choice belong to this family. ARUM also includes ordered treatments, for which $u_{it} \equiv \sigma(t)u_i$ for some increasing positive function σ.

In a randomized experiment with perfect compliance, we would have $U_z(t') = -\infty$ and $U_{z'}(t) = -\infty$. With imperfect compliance, these mean values are finite; if for instance $u_{it'} - u_{it} > U_z(t) - U_z(t')$, individual i will get into treatment t' when drawing z would normally assign her to t.

Imposing an ARUM structure will greatly simplify our discussion of treatment assignment. It incorporates a substantial restriction, however. Suppose that observation i has

treatment values t under z and t' under z'. By the ARUM structure, this implies

$$U_z(t) + u_{it} \geqslant U_z(t') + u_{it'}$$
$$U_{z'}(t') + u_{it'} \geqslant U_{z'}(t) + u_{it}.$$

Combining these two restrictions implies an "increasing differences" property:

$$U_{z'}(t') - U_{z'}(t) \geqslant U_z(t') - U_z(t).$$

This inequality in turn is incompatible with the existence of an observation j that has treatment values t' under z and t under z'. Thus we rule out "direct two-way flows": if a change in the value of an instrument causes an observation to shift from a treatment value t to a treatment value t', it can cause no other observation to switch from t' to t. The argument above is a special case of the general discussion in Heckman and Pinto (2018); their Theorem T-3 shows that the treatment assignment models that satisfy unordered monotonicity for each pair of instrument values can be represented as an ARUM. Not all ARUM models satisfy unordered monotonicity, however; unordered monotonicity excludes a more general class of two-way flows. We will illustrate this point on one of our leading examples in Section 3.3.

1.2 Assignment to Treatment

Assumption 2 defines the class of models of assignment to treatment that we analyze in this paper.

Assumption 2 (ARUM)**.** *The treatment assignment model consists of:*

1. *a finite set* $\mathcal{T} = \{0, 1, \ldots, |\mathcal{T}| - 1\}$*;*
2. *a finite set of instrument values* $\mathcal{Z} = \{0, 1, \ldots, |\mathcal{Z}| - 1\}$*;*
3. *an ARUM model of treatment:*

$$T_i(z) = \arg\max_{t \in \mathcal{T}} (U_z(t) + u_{it}),$$

where the vector $(u_{it})_{t \in \mathcal{T}}$ *is distributed independently of* Z_i *and has an absolutely continuous distribution with full support on* $\mathbb{R}^{|\mathcal{T}|}$*.*

We will often refer to the $U_z(t)$ as "mean values". This is only meant to simplify the exposition; it is consistent with, but need not refer to, preferences on the part of the agent.

Note that when $\mathcal{T} = \{0, 1\}$, Assumption 2 is just the standard monotonicity assumption, with a threshold-crossing rule

$$T_i(z) = \mathbb{1}(u_{i0} - u_{i1} \leqslant U_z(1) - U_z(0)).$$

If we add a third treatment value so that $\mathcal{T} = \{0, 1, 2\}$, the ARUM assumption starts to bite as it excludes direct two-way flows in the treatment model. However, the combination of Assumptions 1 and 2 is far from sufficient to identify interesting treatment effects in general. In order to better understand what is needed, we now resort to the notion of *response-groups* of observations, whose members share the same mapping from instruments z to treatments t. We first state a general definition[8].

Definition 1 (Response-vectors and Response-groups)**.** Let R be an element of the Cartesian product $\mathcal{T}^{\mathcal{Z}}$ and $R(z) \in \mathcal{T}$ denote its component for instrument value $z \in \mathcal{Z}$.

- Observation i has (elemental) *response-vector* R if and only if for all $z \in \mathcal{Z}$, $T_i(z) = R(z)$. The set C_R denotes the set of observations with response-vector R and we call it a *response-group*.
- We extend the definition in the natural way to incompletely specified mappings, where each $R(z)$ is a subset of $\mathcal{T}$. We call the corresponding response-vectors and response-groups *composite*. In particular, if $R(z) = \mathcal{T}$ we denote it by an asterisk in the corresponding position.

To illustrate, consider the binary instrument/binary treatment case. It has a priori $2^2 = 4$ response vectors, $R \in \{00, 01, 10, 11\}$ with corresponding response-groups $C_{00}, C_{01}, C_{10}, C_{11}$. In this notation, the first number refers to a treatment value with $z = 0$ and the second number with $z = 1$. For instance, C_{01} refers to those with $T_i(0) = 0$ and $T_i(1) = 1$, while the composite response-group C_{*1}, for which $R(0) = \{0, 1\}$, represents the union of C_{01} and C_{11} The LATE-monotonicity assumption implies that either C_{01} or C_{10} is empty.

2 Targeting

We start by introducing additional assumptions on the underlying treatment model. We will illustrate these assumptions on three examples: the "binary instrument model" or the "$2 \times T$" model; the "3×3 model"; and a generalized Roy model. We first define them briefly.

[8]This is analogous to the definitions in Heckman and Pinto (2018).

Example 1 (The binary instrument $(2 \times T)$ model)**.** $\mathcal{T} = \{0, 1, \ldots, T-1\}$ and $\mathcal{Z} = \{0, 1\}$. This could for instance represent an intent-to-treat model, where agents in the control group $Z = 0$ are not treated $(T = 0)$ and agents with $Z = 1$ self-select the type of the treatment $T \geqslant 1$ or opt out altogether $(T = 0)$. $\square$

When $|\mathcal{T}| = 3$, treatment assignment can be represented in the $(u_{i1} - u_{i0}, u_{i2} - u_{i0})$ plane. The points of coordinates $P_z = (U_z(0) - U_z(1), U_z(0) - U_z(2))$ play an important role as for a given z,

- $T_i(z) = 0$ to the south-west of P_z;
- $T_i(z) = 1$ to the right of P_z and below the diagonal that goes through it;
- $T_i(z) = 2$ above P_z and above the diagonal that goes through it.

Treatment assignment is illustrated in Figure 1 for a given z, where the origin is in P_z. We will make recurrent use of this type of figure.

Example 2 $(3 \times 3$ model)**.** Assume that $\mathcal{Z} = \{0, 1, 2\}$ and $\mathcal{T} = \{0, 1, 2\}$. As a leading example, Kirkeboen, Leuven, and Mogstad (2016) investigate the 3×3 model in order to analyze the effect of students' choice of field of study on their earnings; each instrument value shifts the eligibility of a student for a given field. We will return to this application in Section 3.4.

Finally, our framework also includes multivalued generalized Roy models (see Eisenhauer, Heckman, and Vytlacil (2015)).

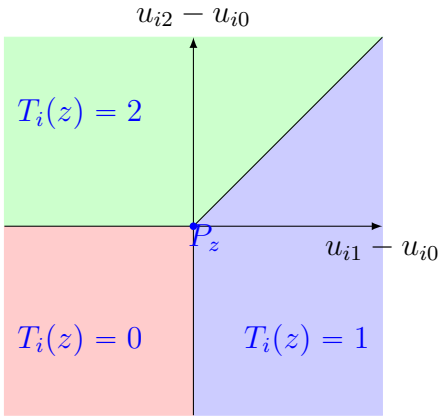
Example 3 (A Generalized Roy Model)**.** Suppose that agents choose occupations $t = 0, \ldots, |\mathcal{T}| - 1$ on the basis of their expected wages $w_i(t) = \bar{w}(t) + \eta_{it}$, net of labor disutilities that depend on the values of the instruments:

$$T_i(z) = \arg \max_{t=0,\ldots,|\mathcal{T}|-1} (w_i(t) - d_i(z, t))$$

where $d_i(z, t) = \bar{d}_z(t) + v_{it}$. Potential wages are $Y_i(t) = w_i(t) + \varepsilon_{it}$. We observe Z_i, the chosen occupation $T_i = T_i(Z_i)$, and realized wages $Y_i = Y_i(T_i)$. If the vector of variables $\{(\eta_{it}, v_{it})\}_{t=0}^{|\mathcal{T}|-1}$ is independent of Z_i, this is an ARUM model with $U_z(t) = \bar{w}(t) - \bar{d}_z(t)$ and $u_{it} = \eta_{it} - v_{it}$.

"Targeting" will be the common thread in our analysis. Just as in general economic discussions a policy measure may target a particular outcome, we will speak of instruments (in the econometric sense) targeting the assignment to a particular treatment.

Figure 1: Treatment assignment for $|T| = 3$ for given z



Under Assumption 2, assignment to treatment is governed by the differences in mean values $(U_z(t) - U_z(\tau))$ and by the differences in unobservables $u_{it} - u_{i\tau}$. Only the former depend on the instrument. From now on, we assume that there is a reference treatment value t_0 whose mean utility does not depend does not depend on the value of the instrument:

Assumption 3 (Reference Treatment)**.** *There exists $t_0 \in \mathcal{T}$ such that $z \in \mathcal{Z} \to U_z(t_0)$ is constant. Without loss of generality, we renumber treatment values so that $t_0 = 0$; and we normalize utilities with $U_z(0) = 0$ for all $z \in \mathcal{Z}$.*

In many applications, $t = 0$ is a "no-treatment" value, and instruments only change the mean utilities of the other treatments. For instance, tuition subsidies, investment credits, and invitations to training programs have no effect for those who do not attend college, do not invest, or choose not to train. Assumption 3 seems natural in such cases[9]. For a counter-example, consider a program of unconditional cash transfers with different values z, for which we observe the purchases t of several categories of goods the following month. If a household decides to save the transfer $(t = 0)$, its mean (discounted) utility will still depend on the value of the transfer z that it received[10].

Given Assumption 3, we will say that an instrument value z *targets* a treatment value t if it maximizes the mean utility $U_z(t) - U_z(0) = U_z(t)$.

Definition 2 (Targeted Treatments and Targeting Instruments)**.** For any $z \in \mathcal{Z}$ and $t \in \mathcal{T}$, let

$$\bar{U}(t) \equiv \max_{z \in \mathcal{Z}} U_z(t) \quad \text{and} \quad Z^*(t) \equiv \arg\max_{z \in \mathcal{Z}} U_z(t)$$

[9]In the generalized Roy model (Example 3), it holds if the disutility of occupation 0 does not depend on the values of the instruments.

[10]In that case one could define targeting with the function $\tilde{U}_z(t) = U_z(t) - U_z(0)$. We have not explored the consequences of this alternative definition.

denote the maximum value of $U_z(t)$ over $z \in \mathcal{Z}$ and the set of maximizers, respectively. If $Z^*(t)$ is not all of $\mathcal{Z}$, then we will say that the instrument values $z \in Z^*(t)$ *target* treatment value t; and we write $t \in T^*(z)$. We denote by $\mathcal{T}^*$ the set of targeted treatments and $\mathcal{Z}^* = \bigcup_{t \in \mathcal{T}^*} Z^*(t)$ the set of targeting instruments.

Definition 2 calls for several remarks. First, Assumption 3 implies that $Z^*(0) = \mathcal{Z}$. Therefore $t = 0$ is not in $\mathcal{T}^*$; the set $\mathcal{T}^*$ may exclude other treatment values, however. If a treatment value t is not targeted ($t \notin \mathcal{T}^*$), by definition the function $z \to U_z(t)$ is constant over $z \in \mathcal{Z}$, with value $\bar{U}(t)$. If an instrument value z does not target any treatment ($z \notin \mathcal{Z}^*$), then $U_z(t) < \bar{U}(t)$ for every $t \in \mathcal{T}^*$. While non-targeted treatment values ($t \in \mathcal{T} \setminus \mathcal{T}^*$) have mean values that do not respond to changes in the instruments, these mean values may and in general will differ across treatments. The probability that an individual observation takes a treatment $t \in \mathcal{T} \setminus \mathcal{T}^*$ also generally depends on the value of the instrument.

It is important to note here that the values $U_z(t)$ and therefore the targeting maps Z^* and T^* are not observable; any assumption on targeting instruments and targeted treatments must be a priori and context-dependent. As we will see, these prior assumptions sometimes have consequences that can be tested.

Now suppose that each z consists of a set of (possibly zero or negative) subsidies $S_z(t)$ for treatments $t \in \mathcal{T}$. If there is a no-subsidy regime $z = 0$ with $S_0(t) = 0$ for all t, it seems natural to write the mean value as $U_z(t) = U_0(t) + S_z(t)$. Then for any treatment t, the set $Z^*(t)$ consists of the instrument values z that subsidize t most heavily. As this illustration suggests, the sets $Z^*(t)$ may not be singletons, and they may well intersect. We now introduce a more restrictive definition that rules out these two possibilities.

Definition 3 (One-to-one targeting)**.** Targeting is *one-to-one* when both $Z^* : \mathcal{T}^* \to \mathcal{Z}^*$ and $T^* : \mathcal{Z}^* \to \mathcal{T}^*$ are functions.

Under one-to-one targeting, we will often write "$z = t$" if z targets t; this is without loss of generality. Let us illustrate these varieties of targeting on Example 2.

Table 1: Values of $U_z(t)$ in the 3×3 model

	$t = 0$	$t = 1$	$t = 2$
$z = 0$	0	a	d
$z = 1$	0	b	e
$z = 2$	0	c	f

Example 2 continued. Table 1 shows the values of $U_z(t)$ in the 3×3 model of Example 2. Suppose that $t = 1$ is targeted; choose some z that targets it and relabel it as $z = 1$. This

means that

$$b \geqslant \max(a, c) \text{ and } b > \min(a, c).$$

If $t = 2$ is also targeted by some $z \neq 1$, we relabel this instrument value as $z = 2$. This gives

$$f \geqslant \max(d, e) \text{ and } f > \min(d, e).$$

Finally, if targeting is one-to-one we have $b > \max(a, c)$ and $f > \max(d, e)$.

2.1 Consequences of One-to-One Targeting

In this subsection, we impose

Assumption 4 (One-to-one Targeting)**.** *Targeting is one-to-one.*

Remember that under Assumption 4, we can relabel instrument values so that if t is targeted, then it is targeted by $z = t$. Moreover, $t^*(z)$ must equal z.

This implies some useful restrictions on response-groups.

Proposition 1 (Response-groups under one-to-one targeting)**.** *Under Assumptions 1, 2, and 4, take a targeted treatment $t \in \mathcal{T}^*$.*

(i) If an observation i has $T_i(t) = 0$, then it never receives treatment t: $T_i(z) \neq t$ for all $z \in \mathcal{Z}$.

(ii) As a consequence, all response-groups C_R with $R(t) = 0$ and $R(z) = t$ for some $z \neq t$ are empty.

Example 2 (continued) Return to the 3×3 model and to Table 1. Suppose that both $t = 1$ and $t = 2$ are targeted. Under the conditions of Proposition 1, we have $b > \max(a, c)$ and $f > \max(d, e)$.

Since the points P_z have coordinates $(-U_z(1), -U_z(2))$,

- $P_1 = (-b, -e)$ must lie to the left of $P_0 = (-a, -d)$ and of $P_2 = (-c, -f)$,
- P_2 must lie below P_0 and P_1.

This is easily rephrased in terms of the response-vectors of definition 1. First note that in the 3×3 case, there are a priori $3^3 = 27$ response-vectors, $R = 000$ to $R = 222$, with corresponding response-groups C_{000} to C_{222}. Groups C_{ddd} are "always-takers"[11] of treatment

[11] Observations in group C_{000} are usually called the "never-takers". We prefer not to break the symmetry in our notation. We hope this will not cause confusion.

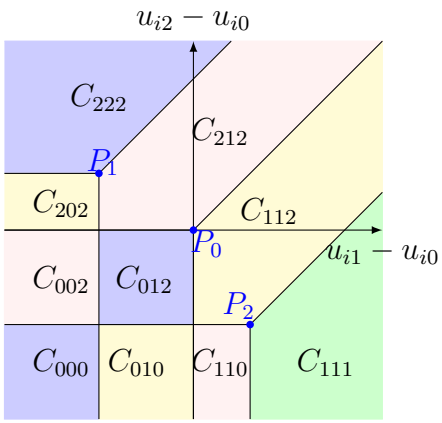
value d. All other groups are "compliers" of some kind, in that their treatment changes under some changes in the instrument. We will also pay special attention to some non-elemental groups. For instance, C_{0*2} will denote the group who is assigned treatment 0 under $z = 0$ and treatment 2 under $z = 2$, and any treatment under $z = 1$. That is,

$$C_{0*2} = C_{002} \bigcup C_{012} \bigcup C_{022}.$$

Assumptions 2 and 4 together imply the emptiness of four composite groups out of the 27 possible. For any treatment value τ, Proposition 1(ii) rules out group $C_{10\tau}$ since this group has $R(1) = 0$ and $R(0) = 1$. It rules out $C_{\tau 01}$ as $R(1) = 0$ and $R(2) = 1$. This eliminates the composite groups C_{10*} and C_{*01}. The same argument applies to composite groups C_{*20} and C_{2*0}, which have $R(2) = 0$ and $R(1) = 2$ or $R(2) = 2$.

These four composite groups correspond to 10 elemental groups[12]. This still leaves us with 17 elemental groups, and potentially complex assignment patterns. Consider for instance Figure 2. It shows one possible configuration for the 3×3 model; the positions for P_0, P_1 and P_2 are consistent with Assumptions 2 and 4.

Figure 2: A 3×3 example



The number of distinct response-groups (ten in this case) and the contorted shape of the C_{212} and C_{112} groups in Figure 2 point to the difficulties we face in identifying response-groups without further assumptions. Moreover, this is only one possible configuration: other cases exist, which would bring up other response-groups.

Heckman and Pinto (2018, pp. 16–20), Pinto (2021), and Kirkeboen, Leuven, and Mogstad (2016) also studied the 3×3 model; they proposed sets of assumptions that identify some treatment effects. The example in Heckman and Pinto (2018, pp. 16–20) is rather specific.

[12]Specifically, they are: $C_{100}, C_{101}, C_{102}, C_{001}, C_{201}, C_{020}, C_{120}, C_{220}, C_{200}$, and C_{210}.

We show in Appendix E how to apply our framework to the Moving to Opportunity experiment studied in Pinto (2021). The setup in Kirkeboen, Leuven, and Mogstad (2016) is most similar to ours; we will return to the differences between our approach and theirs in Section 3.3. □

2.2 Strict Targeting

Figure 2 suggests that if we could make sure that P_1 is directly to the left of P_0, the shape of C_{212} would become nicer—and group C_{202} would be empty. Bringing P_2 directly under P_0 would have a similar effect. This translates directly into assumptions on the dependence of the $U_z(t)$ on the instruments: the first one imposes $d = e$ and the second one imposes $a = c$. This can be interpreted as policy regime $z = 1$ (resp. $z = 2$) subsidizing treatment $t = 1$ (resp. $z = 2$) only. To return to the general model, there are applications in which the instruments $z \in Z^*(t)$, which maximize $U_z(t)$, do not shift assignment between the other values of the treatment. The following definition is a direct extension of this discussion.

Definition 4 (Strict Targeting of Treatment t)**.** Take any targeted treatment value $t \in \mathcal{T}^*$. It is strictly targeted if the function $z \in \mathcal{Z} \to U_z(t)$ takes the same value for all instruments that do not target t (the values $z \notin Z^*(t)$). We denote this common value by $\underline{U}(t)$, and we will say of the instrument values $z \in Z^*(t)$ that they strictly target t.

Suppose for instance the data comes from a randomized experiment, where the instrument value $z = t$ targets treatment t. If compliance is imperfect, an individual will trade off the benefits from switching to a treatment $t' \neq t$ with the costs of the effort required. Strict targeting obtains when the cost of switching to t' do not depend on the value of t.

Under strict targeting, turning on instrument $z \in Z^*(t)$ promotes treatment t without affecting the mean values $U_z(t')$ of other treatment values t'. This explains our use of the term "strict targeting". In this ARUM specification, an instrument in $Z^*(t)$ plays the same role as a price discount on good t in a model of demand for goods whose mean values only depend on their own prices. In the language of program subsidies, all $z \in Z^*(t)$ subsidize t at the same high rate, and all other instrument values offer the same, lower subsidy.

Note that strict targeting only bites if $\mathcal{Z}$ contains at least three instrument values. If $|\mathcal{Z}| = 2$ (one binary instrument, as in our Example 1) and say $z = 1$ targets t, then $\mathcal{Z} \backslash Z^*(t)$ can only consist of $z = 0$ and Assumption 5 trivially holds.

Finally, we should emphasize that one-to-one targeting and strict targeting are logically independent assumptions: neither one implies the other. Consider the 3×3 model of Example 2 under one-to-one targeting; strict targeting only holds for $t = 1$ if $a = c$, and for $t = 2$

if $d = e$. On the other hand, the 3×3 model with $b > a = c$ and $e > d = f$ satisfies strict targeting but not one-to-one targeting, as $z = 1$ targets both $t = 1$ and $t = 2$.

2.3 Consequences of Strict Targeting

Now consider the general model. If a treatment t is strictly targeted, then $U_z(t)$ can only take one of two values: $\bar{U}(t)$ if z targets t, and $\underline{U}(t)$ otherwise. By definition, if t is not targeted then the value of $U_z(t)$ does not depend on z; we also denote it $\underline{U}(t)$. We will assume in this subsection that all targeted treatments are strictly targeted:

Assumption 5 (Strict targeting)**.** *If t is in $\mathcal{T}^*$, then t is strictly targeted.*

Under strict targeting, the values of $U_z(t)$ are given in Table 2.

Table 2: Values of $U_z(t)$ under strict targeting

	$t \in T^*(z)$	$t \notin T^*(z)$
$z \in Z^*$	$\bar{U}(t)$	$\underline{U}(t)$
$z \notin Z^*$	$\underline{U}(t)$	

Consider an observation i under strict targeting. If it is assigned an instrument value z, it can end up with one of the treatment values t that z targets (if any), with a value $\bar{U}(t) + u_{it}$ in the ARUM. Alternatively, if its treatment is some t' that z does not target, then the ARUM value will be $\underline{U}(t') + u_{it'}$. This motivates the following definition.

Definition 5 (Top targeted and top alternative treatments)**.** Take any observation i in the population.

(i) For any targeting instrument $z \in \mathcal{Z}^*$, let

$$V_i^*(z) = \max_{t \in T^*(z)} (\bar{U}(t) + u_{it})$$

and $T_i^*(z) \subset T^*(z)$ denote the set of maximizers. We call the elements of $T_i^*(z)$ the *top targeted treatments* for observation i under instrument value z.

(ii) Also define

$$\underline{V}_i = \max_{t \in \mathcal{T}} (\underline{U}(t) + u_{it})$$

and let $\underline{T}_i \subset \mathcal{T}$ denote the set of maximizers. We call the elements of $\underline{T}_i$ the *top alternative treatments* for observation i.

(iii) The sets $T_i^*(z)$ and $\underline{T}_i$ are singletons[13] with probability 1; we let $t_i^*(z)$ and $\underline{t}_i$ denote the top targeted treatment and the top alternative treatment.

The term "top alternative treatment" may read like a misnomer since the maximization runs over all treatment values. The following result justifies it; more importantly, it shows that strict targeting imposes a lot of structure on the mapping from instruments to treatments.

Proposition 2 (Response groups under strict targeting)**.** *Let Assumptions 1, 2, 3, and 5 hold. Let i be any observation in the population. For any instrument value z, $T_i(z)$ is either the top targeted treatment or the top alternative treatment. If z is not a targeting instrument, $T_i(z)$ can only be the top alternative treatment. That is:*

(i) if $z \in \mathcal{Z}^*$, then $T_i(z)$ is $t_i^*(z)$ if $V_i^*(z) > \underline{V}_i$; if $V_i^*(z) < \underline{V}_i$, then $T_i(z) = \underline{t}_i$ and $\underline{t}_i$ is not targeted by z.

(ii) if $z \notin \mathcal{Z}^*$, then $T_i(z)$ is $\underline{t}_i$.

Note that in a sense, all instrument values in $\mathcal{Z} \setminus \mathcal{Z}^*$ are equivalent under strict targeting. If z and z' are both in $\mathcal{Z}^*$, then the functions U_z and $U_{z'}$ coincide on all of $\mathcal{T}$ and the counterfactual treatments $T_i(z)$ and $T_i(z')$ must be in $\underline{T}_i$ for any observation i.

In the 2×2 model, we have $\bar{U}(1) = U_1(1)$ and $\underline{U}(0) = \underline{U}(1) = 0$. A complier is an observation $i \in C_{01}$; it is in treatment arm $t = 0$ when $z = 0$ and in $t^*(1) = 1$ when $z = 1$. In our more general model, it seems natural to define a *t-complier* as an observation i that is in treatment arm t when assigned an instrument value z such that $t_i^*(z) = t$, and only then. This is, clearly, a composite group. Take the 3×3 model as an example, and assume that $t^*(1) = 1$. Then the set of 1-compliers consists of the five response-groups C_{010}, C_{012}, C_{111}, C_{112}, and C_{212}.

2.4 Strict one-to-one targeting

We now impose one-to-one targeting (Assumption 4) as well as strict targeting. Under one-to-one targeting, the sets $Z^*(t)$ and $T^*(z)$ are singletons; and each targeting instrument z can be relabeled as the treatment value $t = t_i^*(z)$ that it targets.

Corollary 1 (Treatment assignment under strict, one-to-one targeting)**.** *Take any observation i. Let A_i be the (possibly empty) subset of $t \in \mathcal{T}^*$ such that $T_i(t) = t$. Then under Assumptions 1 to 5,*

[13] Note that this follows from our assumption that the distribution of the random vector $(u_{it})_{t \in \mathcal{T}}$ is absolutely continuous; however, it does not extend to the sets $Z^*(t)$ and $T^*(z)$, which can still have several elements.

1. *$T_i(t) = \underline{t}_i$ for all $t \in \mathcal{T} \setminus A_i$;*

2. *if $\underline{t}_i$ is a targeted treatment, it must belong to A_i.*

The pair $(A_i, \underline{t}_i)$ defines an elemental response group which we denote $C(A_i, \underline{t}_i)$. The family of sets $\{C(A,t) \mid A \subset \mathcal{T}^*, t \notin \mathcal{T}^* \setminus A\}$ form a partition of the set of observations.

Note that the $C(A,t)$ notation is just a shortcut: every $C(A,t)$ is an elemental group, and every elemental group is a $C(A,t)$. If for instance $|\mathcal{T}| = 6$, it is just more convenient to write $C(\{1,3\},2)$ than to write C_{212322}.

If $\underline{t}_i \notin \mathcal{T}^*$ and the set A_i is non-empty, then the observation i is what one could call a *strict A_i-complier*: when the value of the instrument moves from $\mathcal{Z} \setminus A_i$ to $t \in A_i$, observation i switches from its top alternative treatment $\underline{t}_i$ to the treatment t. In the 3-by-3 model with $\mathcal{T}^* = \{1,2\}$, there are three groups of strict compliers: $C_{010} = C(\{1\},0)$, $C_{002} = C(\{2\},0)$, and $C_{012} = C(\{1,2\},0)$.

Strict one-to-one targeting brings us very close to the main identifying assumption in Heckman and Vytlacil (2007b, Assumption B-2a, p. 5006): the indicator variable $\mathbb{1}(Z = t)$ can be used as the $Z^{[t]}$ in their assumption. Heckman and Vytlacil use their Assumption B-2a to identify the effect of the preferred treatment t relative to the next-best treatment. Their complier group consists of those individuals who choose treatment t under $Z = z$ and another treatment under $Z = z'$. This can be a very heterogeneous group, as our examples will show. To paraphrase Heckman and Vytlacil (2007b, p. 5013): the mean effect of treatment t versus the next best option is a weighted average over $t' \in \mathcal{T} \setminus \{t\}$ of the effect of treatment t versus treatment t', conditional on t' being the next best option, weighted by the probability that t' is the next best option. In contrast, we seek a complete characterization of all treatment effects that can be identified under this set of assumptions.

3 Identification

Now that we have characterized response-groups, we seek to identify the probabilities of the corresponding response-groups in the treatment model. Let $P(t|z) \equiv \Pr(T_i = t | Z_i = z)$ denote the generalized propensity score. Under strict, one-to-one targeting, the response-groups are easily enumerated.

Proposition 3 (Counting response-groups under strict, one-to-one targeting)**.** *Suppose that p treatment values are targeted and q are not. Under Assumptions 1 to 5, the number of response-groups is $N \equiv (p + 2q) \times 2^{p-1}$.*

As the probabilities of the response-groups must sum to one, we have $(N-1)$ unknowns. The data gives us the generalized propensity scores $P(t|z)$ for $(t,z) \in \mathcal{T} \times \mathcal{Z}$. The adding-up constraints $\sum_{t \in \mathcal{T}} P(t|z) = 1$ for each $z \in \mathcal{Z}$ reduce the count of independent data points to $(|\mathcal{T}|-1) \times |\mathcal{Z}| = (p+q-1)(p+1)$.

Table 3: Identifying the sizes of the response groups under strict, one-to-one targeting

	$\mathcal{T}$	p	q	Unknowns	Equations	Required
LATE	({0,1}	1	1	2	2	0
Example 1	{0,1,...,$\|\mathcal{T}\|-1$}	1	$\|\mathcal{T}\|-1$	$2(\|\mathcal{T}\|-1)$	$2(\|\mathcal{T}\|-1)$	0
Example 2	{0,1,2}	2	1	7	6	1

Table 3 shows some values of the number of equations and the number of unknowns $(N-1)$ for three examples. The first row has $|\mathcal{T}| = |\mathcal{Z}| = 2$; it generates the standard LATE case, where the response group consists of never-takers (C_{00}), compliers (C_{01}), and always-takers (C_{11}). The second row is another case of exact identification. The third row shows that one restriction is required to identify the sizes of the response-groups for the 3×3 model. More generally, the degree of underidentification increases exponentially with the number of targeted treatments p. The probabilities of the different groups are linked to the generalized propensity scores by a system of linear equations. Under strict, one-to-one targeting, this system takes a simple form, as mentioned in Section 3.

Proposition 4 (Identifying equations for group sizes under strict, one-to-one targeting)**.** *Under Assumptions 1 to 5, the generalized propensity scores satisfy the following system of equations, for all* $(z,t) \in \mathcal{Z} \times \mathcal{T}$*:*

$$P(t|z) = \sum_{A \subset \mathcal{T}^* \setminus \{z\}} \mathbb{1}(t \in \bar{A}) \Pr(i \in C(A,t)) + \sum_{A \subset \mathcal{T}^*} \mathbb{1}(t = z \in A) \sum_{\tau \in \bar{A}} \Pr(i \in C(A,\tau)) \tag{3.1}$$

where we denote $\bar{A} \equiv (\mathcal{T} \setminus \mathcal{T}^*) \bigcup A$.

While this may look cryptic, it is directly related to Corollary 1: the first line corresponds to $z \in \mathcal{T} \setminus A_i$ and $t = \underline{t}_i$, and the second line corresponds to $t = z \in A_i$. The set $\bar{A}_i \equiv (\mathcal{T} \setminus \mathcal{T}^*) \bigcup A_i$ contains the non-targeted treatments and those for which $T_i(t) = t$.

To simplify the exposition, we introduce one more element of notation. For any $z \in \mathcal{Z}$ and $t \in \mathcal{T}$, we define the *conditional average outcome* by $\bar{E}_z(t) \equiv \mathbb{E}(Y_i \mathbb{1}(T_i = t)|Z_i = z)$. For any response-group C and treatment value $t \in \mathcal{T}$, we define the *group average outcome*

as $\mathbb{E}(Y_i(t)|i \in C)$. While the conditional average outcomes $\bar{E}_z(t)$ are directly identified from the data, the group average outcomes of course are not. We do know that some of them are zero; and that they combine with the group probabilities to form the conditional average outcomes. We will repeatedly use the following identity from Heckman and Pinto (2018, Theorem T-1):

Lemma 1 (Group- and conditional average outcomes—Theorem T-1 of Heckman and Pinto (2018))**.** *Let $z \in \mathcal{Z}$ and $t \in \mathcal{T}$. Then*

$$\bar{E}_z(t) = \sum_{C=C_R \mid R(z)=t} \mathbb{E}(Y_i(t) \mid i \in C) \Pr(i \in C).$$

In addition,

$$\mathbb{E}(Y_i \mid Z_i = z) = \sum_{t \in \mathcal{T}} \bar{E}_z(t).$$

Under strict, one-to-one targeting, the set of response-groups $C = C_R$ such that $R(z) = t$ is as enumerated in Proposition 4: it consists of

- all $C(A, t)$ such that $A \subset \mathcal{T}^* \setminus \{z\}$ and $t \in \bar{A}$;
- and, if $t = z$, all $C(A, \tau)$ for $z \in A \subset \mathcal{T}^*$ and $\tau \in \bar{A}$.

The combination of Lemma 1 and of either Proposition 2 (under strict targeting) or Proposition 4 (under strict, one-to-one targeting) does *not* exhaust the empirical content of the model. A succession of papers[14] has given necessary and sometimes sufficient conditions for data to be rationalized under an instrument exclusion restriction. Most recently, Bai and Tabord-Meehan (2024) characterized the sharp testable implications of ARUM under joint independence[15].

To simplify the exposition, we will state our results in terms of effects of the treatment on the expectation of the outcomes; they hold, however, for any measurable function of the outcome $f(Y)$. Note that if we chose $f(Y) = \mathbb{1}(Y \leqslant t)$ for some value t, we would identify the effects of the treatment on the cumulative distribution function of the outcome. By inversion, we would recover the quantile treatment effects.

3.1 Positive Selection

We will sometimes make use of an identifying assumption that we call *positive selection*. It obtains when for some treatment value t and response groups $C \neq C'$ that sometimes

[14] See Balke and Pearl (1997), Kitagawa (2015), Mourifié and Wan (2017), Kédagni and Mourifié (2020) and Sun (2023).

[15] That is, when Z_i is independent of $\{(Y_i(t), T_i(z)) : (t, z) \in \mathcal{T} \times \mathcal{Z}\}$.

choose t,[16] we have $\mathbb{E}(Y_i(t)|i \in C) \leqslant \mathbb{E}(Y_i(t)|i \in C')$. The identifying power of positive selection depends on the context; we will illustrate it in Corollaries 2 and 3 as well as in our application to Head Start in Section 4. A slightly different definition would replace $Y_i(t)$ with a treatment effect $Y_i(t) - Y_i(t')$; we explore this variant in Corollary 4.

3.2 The Binary Instrument Model

Recall that with a binary instrument, strict targeting is trivially satisfied. Under one-to-one targeting, Proposition 4 can be applied directly to some of the rows of Table 3.

3.2.1 Identification Under One-to-one Targeting

The second row of Table 3 shows that the group probabilities are just identified in our Example 1 under strict, one-to-one targeting. Proposition 4 gives $2(T-1)$ independent equations: for $t \neq 1$,

$$P(t|0) = \Pr(i \in C(\varnothing, t)) + \Pr(i \in C(\{1\}, t)) \text{ and } P(t|1) = \Pr(i \in C(\varnothing, t)).$$

Moreover, $C(\varnothing, t) = C_{tt}$ for $t \neq 1$ and $C(\{1\}, t) = C_{t1}$ for all t.

Note that when z changes from 0 to 1, the only observations that change treatment are in C_{t1} for $t \neq 1$. Since the corresponding C_{1t} group is empty, there are no "two-way flows" and this model satisfies the unordered monotonicity property of Heckman and Pinto (2018). Proposition 5 gives explicit formulæ for the probabilities of all $(2|\mathcal{T}| - 1)$ response groups.

Proposition 5 (Response-group probabilities in Example 1 under one-to-one targeting)**.** *Under Assumptions 1 to 5, the following probabilities are identified:*

$$\begin{aligned} \Pr(C_{11}) &= P(1|0), \\ \Pr(C_{tt}) &= P(t|1) \text{ for } t \neq 1, \\ \Pr(C_{t1}) &= P(t|0) - P(t|1) \text{ for } t \neq 1. \end{aligned} \tag{3.2}$$

Since $\Pr(C_{t1}) \geqslant 0$, the model has $(|T| - 1)$ simple testable predictions: $P(t|0) \geqslant P(t|1)$ for $t \neq 1$. While all the response group probabilities are point-identified, only some group average outcomes are point identified without further restrictions, as shown by Proposition 6.

[16] That is, for which $\Pr(T_i = t|i \in C)$ and $\Pr(T_i = t|i \in C')$ are nonzero.

Proposition 6 (Group average outcomes in Example 1 under one-to-one targeting)**.** *Under Assumptions 1 to 5, the following group average outcomes are point-identified:*

$$\begin{aligned}
\mathbb{E}\left[Y_i(1)|i \in C_{11}\right] &= \frac{\bar{E}_0(1)}{P(1|0)}, \\
\mathbb{E}\left[Y_i(t)|i \in C_{tt}\right] &= \frac{\bar{E}_1(t)}{P(t|1)} \text{ for } t \neq 1, \\
\mathbb{E}\left[Y_i(t)|i \in C_{t1}\right] &= \frac{\bar{E}_0(t) - \bar{E}_1(t)}{P(t|0) - P(t|1)} \text{ for } t \neq 1.
\end{aligned}$$

However, if $T > 2$ the standard Wald estimator only identifies a convex combination of the LATEs on the complier groups C_{t1}:

$$\begin{aligned}
\frac{\mathbb{E}(Y_i|Z_i = 1) - \mathbb{E}(Y_i|Z_i = 0)}{\Pr(T_i = 1|Z_i = 1) - \Pr(T_i = 1|Z_i = 0)} &= \frac{(\bar{E}_1(1) - \bar{E}_0(1)) - \sum_{t \neq 1}(\bar{E}_0(t) - \bar{E}_1(t))}{P(1|1) - P(1|0)} \\
&= \sum_{t \neq 1} \alpha_t \mathbb{E}\left[Y_i(1) - Y_i(t)|i \in C_{t1}\right], \qquad (3.3)
\end{aligned}$$

where the weights $\alpha_t = \Pr(i \in C_{t1}|i \in \bigcup_{\tau \neq 1} C_{\tau 1}) = (P(t|0) - P(t|1))/(P(1|1) - P(1|0))$ are identified, positive, and sum to 1. If $T = 2$, we have $\alpha_0 = 1$ and the familiar LATE formula

$$\mathbb{E}(Y_i(1) - Y_i(0)|i \in C_{01}) = \frac{\mathbb{E}(Y_i|Z_i = 1) - \mathbb{E}(Y_i|Z_i = 0)}{\Pr(T_i = 1|Z_i = 1) - \Pr(T_i = 1|Z_i = 0)}.$$

Proposition 6 shows that we only identify a known convex combination of the $(|\mathcal{T}| - 1)$ LATEs[17]. This formula is reminiscent of Angrist and Imbens (1995, Theorem 1), which deals with a different model in which treatments are ordered. It is possible to re-derive our identification results in Propositions 5 and 6 using the general framework of Heckman and Pinto (2018). We provide details in Appendix D.

So far, we only imposed restrictions on the process by which treatment values are assigned to observations; this is what Goff (2024b) calls an "outcome-agnostic" approach in that it only assumes that the instruments are excluded from the outcome equations. It is possible to bound the average treatment effects in a straightforward manner if we assume that the support of the outcomes is known and bounded. One could instead add restrictions to achieve point identification of average treatment effects for the compliers. Assuming that the ATEs are all equal is one obvious solution. Another one is to assume some degree of homogeneity of group average outcomes. Alternatively, we may consider weaker conditions under which

[17]We use the term "LATEs" for the average treatment effects on the various complier groups. Throughout the remainder of the paper, we assume, as is standard, that probability differences appearing in the denominator of estimands are always nonzero.

the average treatment effects for the compliers are only partially identified. We explore these ideas below.

3.2.2 Adding Identification Constraints

Consider the binary instrument model with $T \geqslant 3$.

Beyond One-to-one Targeting First note that the probabilities of the response-groups can be identified under weaker restrictions than one-to-one targeting. Suppose for instance that $z = 1$ targets all treatment values $t \geqslant 1$: we have $U_1(t) > U_0(t)$ for all $t \geqslant 1$. Then the complier groups C_{t0} for $t \geqslant 1$ must be empty. To see this, suppose that $T_i(0) = t \geqslant 1$. This implies $U_0(t) + u_{it} > U_0(0) + u_{i0} = u_{i0}$. Adding up these inequalities gives $U_1(t) + u_{it} > u_{i0}$, and $T_i(1)$ cannot be 0.

All other groups $C_{tt'}$ may exist. This leaves $|\mathcal{T}|(|\mathcal{T}| - 1)$ unknown group probabilities, which is $|\mathcal{T}|/2$ times more than the $2(|\mathcal{T}| - 1)$ propensity scores we observe. We need $(|\mathcal{T}| - 1)(|\mathcal{T}| - 2)$ additional constraints to point-identify all group probabilities.

Single-peaked Mean Utilities Now suppose that mean utilities are "single-peaked" in the sense that the function $t \to U_1(t) - U_0(t)$ is decreasing over $t = 1, \ldots, T - 1$. This would be a reasonable assumption if $z = 1$ makes treatment $t = 1$ more attractive and the treatments $t > 1$ are ordered by their proximity to $t = 1$.

If this holds, then the same argument as above shows that the response groups $C_{tt'}$ must be empty when $t' > t \geqslant 1$. This eliminates $(|\mathcal{T}| - 1)(|\mathcal{T}| - 2)/2$ response groups; we divided by two the number of additional identification constraints that we need.

Positive Selection The binary instrument model gives a first example of the power of the positive selection defined in Section 3.1. Take $\tau \neq 1$ and consider the complier groups $C_{\tau 1}$: they all have $t = 1$ when $z = 1$, but they shift to it from different treatment values τ under $z = 0$. Depending on the context, there may be a plausible reason to order the corresponding group average outcomes when $t = 1$. Suppose for instance that $T = 3$, and that

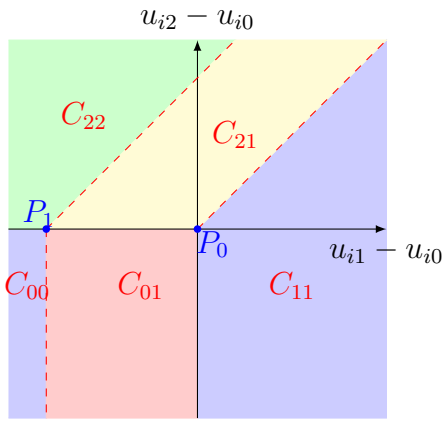
$$\mathbb{E}\left[Y_i(1) | i \in C_{01}\right] \leqslant \mathbb{E}\left[Y_i(1) | i \in C_{21}\right]. \tag{3.4}$$

In this 2×3 model under one-to-one targeting, there are five response-groups, as illustrated in Figure 3. Proposition 6 shows that the Wald estimator only identifies

$$\alpha_0 \mathbb{E}\left[Y_i(1) - Y_i(0) | i \in C_{01}\right] + (1 - \alpha_0)\mathbb{E}\left[Y_i(1) - Y_i(2) | i \in C_{21}\right],$$

where $\alpha_0 = (P(0|0) - P(0|1))/(P(1|1) - P(1|0))$ is point-identified. Corollary 2 shows that adding inequality (3.4) yields bounds on the corresponding LATEs.

Figure 3: A 2×3 model with one targeted treatment



Corollary 2 (Positive selection and treatment effects in the 2×3 model under one-to-one targeting)**.** *If*

$$\mathbb{E}\left[Y_i(1)|i \in C_{01}\right] \leqslant \mathbb{E}\left[Y_i(1)|i \in C_{21}\right], \tag{3.5}$$

then the local average treatment effects for C_{01} and C_{21} are partially identified:

$$\begin{aligned} \mathbb{E}\left[Y_i(1) - Y_i(0)|i \in C_{01}\right] &\leqslant \frac{\bar{E}_1(1) - \bar{E}_0(1)}{P(1|1) - P(1|0)} - \frac{\bar{E}_0(0) - \bar{E}_1(0)}{P(0|0) - P(0|1)}, \\ \mathbb{E}\left[Y_i(1) - Y_i(2)|i \in C_{21}\right] &\geqslant \frac{\bar{E}_1(1) - \bar{E}_0(1)}{P(1|1) - P(1|0)} - \frac{\bar{E}_0(2) - \bar{E}_1(2)}{P(2|0) - P(2|1)}, \end{aligned} \tag{3.6}$$

Moreover, (3.5) *implies the following testable prediction:*

$$\frac{\bar{E}_0(2) - \bar{E}_1(2)}{P(2|0) - P(2|1)} \geqslant \frac{\bar{E}_0(0) - \bar{E}_1(0)}{P(0|0) - P(0|1)}. \tag{3.7}$$

If (3.5) *holds at equality, then the two statements in* (3.6) *and the testable prediction in* (3.7) *also become equalities, and the two LATEs are point-identified.*

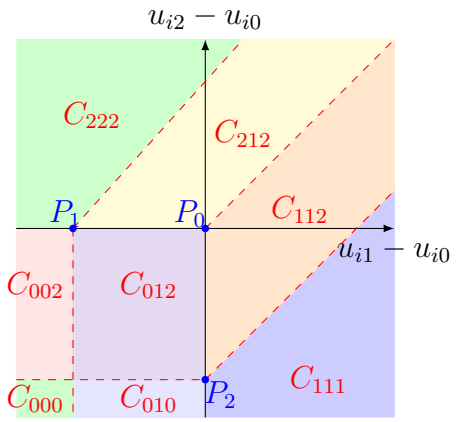
The lower bounds on the local average treatment effects for C_{01} and C_{21} may not be sharp; on the other hand, they are easy to estimate from sample averages. It is a topic for future research to obtain the sharp bounds and develop a corresponding method for estimation and inference.

3.3 The 3×3 Model

Let us now turn to the 3×3 model of Example 2, where $\mathcal{Z}^* = \mathcal{T}^* = \{1, 2\}$ and $\mathcal{Z} = \mathcal{T} = \{0, 1, 2\}$. We assume strict one-to-one targeting: for all of our results in this section, we impose Assumptions 1 - 5; $z = 1$ targets $t = 1$ and $z = 2$ targets $t = 2$.

The set A in Corollary 1 can be $\varnothing, \{1\}, \{2\}$, or $\{1, 2\}$, with corresponding values of t in $\{0\}, \{0, 1\}, \{0, 2\}$ or $\{0, 1, 2\}$ respectively. The set $c(\varnothing, 0)$ corresponds to the never-takers C_{000}. For $A = \{1\}$ we get C_{010} and C_{111}, and for $A = \{2\}$ we get C_{002} and C_{222}. Finally, with $A = \{1, 2\}$ we have C_{012}, C_{112}, and C_{212}.

Figure 4: Strictly one-to-one targeted treatment in the 3×3 model



These eight elemental response groups are illustrated in Figure 4, again with the origin in P_0. Comparing Figure 4 with Figure 2 shows the identifying power of Assumption 5. Table 4 shows which groups take $T_i = t$ when $Z_i = z$.

Table 4: Response Groups of Example 2

	$T_i(z) = 0$	$T_i(z) = 1$	$T_i(z) = 2$
$z = 0$	$C_{000} \bigcup C_{010} \bigcup C_{002} \bigcup C_{012}$	$C_{111} \bigcup C_{112}$	$C_{222} \bigcup C_{212}$
$z = 1$	$C_{000} \bigcup C_{002}$	$C_{111} \bigcup C_{010} \bigcup C_{012} \bigcup C_{112} \bigcup C_{212}$	C_{222}
$z = 2$	$C_{000} \bigcup C_{010}$	C_{111}	$C_{222} \bigcup C_{002} \bigcup C_{012} \bigcup C_{112} \bigcup C_{212}$

Unlike the 2×3 model, even under strict one-to-one targeting the 3×3 model does not satisfy unordered monotonicity. One could show it with the matrix algebra in Heckman and Pinto (2018).[18] It is more straightforward to note that when thee instrument value changes

[18] See Appendix D for details.

from $z = 1$ to $z = 2$, observations in C_{010} move to treatment value 0, while observations in C_{002} leave treatment 0. This is the definition of a two-way flow, which violates unordered monotonicity. Since the 3×3 model has three instrument values and only two targeted treatments, Bai, Huang, Moon, Shaikh, and Vytlacil (2024, Example 4.7) shows that it satisfies their weaker general monotonicity assumption. As a consequence, the average potential outcomes $\mathbb{E}[Y_i(d)]$ can only be restricted by identification at infinity arguments.

3.3.1 Identification in the 3×3 Model

We know from the third row of Table 3 that one restriction is missing to point-identify the probabilities of all eight response-groups. The following proposition shows that the probabilities of four of the eight elemental groups are point-identified: two groups of always-takers, and two groups of compliers. The other four probabilities are constrained by three adding-up constraints.

Proposition 7 (Response-group probabilities in the 3×3 model under strict, one-to-one targeting)**.** *The following four probabilities are identified:* $\Pr(C_{111}) = P(1|2)$*,* $\Pr(C_{222}) = P(2|1)$*,* $\Pr(C_{112}) = P(1|0) - P(1|2)$*, and* $\Pr(C_{212}) = P(2|0) - P(2|1)$*. The remaining four response group probabilities are partially-identified and can be parameterized as:* $\Pr(C_{000}) = p$*,* $\Pr(C_{002}) = P(0|1) - p$*,* $\Pr(C_{010}) = P(0|2) - p$*, and* $\Pr(C_{012}) = P(0|0) - P(0|1) - P(0|2) + p$*, where the unknown* p *satisfies* $\max\{0, P(0|1) + P(0|2) - P(0|0)\} \leqslant p \leqslant \min\{1, P(0|1), P(0|2)\}$*.*

As before, the model has the following testable implications: $P(1|1) \geqslant P(1|0) \geqslant P(1|2)$, $P(2|2) \geqslant P(2|0) \geqslant P(2|1)$, and $P(0|0) \geqslant \max(P(0|1), P(0|2))$. The following proposition identifies a number of group average outcomes[19].

Proposition 8 (Group average outcomes in the 3×3 model under strict, one-to-one targeting)**.**

[19] Again, these could also be derived using the general framework of Heckman and Pinto (2018), even though the unordered monotonicity assumption is not satisfied—see Appendix D.

The following group average outcomes are point-identified:

$$\mathbb{E}\left[Y_i(0)|i \in C_{000} \bigcup C_{002}\right] = \frac{\bar{E}_1(0)}{P(0|1)}, \qquad \mathbb{E}\left[Y_i(0)|i \in C_{000} \bigcup C_{010}\right] = \frac{\bar{E}_2(0)}{P(0|2)},$$

$$\mathbb{E}\left[Y_i(1)|i \in C_{111}\right] = \frac{\bar{E}_2(1)}{P(1|2)}, \qquad \mathbb{E}\left[Y_i(2)|i \in C_{222}\right] = \frac{\bar{E}_1(2)}{P(2|1)},$$

$$\mathbb{E}\left[Y_i(0)|i \in C_{010} \bigcup C_{012}\right] = \frac{\bar{E}_0(0) - \bar{E}_1(0)}{P(0|0) - P(0|1)}, \quad \mathbb{E}\left[Y_i(0)|i \in C_{002} \bigcup C_{012}\right] = \frac{\bar{E}_0(0) - \bar{E}_2(0)}{P(0|0) - P(0|2)},$$

$$\mathbb{E}\left[Y_i(1)|i \in C_{010} \bigcup C_{012} \bigcup C_{212}\right] = \frac{\bar{E}_1(1) - \bar{E}_0(1)}{P(1|1) - P(1|0)}, \qquad \mathbb{E}\left[Y_i(1)|i \in C_{112}\right] = \frac{\bar{E}_0(1) - \bar{E}_2(1)}{P(1|0) - P(1|2)},$$

$$\mathbb{E}\left[Y_i(2)|i \in C_{002} \bigcup C_{012} \bigcup C_{112}\right] = \frac{\bar{E}_2(2) - \bar{E}_0(2)}{P(2|2) - P(2|0)}, \qquad \mathbb{E}\left[Y_i(2)|i \in C_{212}\right] = \frac{\bar{E}_0(2) - \bar{E}_1(2)}{P(2|0) - P(2|1)}.$$

By itself, Proposition 8 does not allow us to identify an average treatment effect for *any* (even composite) response-group. Suppose for instance that we want to identify $\mathbb{E}(Y_i(1) - Y_i(0)|i \in C)$ for some group C. Then C needs to exclude C_{111}, C_{112}, and C_{212}, since $\mathbb{E}(Y_i(0)|i \in C')$ is not identified for any group C' that contains C_{111}, C_{112}, or C_{212}. Since we only know the mean outcome of treatment 1 for groups that contain one of these three elemental groups, the conclusion follows.

3.3.2 Using Positive Selection

Note that if we assumed $\mathbb{E}(Y_i(1)|i \in C_{112}) = \mathbb{E}(Y_i(1)|i \in C_{212})$, then we could combine the two equations in the fourth displayed line of Proposition 8 and the probabilities of C_{112} and C_{212} (which are point-identified by Proposition 7) to obtain $\mathbb{E}(Y_i(1)|i \in C_{010} \bigcup C_{012})$. This would point-identify the average effect of treatment 1 vs treatment 0 on this composite complier group C_{01*}. While this assumption may be overly strong, it seems natural to impose that $Y_i(\tau)$ is on average larger in a response group that has $t = \tau$ for more values of z. Assumption 6 formalizes this intuition in our setting.

Assumption 6 (Positive selection in the 3×3 model)**.** *Either or both of the following assumptions hold:*

$$\mathbb{E}\left[Y_i(1)|i \in C_{112}\right] \geqslant \mathbb{E}\left[Y_i(1)|i \in C_{212}\right], \tag{3.8}$$

$$\mathbb{E}\left[Y_i(2)|i \in C_{212}\right] \geqslant \mathbb{E}\left[Y_i(2)|i \in C_{112}\right]. \tag{3.9}$$

Assumption 6 states a form of positive selection into treatment, as defined in Section 3.1. Consider Equation (3.8) for instance. It says that within the group of "12-compliers" $C_{*12} = C_{012} \bigcup C_{112} \bigcup C_{212}$, those observations with $T(0) = 1$ have a larger average counterfactual

$Y(1)$ than those with $T(0) = 2$. Corollary 3 shows that this gives bounds on the local average treatment effects for C_{01*}-compliers, with a similar result for Equation (3.9) and C_{0*2}-compliers.

Corollary 3 (Identifying treatment effects in the 3×3 model)**.** *1. Under* (3.8)*, the local average treatment effect*

$$\mathbb{E}\left[Y_i(1) - Y_i(0) | i \in C_{01*}\right]$$

is at least as large as

$$\frac{(\bar{E}_1(1) - \bar{E}_0(1)) - (\bar{E}_0(0) - \bar{E}_1(0))}{P(0|0) - P(0|1)} - \frac{\bar{E}_0(1) - \bar{E}_2(1)}{P(1|0) - P(1|2)} \frac{P(2|0) - P(2|1)}{P(0|0) - P(0|1)}.$$

2. Under (3.9)*, the local average treatment effect*

$$\mathbb{E}\left[Y_i(2) - Y_i(0) | i \in C_{0*2}\right]$$

is at least as large as

$$\frac{(\bar{E}_2(2) - \bar{E}_0(2)) - (\bar{E}_0(0) - \bar{E}_2(0))}{P(0|0) - P(0|2)} - \frac{\bar{E}_0(2) - \bar{E}_1(2)}{P(2|0) - P(2|1)} \frac{P(1|0) - P(1|2)}{P(0|0) - P(0|2)}.$$

3. In both 1 and 2, "at least as large" can be replaced with "equals" if the corresponding inequality in Assumption 6 is an equality.

3.3.3 When is Positive Selection Plausible?

Let us focus on (3.9). Given strict one-to-one targeting, C_{112} is defined by

$$\underline{U}(1) - \bar{U}(2) \leqslant u_{i2} - u_{i1} \leqslant \underline{U}(1) - \underline{U}(2),\ u_{i1} - u_{i0} \geqslant -\underline{U}(1).$$

C_{212} is defined by

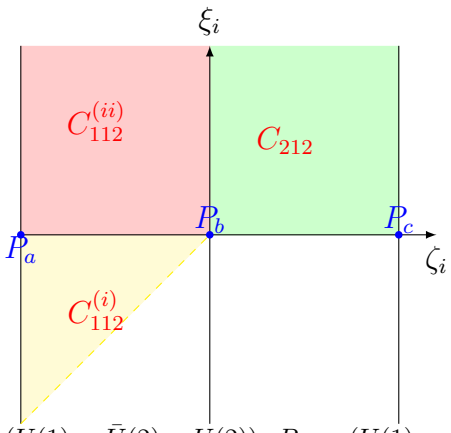
$$\underline{U}(1) - \underline{U}(2) \leqslant u_{i2} - u_{i1} \leqslant \bar{U}(1) - \underline{U}(2),\ u_{i2} - u_{i0} \geqslant -\underline{U}(2).$$

To simplify notation, define $\zeta_i = u_{i2} - u_{i1}$ and $\xi_i = u_{i2} - u_{i0}$, so that $u_{i1} - u_{i0} = \xi_i - \zeta_i$. The inequalities above can be rewritten as

- for C_{112}: $\underline{U}(1) - \bar{U}(2) \leqslant \zeta_i \leqslant \underline{U}(1) - \underline{U}(2)$, $\xi_i - \zeta_i \geqslant -\underline{U}(1)$;
- for C_{212}: $\underline{U}(1) - \underline{U}(2) \leqslant \zeta_i \leqslant \bar{U}(1) - \underline{U}(2)$, $\xi_i \geqslant -\underline{U}(2)$.

Figure 5 plots these two groups on the $\zeta_i \times \xi_i$ plane. Group C_{212} corresponds to the top-right (infinite) rectangle and group C_{112} is partitioned into the two subgroups: $C_{112}^{(i)}$ is a bottom-left triangle and $C_{112}^{(ii)}$ is a top-left (infinite) rectangle.

Figure 5: Positive Selection in the Generalized 3×3 Roy model



Notes: $P_a = (\underline{U}(1) - \bar{U}(2), -\underline{U}(2))$, $P_b = (\underline{U}(1) - \underline{U}(2), -\underline{U}(2))$, and $P_c = (\bar{U}(1) - \underline{U}(2), -\underline{U}(2))$.

Now suppose that

Assumption 7 (Positive Codependence)**.** *(1) If A and B are two measurable sets in the (ζ, ξ) plane such that*

$$(\zeta, \xi) \in A \text{ and } (\zeta', \xi') \in B \implies (\zeta \leqslant \zeta' \text{ and } \xi' \leqslant \xi'),$$

then $\mathbb{E}(Y_i(2)|(\zeta_i, \xi_i) \in A) \leqslant \mathbb{E}(Y_i(2)|(\zeta_i, \xi_i) \in B)$.

(2) If A and B are two measurable sets on the real line such that

$$\zeta \in A \text{ and } \zeta' \in B \implies \zeta \leqslant \zeta'$$

then $\mathbb{E}(Y_i(2)|\zeta_i \in A, \xi_i \geqslant b) \leqslant \mathbb{E}(Y_i(2)|\zeta_i \in B, \xi_i \geqslant b)$ for all b.

As shown in Figure 5, every point in $C_{112}^{(i)}$ has lower values of both ζ_i and ξ_i than any point in C_{212}. Therefore, by Assumption 7(1), the expected value of $Y_i(2)$ in this triangle is smaller than $\mathbb{E}(Y_i(2)|i \in C_{212})$. Every point in $C_{112}^{(ii)}$ has is a smaller value of ζ_i than at any point in C_{212} (fixing the value of $\xi_i \geqslant -\underline{U}(2)$ on both sides). Assumption 7(2) implies that the expected value of $Y_i(2)$ in this rectangle again is smaller than $\mathbb{E}(Y_i(2)|i \in C_{212})$. Combining these two inequalities gives $\mathbb{E}(Y_i(2)|i \in C_{212}) \geqslant \mathbb{E}(Y_i(2)|i \in C_{112})$, that is (3.9).

Assumption 7 seems weak. Because both $\zeta_i = u_{i2} - u_{i1}$ and $\xi_i = u_{i2} - u_{i0}$ are increasing in u_{i2}, this assumption aligns well with the concept of positive selection. Suppose for instance

that

$$\mathbb{E}(Y_i(2)|u_{i0}, u_{i1}, u_{i2}) - \mathbb{E}Y_i(2) = a_0 u_{i0} + a_1 u_{i1} + a_2 u_{i2},$$

where a_0, a_1, and a_2 are some constants, and that (u_{i0}, u_{i1}, u_{i2}) are jointly normal and mutually uncorrelated with the common mean 0 and the common variance 1. We interpret (u_{i0}, u_{i1}, u_{i2}) as the underlying primitive random variables that are normalized to have mean 0 and variance 1. It is easy to derive[20] that

$$\mathbb{E}(Y_i(2)|\zeta_i, \xi_i) - \mathbb{E}Y_i(2) = \frac{a_2 + a_0 - 2a_1}{3}\zeta_i + \frac{a_2 + a_1 - 2a_0}{3}\xi_i.$$

Hence, in this example, Assumption 7 holds if and only if $a_2 + a_0 \geqslant 2a_1$ and $a_2 + a_1 \geqslant 2a_0$. In summary, a sufficiently large value of a_2 induces positive selection, generating patterns similar to those of comparative advantage in generalized Roy models.

3.4 What do the IV estimators identify in the 3×3 model?

Kirkeboen, Leuven, and Mogstad (2016, hereafter KLM) used a 3×3 model to study the impact of the field of study on later earnings. Their Proposition 2 characterizes what two-stage least squares (TSLS) estimators identify under different sets of assumptions. The least stringent version combines a monotonicity assumption (Assumption 4 in KLM) and condition (iii) in their Proposition 2, which they call "irrelevance and information on next-best alternatives". "Irrelevance" is a set of exclusion restrictions, while "information on next-best alternatives" assumes the availability of additional data.

3.4.1 Monotonicity and Irrelevance

While we take quite a different path, our strict one-to-one targeting assumption turns out to yield exactly the same identifying restrictions as the combination of monotonicity and irrelevance in KLM. We show it in Appendix C.

This set of assumptions in itself is too weak to give two-stage least squares estimates a simple interpretation. To see this, let β_1 and β_2 be the probability limits of the coefficients in a regression of Y_i on the indicator variables $\mathbb{1}(T_i = 1)$ and $\mathbb{1}(T_i = 2)$, with instruments Z_i. Remember from Table 4 that under strict one-to-one targeting, five response-groups have $T(1) = 1$:

1. the always-takers C_{111};

2. the "intermediate" group C_{112}, which has $T(z) = 1$ unless $z = 2$;

[20] See Appendix G for details.

3. the three groups C_{010}, C_{012}, and C_{212}, which have $T(z) = 1$ if and only of $z = 1$.

A similar distinction applies to the groups that have $T(2) = 2$; it motivates Definition 6.

Definition 6 (1-compliers and 2-compliers)**.** We call

$$\mathcal{C}_1 = C_{010} \bigcup C_{012} \bigcup C_{212},$$

the 1-compliers group and

$$\mathcal{C}_2 = C_{002} \bigcup C_{012} \bigcup C_{112}$$

the 2-compliers group.

The β_1 and β_2 coefficients turn out to be weighted averages of the LATEs on these two groups and on the intermediate groups C_{112} and C_{212}.

Proposition 9 (TSLS in the 3×3 model under strict, one-to-one targeting)**.** *The parameters β_1 and β_2 satisfy*

$$\begin{pmatrix} \Pr(i \in \mathcal{C}_1) & -\Pr(i \in C_{212}) \\ -\Pr(i \in C_{112}) & \Pr(i \in \mathcal{C}_2) \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} = \begin{pmatrix} \mathbb{E}[\{Y_i(1) - Y_i(0)\}\mathbb{1}(i \in \mathcal{C}_1)] - \mathbb{E}[\{Y_i(2) - Y_i(0)\}\mathbb{1}(i \in C_{212})] \\ \mathbb{E}[\{Y_i(2) - Y_i(0)\}\mathbb{1}(i \in \mathcal{C}_2)] - \mathbb{E}[\{Y_i(1) - Y_i(0)\}\mathbb{1}(i \in C_{112})] \end{pmatrix}.$$

Proposition 9 implies that β_1 and β_2 are weighted averages of the four local average treatment effects on the right-hand side of this system of two equations. The weights are functions of the four probabilities on the left-hand side, which are point identified by Proposition 7. However, these weights may be positive or negative. This complicates interpretation further[21].

3.4.2 Additional Assumptions

Next-best alternatives Using the additional information on next-best alternatives in KLM amounts, in our notation, to dropping the "intermediate" response-groups C_{212} and C_{112} from the data. Then the system of equations in Proposition 9 becomes diagonal and it yields

$$\beta_1 = \mathbb{E}[Y_i(1) - Y_i(0) | i \in \mathcal{C}_1],$$
$$\beta_2 = \mathbb{E}[Y_i(2) - Y_i(0) | i \in \mathcal{C}_2],$$

[21] Mogstad, Torgovitsky, and Walters (2021) give a set of assumptions under which the weights are positive in a model with multiple binary instruments.

where now $\mathcal{C}_1$ reduces to $C_{010} \bigcup C_{012}$ and $\mathcal{C}_2$ reduces to $C_{002} \bigcup C_{012}$. This is exactly Proposition 2 (iii) of KLM. Alternatively, one may simply assume that the response-groups C_{212} and C_{112} are empty. This is the path taken by Bhuller and Sigstad (2024)[22].

Positive Selection Additional information of the type used by KLM often is not available. Moreover, assuming away C_{112} and C_{212} seems rather strong. On the other hand, reasonable assumptions can be used to generate bounds on the local average treatment effects for 1-compliers and 2-compliers. Corollary 4 illustrates this.

Corollary 4 (TSLS in the 3×3 model under strict, one-to-one targeting)**.** *Assume that*

$$\mathcal{D} \equiv \Pr(i \in \mathcal{C}_1) \Pr(i \in \mathcal{C}_2) - \Pr(i \in C_{212}) \Pr(i \in C_{112}) \neq 0. \tag{3.10}$$

Let

$$\mathcal{D}_1 \equiv \mathbb{E}\left(Y_i(1) - Y_i(0) | i \in \mathcal{C}_1\right) - \mathbb{E}\left(Y_i(1) - Y_i(0) | i \in C_{112}\right)$$

and

$$\mathcal{D}_2 \equiv \mathbb{E}\left(Y_i(2) - Y_i(0) | i \in \mathcal{C}_2\right) - \mathbb{E}\left(Y_i(2) - Y_i(0) | i \in C_{212}\right).$$

If $\mathcal{D}_1 \mathcal{D}_2 > 0$, then $\beta_1 - \mathbb{E}\left(Y_i(1) - Y_i(0) | i \in \mathcal{C}_1\right)$ and $\beta_2 - \mathbb{E}\left(Y_i(2) - Y_i(0) | i \in \mathcal{C}_2\right)$ have the sign of $\mathcal{D}$.

Note that the KLM result of the previous paragraph is the limit case where $\mathcal{D}_1 = \mathcal{D}_2 = 0$.

The regularity condition (3.10) ensures that the 2×2 matrix that premultiplies $(\beta_1, \beta_2)'$ in Proposition 9 be invertible[23]. To interpret the assumptions on signs, suppose that $\mathcal{D}_1$ is positive. Since $\mathcal{C}_1 = C_{010} \bigcup C_{012} \bigcup C_{212}$, the positivity of $\mathcal{D}_1$ states that the average effect of treatment 1 on $C_{010} \bigcup C_{012} \bigcup C_{212}$ is at least as large as on C_{112}. This is a form of positive selection that is in the same spirit as (but different from) Assumption 6. If this form of positive selection holds for both treatments, then the TSLS estimates overestimate the LATEs on the corresponding compliers if $\mathcal{D} > 0$, and they underestimate them if $\mathcal{D} < 0$.

To summarize, the TSLS estimators in the 3×3 model are difficult to interpret unless additional information is available and/or some additional assumptions are imposed. If the groups C_{112} and C_{212} are indeed empty, then both the TSLS estimators and those we obtained in Corollary 3 should identify the LATEs on the 1- and 2-compliers. Comparing their values is a useful (if informal) way of testing the assumptions and of exploring further the heterogeneity of the treatment effects.

[22] See their Corollary 5 and Table 1 for details.

[23] It holds if C_{212} and C_{112} have positive probability and either $C_{010} \bigcup C_{012}$ or $C_{002} \bigcup C_{012}$ has positive probability.

4 Empirical Example: The Head Start Impact Study

We now reexamine the Kline and Walters's (2016) analysis of the Head Start Impact Study (HSIS) using our framework. We use exactly the same data as they did; we only apply different identifying assumptions.

Head Start is a federal program in the US that addresses various factors affecting children's development in low-income families. It provides early childhood education (hereafter "preschool") and health and nutrition services. HSIS was a longitudinal study conducted from 2002 to 2010 to assess the program's impact on cognitive, social-emotional, and health outcomes. It focused on 84 communities where the demand for Head Start services was larger than the supply. HSIS randomly assigned about 5,000 three and four year old preschool children to either a treatment group which was offered Head Start services, or a control group which received no such offer. Children in either group could also attend other preschool centers if offered a slot

The structure of HSIS is identical to that of Example 1: it is a 2×3 model. The treatments here consist of no preschool (n), Head Start (h), and other preschool centers (c): $\mathcal{T} = \{n, h, c\}$. The instrument is binary, with a control group ($z = 0$) and a group that is offered admission to Head Start ($z = 1$). The outcome variable is test scores, measured in standard deviations from their mean.

In the terminology of this paper, treatment assignment satisfies strict, one-to-one targeting: strict targeting as the instrument is binary, and one-to-one targeting as $z = 1$ only targets Head Start[24]. Figure 6 reproduces Figure 3 in this setting. In addition to the three always-taker groups C_{nn}, C_{cc}, and C_{hh}, there are two complier groups: C_{nh}, and C_{ch}. In Sections 4.1 and 4.2, we focus on the LATEs on the two complier groups C_{nh} and C_{ch}. Section 4.3 embeds the model into a larger, 3×3 model in order to evaluate the marginal value of the public funds used in Head Start.

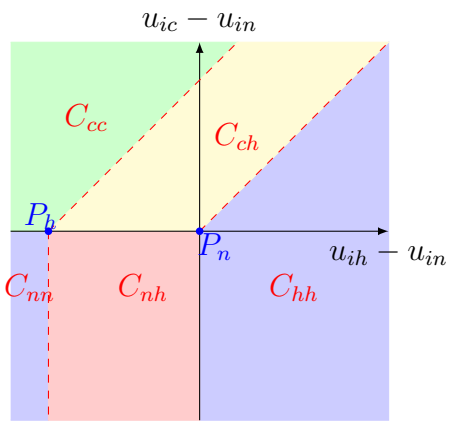
4.1 Group proportions and counterfactual means

Our estimates of the proportions of the two complier groups in the sample use (3.2) in Proposition 5; they are shown in Panel A of Table 5. As expected, they coincide with those in Kline and Walters (2016).

Panel B of Table 5 shows the counterfactual means of test scores for the complier groups, as per Proposition 6. While $\mathbb{E}[Y_i(n)|i \in C_{nh}]$ is negative, $\mathbb{E}[Y_i(c)|i \in C_{ch}]$ is above 0.1 standard deviation—not a negligible value in this field. This suggests that some of the

[24]Kamat (2024) analyzes HSIS using a different approach that focuses on how the choice sets available to a child vary with the value of the instrument.

Figure 6: The Kline and Walters (2016) Model of Preschool Choice



children who enter Head Start would have been at a good preschool otherwise. Kline and Walters (2016) call this pattern the "substitution effect" of Head Start. However, Kline and Walters (2016) do not report estimates of $\mathbb{E}[Y_i(n)|i \in C_{nh}]$ and $\mathbb{E}[Y_i(c)|i \in C_{ch}]$.

4.2 Treatment Effects

To fully measure the substitution effect, one needs to identify $\mathbb{E}\left[Y_i(h)|i \in C_{nh}\right]$ and $\mathbb{E}\left[Y_i(h)|i \in C_{ch}\right]$. However, we know from Proposition 6 that they are only partially identified by

$$\alpha_0 \mathbb{E}\left[Y_i(h)|i \in C_{ch}\right] + (1-\alpha_0)\mathbb{E}\left[Y_i(h)|i \in C_{nh}\right] = \frac{\mathbb{E}\left[Y_i \mathbb{1}(T_i = h)|Z_i = 1\right] - \mathbb{E}\left[Y_i \mathbb{1}(T_i = h)|Z_i = 0\right]}{P(h|1) - P(h|0)}.$$

where $\alpha_0 = (P(c|0) - P(c|1))/(P(h|1) - P(h|1))$. This is exactly the formula on Kline and Walters (2016, pp.1811): as they point out, the LATE for Head Start is a weighted average of "subLATEs" with weights determined by the proportion of C_{ch} among compliers, which is identified from the data[25].

Kline and Walters (2016) first tried to identify $\mathbb{E}[Y_i(h) - Y_i(c)|i \in C_{ch}]$ and $\mathbb{E}[Y_i(h) - Y_i(n)|i \in C_{nh}]$ separately using interactions of the instrument with covariates or experimental sites. They acknowledged the limitations of this approach and resorted to a parametric selection model à la Heckman (1979) instead. They report[26] estimates of the local average treatment effects of 0.370 for C_{nh} and -0.093 for C_{ch}, with respective standard errors 0.088 and 0.154. The resulting point estimate of the difference is quite large, at 0.463 standard deviation.

Our Corollary 2 provides an alternative approach to separating the two treatment effects.

[25] Our α_0 is denoted S_c in their paper.
[26] See Kline and Walters (2016, Table VIII, column (4), full model).

Table 5: Proportions, Counterfactual Means and Treatment Effects by Response Groups

	3-year-olds	4-year-olds	Pooled
Panel A. Proportions of Response Groups via Proposition 5			
Compliers from n to h (C_{nh})	0.505	0.393	0.454
Compliers from c to h (C_{ch})	0.198	0.272	0.232
Panel B. Counterfactual Means of Test Scores via Proposition 6			
$\mathbb{E}[Y_i(n)\|i \in C_{nh}]$	-0.027	-0.116	-0.062
$\mathbb{E}[Y_i(c)\|i \in C_{ch}]$	0.112	0.144	0.129
Panel C. Treatment Effects via Corollary 2			
Upper Bound on $\mathbb{E}[Y_i(h) - Y_i(n)\|i \in C_{nh}]$	0.279	0.285	0.278
	(0.063)	(0.076)	(0.050)
Lower Bound on $\mathbb{E}[Y_i(h) - Y_i(c)\|i \in C_{ch}]$	0.140	0.025	0.087
	(0.089)	(0.097)	(0.063)
Upper Bound on	0.139	0.260	0.191
$\mathbb{E}[Y_i(h) - Y_i(n)\|i \in C_{nh}] - \mathbb{E}[Y_i(h) - Y_i(c)\|i \in C_{ch}]$	(0.098)	(0.115)	(0.071)

Notes: Head Start (h), other centers (c), no preschool (n). Standard errors in parentheses are clustered at the Head Start center level.

Given that compliers coming from other preschools (C_{ch}) had better test scores than compliers not originally in preschools (C_{nh}), it seems reasonable to assume that they also have better test scores under Head Start:

$$\mathbb{E}\left[Y_i(h)|i \in C_{ch}\right] \geqslant \mathbb{E}\left[Y_i(h)|i \in C_{nh}\right]. \tag{4.1}$$

This is a "positive selection" that fits within the framework of Corollary 2. It can be derived in a simple model in which preschools, and especially Head Start, improve the outcomes of some students; and students choose schools as a function of their expected outcome. We show in Appendix B that this model generates positive selection under reasonable assumptions. The pooled cohort estimates in Panel C of Table 5 indicate that the upper bound on $\mathbb{E}[Y_i(h) - Y_i(n)|i \in C_{nh}]$ is 0.28 and the lower bound on $\mathbb{E}[Y_i(h) - Y_i(n)|i \in C_{ch}]$ is 0.09. The difference between these two numbers gives an upper bound of 0.19 for the difference of these two LATEs, with a standard error of 0.07. The testable prediction (3.7) implied by positive selection translates here into the non-negativity of the upper bound; we cannot reject it at any reasonable level. Conversely, negative selection (reverting the inequality (4.1)) would make 0.19 a *lower* bound for the difference of the LATEs. At the same time, it would imply that the lower bound is negative; this is soundly rejected by the data.

Our upper bound of 0.19 is much lower than the point estimate reported by Kline and Walters (2016). In fact, our 95% and 99% one-sided confidence intervals for

$$\mathbb{E}[Y_i(h) - Y_i(n)|i \in C_{nh}] - \mathbb{E}[Y_i(h) - Y_i(c)|i \in C_{ch}]$$

are $(-\infty, 0.308)$ and $(-\infty, 0.356)$. We conclude that the 0.463 estimate in Kline and Walters (2016) may overstate the difference between the two complier groups: it can only be rationalized under negative selection, which is a much less plausible assumption.

4.3 Expanding Access to Head Start

Kline and Walters (2016) sought to evaluate the welfare effect of increasing the number of slots in Head Start, as summarized by the marginal value of public funds (MVPF). They note that any expansion of Head Start will vacate some slots at competing preschools, which are oversubscribed. The relaxation of this rationing must be counted as an effect of Head Start expansions. This is what they call "rationed substitutes"[27].

The children who move from $T_i = n$ to $T_i = c$ when a slot is vacated by a child who moves to Head Start consitute a C_{nc} group that is ruled out by the 2×3 model. These children increase their grades by $Y_i(c) - Y_i(n)$, whose average generates a LATE that we denote LATE_{nc}. Equation (9) in Kline and Walters (2016, p. 1816) shows that the value of LATE_{nc} is a crucial input in the computation of the MVPF of a Head Start expansion. Identifying it requires either data on offers to all preschools, which Kline and Walters (2016) do not have[28], or additional modeling assumptions. They used their parametric selection model to construct an estimate for LATE_{nc}. Their estimate of $\text{LATE}_{nc} = 0.294$ results in a high MVPF estimate of 2.02 (see Table IX in their paper).

We take a different approach by embedding the 2×3 model within a 3×3 model. In this richer model, the instrument can take three values: in addition to the control group ($z = 0$) and those offered admission to Head Start ($z = 1$), we have a new group that we denote $z = 2$. This group is only offered admission to competing preschools because some seats were left free by students who moved to Head Start (the C_{ch} group of the binary model). Note that this maintains strict, one-to-one targeting.

Figure 7 shows the resulting treatment assignment, using tildes to denote the complier

[27] See Kline and Walters (2016, Sections V.D and IX.A) for details.

[28] See footnote 19 in their paper.

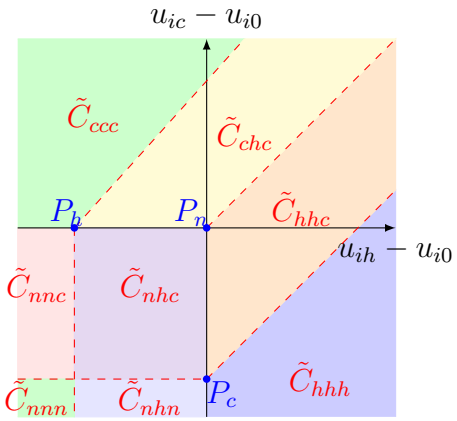
groups of the 3×3 model[29]. Using this notation, LATE_{nc} can be written as

$$\text{LATE}_{nc} = \mathbb{E}[Y_i(c) - Y_i(n) | i \in \tilde{C}_{n*c}],$$

where $\tilde{C}_{n*c} = \tilde{C}_{nnc} \bigcup \tilde{C}_{nhc}$ is the composite group of $n \to c$ compliers. Comparing Figure 7 with Figure 6 shows that the other complier groups of the two models are linked by

$$C_{nh} = \tilde{C}_{nh*} = \tilde{C}_{nhn} \bigcup \tilde{C}_{nhc}$$
$$C_{ch} = \tilde{C}_{ch*} = \tilde{C}_{chc}.$$

Figure 7: Embedding Preschool Choice in a 3×3 Model



Now consider the new group of $n \to c$ compliers. It differs from $\tilde{C}_{chc}$ in that its members will not go to a preschool unless they are offered a slot. We show in Appendix B that the structural model that we used in the binary instrument case predicts the following inequality:

$$\mathbb{E}[Y_i(c) | i \in \tilde{C}_{n*c}] \leqslant \mathbb{E}[Y_i(c) | i \in \tilde{C}_{chc}]. \tag{4.2}$$

Now consider the composite response-groups $\tilde{C}_{n*c} = \tilde{C}_{nnc} \bigcup \tilde{C}_{nhc}$ and $\tilde{C}_{nn*} = \tilde{C}_{nnc} \bigcup \tilde{C}_{nnn}$. As Figure 7 shows, they only differ by the substitution of $\tilde{C}_{nnn}$ for $\tilde{C}_{nhc}$. The former never go to Head Start or to another preschool, while the latter are full compliers. Our structural model generates the inequality

$$\mathbb{E}[Y_i(n) | i \in \tilde{C}_{nnn}] \leqslant \mathbb{E}[Y_i(n) | i \in \tilde{C}_{nhc}]$$

[29] Again, it is just Figure 4 with different notation.

which implies

$$\mathbb{E}[Y_i(n)|i \in \tilde{C}_{nn*}] \leqslant \mathbb{E}[Y_i(n)|i \in \tilde{C}_{n*c}]. \tag{4.3}$$

Inequalities (4.2) and (4.3) again are "positive selection" assumptions that fall under our Corollary 3.

Since $\tilde{C}_{nn*}$ coincides with C_{nn} and $\tilde{C}_{chc}$ is C_{ch}, we already know the values of the right-hand sides of both inequalities. Applying the same logic as in Corollary 3 gives us an upper bound for LATE_{nc}:

$$\text{LATE}_{nc} \leqslant \mathbb{E}[Y_i(c)|i \in \tilde{C}_{chc}] - \mathbb{E}[Y_i(n)|i \in \tilde{C}_{nn*}] = \mathbb{E}[Y_i(c)|i \in C_{ch}] - \mathbb{E}[Y_i(n)|i \in C_{nn}].$$

As the MVPF is an increasing function of LATE_{nc}, this gives us in turn an upper bound on its value [30]. We obtain $\text{LATE}_{nc} \leqslant 0.164$ and $\text{MVPF} \leqslant 1.55$. These upper bounds are noticeably smaller than the point estimates that result from the parametric selection model of Kline and Walters (2016).

Concluding Remarks

We have shown that the idea of targeting is a useful way to analyze models with multivalued treatments and multivalued instruments. Our paper only analyzed discrete-valued instruments and treatments. Some of the notions we used would extend naturally to continuous instruments and treatments: the definitions of targeting, one-to-one targeting, and positive selection would translate directly. Strict targeting, on the other hand, is less appealing in a context in which continuous values may denote intensities. Our earlier paper (Lee and Salanié, 2018) as well as Mountjoy's (2022) can be seen as analyzing continuous-instruments/discrete-treatments models. Extending our analysis to models with continuous treatments is an obvious topic for further research. It would also be interesting to apply the partial identification approach of Mogstad, Santos, and Torgovitsky (2018) in our setting. Finally, there has been a surge of recent interest on understanding the properties of OLS and 2SLS estimands when treatment effects vary with the covariates (Blandhol, Bonney, Mogstad, and Torgovitsky, 2022; Słoczyński, 2022; Goldsmith-Pinkham, Hull, and Kolesár, 2022). We believe that the targeting concept and the identifying assumptions explored in this paper should be relevant in this context and that they merit further investigation.

[30] Online Appendix F derives the formula for the MVPF in this model.

A Proofs for Section 3

Proof of Proposition 1. Suppose that for some $t \in \mathcal{T}^*$, $T_i(Z^*(t)) = 0$. Then $u_{i0} > \bar{U}(t) + u_{it}$. However, if $z \neq Z^*(t)$ then $\bar{U}(t) > U_z(t)$ under Assumption 4. Therefore $u_{i0} > U_z(t) + u_{it}$, and $T_i(z)$ cannot be t. □

Proof of Proposition 2. Recall that $T_i(z)$ maximizes $(U_z(t) + u_{it})$ over $t \in \mathcal{T}$. Under strict targeting, $U_z(t)$ is $\bar{U}(t)$ if $t \in \mathcal{T}^*(z)$ and $\underline{U}(t)$ otherwise.

Proof of (i): Since $\bar{U}(t) > \underline{U}(t)$ if $t \in \mathcal{T}^*$, we have

$$V_i^*(z) > \max_{t \in \mathcal{T}^*(z)} (\underline{U}(t) + u_{it}).$$

This implies that

$$\begin{aligned}
\max_{t \in \mathcal{T}} (U_z(t) + u_{it}) &= \max \left(\max_{t \in \mathcal{T}^*(z)} (\bar{U}(t) + u_{it}), \max_{t \notin \mathcal{T}^*(z)} (\underline{U}(t) + u_{it}) \right) \\
&= \max \left(V_i^*(z), \max_{t \in \mathcal{T}} (\underline{U}(t) + u_{it}) \right) \\
&= \max \left(V_i^*(z), \underline{V}_i \right).
\end{aligned}$$

Moreover, if $\underline{V}_i = \underline{U}_{\underline{t}_i} + u_{i,\underline{t}_i}$ is the maximum and $\underline{t}_i \in \mathcal{T}^*(z)$, then a fortiori $\underline{U}_{\underline{t}_i} + u_{i,\underline{t}_i} > \bar{U}_{\underline{t}_i} + u_{i,\underline{t}_i}$. This gives a contradiction since $\bar{U}_t > \underline{U}_t$ for all strictly targeted t.

Proof of (ii): If $z \notin \mathcal{Z}^*$, then $T^*(z)$ is empty and $V_i^*(z) = -\infty$. □

Proof of Corollary 1. It follows directly from Proposition 2. □

Proof of Proposition 3. Consider an observation i. The set A_i of Corollary 1 is a possibly empty subset of $\mathcal{T}^*$. The top alternative treatment $\underline{t}_i$ can be in A_i or in $\mathcal{T} \setminus \mathcal{T}^*$. If A_i has a elements, this allows for $a + |\mathcal{T}| - |\mathcal{T}^*| = a + q$ values of $\underline{t}_i$. Now every pair $(A_i, \underline{t}_i)$ fully defines a response-group. Since $|\mathcal{Z}^*| = |\mathcal{T}^*| = p$, this gives a total of

$$\sum_{a=0}^{p} \binom{p}{a} (a + q)$$

response-groups. Using the identities

$$\begin{aligned}
\sum_{a=0}^{b} \binom{b}{a} &= (1+1)^b = 2^b \\
\sum_{a=0}^{b} a \binom{b}{a} &= b \times \sum_{a=0}^{b-1} \binom{b-1}{a} = b \times 2^{b-1},
\end{aligned}$$

we obtain a total of $(p + 2q) \times 2^{p-1}$ types. □

Proof of Proposition 4. Take $z \in \mathcal{Z}$ and $t \in \mathcal{T}$. Consider any observation i and the corresponding $A_i \subset \mathcal{T}^*$ and $\underline{t}_i \in A_i \bigcup (\mathcal{T} \setminus \mathcal{T}^*)$. There are only two ways to obtain $T_i(z) = t$:

- if $z \notin A_i$, then $T_i(z) = \underline{t}_i$; therefore $\underline{t}_i = t$. Summing over all subsets A of $\mathcal{T}^*$ that exclude z gives the first term of (3.1).
- if $z \in A_i$ (which implies $z \in \mathcal{T}^*$), we know that $T_i(z) = z$ no matter what the value of $\underline{t}_i$ is; hence t must equal z. Summing over all subsets A that include z and all values of $\underline{t}_i \in A \bigcup (\mathcal{T} \setminus \mathcal{T}^*)$ gives the second line in (3.1).

By construction, each $C(A, t)$ completely defines the mapping from instrument values to treatment values; therefore each $C(A, t)$ is an elemental group. Their union is clearly the set of all observations. If $i \in C(A, t) \bigcup C(A', t')$, then $A' = A = A_i$ by the definition of A_i, and $t' = t = \underline{t}_i$. Therefore the $C(A, t)$ partition the set of observations. □

Proof of Lemma 1. For the sake of completeness, we provide the proof although the first part of the Lemma is the same as Theorem T-1 of Heckman and Pinto (2018) (applied with $\kappa(Y) := Y$). Let

$$E_z(t|C) \equiv \mathbb{E}(Y_i \mathbb{1}(T_i = t)|Z_i = z, i \in C).$$

We start from the sum over all response groups:

$$\bar{E}_z(t) = \sum_C E_z(t|C) \Pr(i \in C).$$

First note that if group C does not have treatment t under instrument z, it should not figure in the sum. Now if $C = C_R$ with $R(z) = t$, we have

$$\begin{aligned} E_z(t|C) &= \mathbb{E}(Y_i \mathbb{1}(T_i = t)|Z_i = z, i \in C) \\ &= \mathbb{E}(Y_i(t)|Z_i = z, i \in C) \\ &= \mathbb{E}(Y_i(t)|i \in C). \end{aligned}$$

The second part of the Lemma is just adding up. □

Proof of Proposition 5. The proof is in the text, with the exception of $\Pr(C_{11}) = P(1|0)$ which follows from the fact that the probabilities add up to 1. □

Proof of Proposition 6. Lemma 1 gives $2|\mathcal{T}|$ equations:

$$
\begin{aligned}
\bar{E}_0(1) &= \mathbb{E}(Y_i(1)|i \in C(\{1\},1)) \Pr(i \in C(\{1\},1)) = \mathbb{E}(Y_i(1)|i \in C_{11}) \Pr(C_{11}) \\
\text{for } t \neq 1,\ \bar{E}_0(t) &= \mathbb{E}(Y_i(t)|i \in C(\varnothing,t)) \Pr(i \in C(\varnothing,t)) \\
&\quad + \mathbb{E}(Y_i(t)|i \in C(\{1\},t)) \Pr(i \in C(\{1\},t)) \\
&= \mathbb{E}(Y_i(t)|i \in C_{tt}) \Pr(C_{tt}) + \mathbb{E}(Y_i(t)|i \in C_{t1}) \Pr(C_{t1}) \\
\bar{E}_1(1) &= \sum_{\tau \in \mathcal{T}} \mathbb{E}(Y_i(1)|i \in C(\{1\},\tau)) \Pr(i \in C(\{1\},\tau)) \\
&= \sum_{\tau \in \mathcal{T}} \mathbb{E}(Y_i(1)|i \in C_{\tau 1}) \Pr(C_{\tau 1}) \\
\text{for } t \neq 1,\ \bar{E}_1(t) &= \mathbb{E}(Y_i(t)|i \in C(\varnothing,t)) \Pr(i \in C(\varnothing,t)) = \mathbb{E}(Y_i(t)|i \in C_{tt}) \Pr(C_{tt}).
\end{aligned}
\tag{A.1}
$$

Since Proposition 5 identifies all type probabilities, the first and fourth equations in (A.1) give directly $\mathbb{E}(Y_i(t)|i \in C_{tt})$ for all t. Then the second equation identifies $\mathbb{E}(Y_i(t)|i \in C_{t1})$ for $t \neq 1$.

By subtraction, we obtain

$$
\begin{aligned}
&(\bar{E}_1(1) - \bar{E}_0(1)) - \sum_{t \neq 1} (\bar{E}_0(t) - \bar{E}_1(t)) \\
&= \sum_{t \neq 1} \mathbb{E}\left[Y_i(1) - Y_i(t)|i \in C_{t1}\right] \Pr(i \in C_{t1}).
\end{aligned}
$$

Combining these results with Proposition 5 and Lemma 1 yields the formula in the Proposition. The denominator

$$
\sum_{t \neq 1} (P(t|0) - P(t|1)) = P(1|1) - P(1|0)
$$

is positive, since all terms in the sum are positive. It follows that all α_t weights are positive and sum to 1. □

Proof of Corollary 2. Recall from (A.1) that when $T = 3$,

$$
\bar{E}_1(1) - \bar{E}_0(1) = \mathbb{E}\left[Y_i(1)|i \in C_{01}\right] \Pr(i \in C_{01}) + \mathbb{E}\left[Y_i(1)|i \in C_{21}\right] \Pr(i \in C_{21}).
$$

Hence, under (3.5) we have

$$\begin{aligned}
&\mathbb{E}\left[Y_i(1)|i \in C_{01}\right]\left\{\Pr(i \in C_{01}) + \Pr(i \in C_{21})\right\} \\
&\leqslant \bar{E}_1(1) - \bar{E}_0(1) \\
&\leqslant \mathbb{E}\left[Y_i(1)|i \in C_{21}\right]\left\{\Pr(i \in C_{01}) + \Pr(i \in C_{21})\right\}.
\end{aligned}$$

The first conclusion of the corollary follows immediately, as

$$\Pr(i \in C_{01}) + \Pr(i \in C_{21}) = P(1|1) - P(1|0).$$

The testable prediction is a direct consequence of this chain of inequalities. □

Proof of Proposition 7. It is straightforward from Figure 4 and Table 4. □

Proof of Proposition 8. By Lemma 1, we obtain

$$\begin{aligned}
\bar{E}_1(0) &= \mathbb{E}\left[Y_i(0)|i \in C_{000} \bigcup C_{002}\right] \Pr(i \in C_{000} \bigcup C_{002}), \\
\bar{E}_2(0) &= \mathbb{E}\left[Y_i(0)|i \in C_{000} \bigcup C_{010}\right] \Pr(i \in C_{000} \bigcup C_{010}), \\
\bar{E}_2(1) &= \mathbb{E}\left[Y_i(1)|i \in C_{111}\right] \Pr(i \in C_{111}), \\
\bar{E}_1(2) &= \mathbb{E}\left[Y_i(2)|i \in C_{222}\right] \Pr(i \in C_{222}), \\
\bar{E}_0(0) - \bar{E}_1(0) &= \mathbb{E}\left[Y_i(0)|i \in C_{010} \bigcup C_{012}\right] \Pr(i \in C_{010} \bigcup C_{012}), \\
\bar{E}_0(0) - \bar{E}_2(0) &= \mathbb{E}\left[Y_i(0)|i \in C_{002} \bigcup C_{012}\right] \Pr(i \in C_{002} \bigcup C_{012}), \\
\bar{E}_1(1) - \bar{E}_0(1) &= \mathbb{E}\left[Y_i(1)|i \in C_{010} \bigcup C_{012} \bigcup C_{212}\right] \Pr(i \in C_{010} \bigcup C_{012} \bigcup C_{212}), \\
\bar{E}_0(1) - \bar{E}_2(1) &= \mathbb{E}\left[Y_i(1)|i \in C_{112}\right] \Pr(i \in C_{112}), \\
\bar{E}_2(2) - \bar{E}_0(2) &= \mathbb{E}\left[Y_i(2)|i \in C_{002} \bigcup C_{012} \bigcup C_{112}\right] \Pr(i \in C_{002} \bigcup C_{012} \bigcup C_{112}), \\
\bar{E}_0(2) - \bar{E}_1(2) &= \mathbb{E}\left[Y_i(2)|i \in C_{212}\right] \Pr(i \in C_{212}).
\end{aligned}$$

Then, the results follows from the fact that all group probabilities are identified. □

Proof of Corollary 3. First note that $C_{01*} = C_{010} \bigcup C_{012}$. Under (3.8), we have

$$\begin{aligned}
\mathbb{E}(Y_i(1)\mathbb{1}(i \in C_{010} \bigcup C_{012})) &= \mathbb{E}(Y_i(1)\mathbb{1}(i \in C_{010} \bigcup C_{012} \bigcup C_{212})) - \mathbb{E}(Y_i(1)\mathbb{1}(i \in C_{212})) \\
&= \mathbb{E}(Y_i(1)\mathbb{1}(i \in C_{010} \bigcup C_{012} \bigcup C_{212})) - \mathbb{E}(Y_i(1)|i \in C_{212}) \Pr(i \in C_{212}) \\
&\geqslant \mathbb{E}(Y_i(1)|i \in C_{010} \bigcup C_{012} \bigcup C_{212}) \times \Pr(i \in C_{010} \bigcup C_{012} \bigcup C_{212}) \\
&\quad - \mathbb{E}(Y_i(1)|i \in C_{112}) \Pr(i \in C_{212}).
\end{aligned}$$

Replacing the probabilities and conditional expectations with their values from Proposition 7 and Proposition 8, we obtain $\Pr(i \in C_{010} \bigcup C_{012}) = P(0|0) - P(0|1)$ and

$$\mathbb{E}(Y_i(1)\mathbb{1}(i \in C_{010} \bigcup C_{012})) \geqslant \bar{E}_1(1) - \bar{E}_0(1) - \frac{\bar{E}_0(1) - \bar{E}_2(1)}{P(1|0) - P(1|2)}(P(2|0) - P(2|1)).$$

Finally, writing

$$\mathbb{E}(Y_i(1) - Y_i(0)|i \in C_{010} \bigcup C_{012}) = \frac{\mathbb{E}(Y_i(1)\mathbb{1}(i \in C_{010} \cup C_{012}))}{\Pr(i \in C_{010} \cup C_{012})} - \frac{\bar{E}_0(0) - \bar{E}_1(0)}{P(0|0) - P(0|1)}$$

gives the result.

The proof under (3.9) is similar: we start from $C_{0*2} = C_{002} \bigcup C_{012}$. Under (3.8), we have

$$\begin{aligned}
\mathbb{E}(Y_i(2)\mathbb{1}(i \in C_{002} \bigcup C_{012})) &= \mathbb{E}(Y_i(2)\mathbb{1}(i \in C_{002} \bigcup C_{012} \bigcup C_{112})) - \mathbb{E}(Y_i(2)\mathbb{1}(i \in C_{112})) \\
&= \mathbb{E}(Y_i(2)\mathbb{1}(i \in C_{002} \bigcup C_{012} \bigcup C_{112})) - \mathbb{E}(Y_i(2)|i \in C_{112}) \Pr(i \in C_{112}) \\
&\geqslant \mathbb{E}(Y_i(2)|i \in C_{002} \bigcup C_{012} \bigcup C_{112}) \times \Pr(i \in C_{002} \bigcup C_{012} \bigcup C_{112}) \\
&\quad - \mathbb{E}(Y_i(2)|i \in C_{212}) \Pr(i \in C_{112}).
\end{aligned}$$

Replacing the probabilities and conditional expectations with their values from Proposition 7 and Proposition 8, we obtain $\Pr(i \in C_{002} \bigcup C_{012}) = P(0|0) - P(0|2)$ and

$$\mathbb{E}(Y_i(2)\mathbb{1}(i \in C_{002} \bigcup C_{012})) \geqslant \bar{E}_2(2) - \bar{E}_0(2) - \frac{\bar{E}_0(2) - \bar{E}_1(2)}{P(2|0) - P(2|1)}(P(1|0) - P(1|2)).$$

Finally, writing

$$\mathbb{E}(Y_i(2) - Y_i(0)|i \in C_{002} \bigcup C_{012}) = \frac{\mathbb{E}(Y_i(2)\mathbb{1}(i \in C_{002} \cup C_{012}))}{\Pr(i \in C_{002} \cup C_{012})} - \frac{\bar{E}_0(0) - \bar{E}_2(0)}{P(0|0) - P(0|2)}$$

gives the result. □

B Positive Selection in Head Start

Let realized grades be given by

$$Y_i(t) = k_t + m_i p_t + \zeta_{it},$$

where $m_i > 0$

$$p_h > p_c > p_n = 0.$$

These conditions imply that children with a larger m_i benefit more from preschool, especially from Head Start; m_i does not play a role if i goes to neither type of preschool. The ζ_{it} shocks are zero mean and idiosyncratic; we suppose that each subject i expects $\mathbb{E}_i(Y_i(t)) = k_t + m_i p_t$. Preference shocks depend positively on expected grades:

$$u_{it} = a_i + b_i \mathbb{E}_i(Y_i(t)) + \varepsilon_{it} = a_i + b_i k_t + b_i m_i p_t + \varepsilon_{it},$$

with $b_i > 0$.

Let us define $v_{it} = u_{it} - u_{in}$; $\eta_{it} = \varepsilon_{it} - \varepsilon_{in}$; and $d_t = k_t - k_n$ for $t = c, h$. With this specification, we have

$$v_{it} = b_i d_t + b_i m_i p_t + \eta_{it}.$$

We assume that b_i, m_i, and the random vectors (η_{ic}, η_{ih}) and $(\zeta_{in}, \zeta_{ic}, \zeta_{ih})$ are mutually independent.

We will use the following lemma:

Lemma 2. *Let $A(\eta_{ic}, \eta_{in}, b_i)$ and $B(\eta_{ic}, \eta_{in}, b_i)$ be random subsets of $\mathbb{R}$ such that*

$$\sup A(\eta_{ic}, \eta_{in}, b_i) \leqslant \inf B(\eta_{ic}, \eta_{in}, b_i)$$

with probability one. Then for $t = c, h$,

$$\mathbb{E}(Y_i(t) \mid m_i \in A(\eta_{ic}, \eta_{in}, b_i)) \leqslant \mathbb{E}(Y_i(t) | m_i \in B(\eta_{ic}, \eta_{in}, b_i)).$$

Proof of Lemma 2. Take $t \in \{c, h\}$. Since $\mathbb{E}(Y_i(t)|m_i = m) = k_t + m p_t$, it is an increasing function of m. Fix $(\eta_{ic}, \eta_{ih}, b_i)$; obviously, the distribution of m_i conditional on $m_i \in B(\eta_{ic}, \eta_{in}, b_i)$ first-order stochastically dominates that of m_i conditional on $m_i \in A(\eta_{ic}, \eta_{in}, b_i)$. Therefore

$$\mathbb{E}(Y_i(t) \mid m_i \in A(\eta_{ic}, \eta_{in}, b_i), \eta_{ic}, \eta_{ih}, b_i) \leqslant \mathbb{E}(Y_i(t) \mid m_i \in B(\eta_{ic}, \eta_{in}, b_i), \eta_{ic}, \eta_{ih}, b_i).$$

Taking the expectation over $(\eta_{ic}, \eta_{ih}, b_i)$ completes the proof. □

B.1 The Binary Instrument Case

In Section 4.2, subjects who are assigned $z = 1$ receive a Head Start offer; those with $z = 0$ do not. The complier group C_{ch} has

$$\begin{aligned} \underline{U}_c + v_{ic} &\geqslant \max(0, \underline{U}_h + v_{ih}), \\ \bar{U}_h + v_{ih} &\geqslant \max(0, \underline{U}_c + v_{ic}). \end{aligned}$$

and the complier group C_{nh} has

$$\begin{aligned} 0 &\geqslant \max(\underline{U}_c + v_{ic}, \underline{U}_h + v_{ih}), \\ \bar{U}_h + v_{ih} &\geqslant \max(0, \underline{U}_c + v_{ic}). \end{aligned}$$

Note that $v_{ic} \geqslant -\underline{U}_c$ in C_{ch} and $v_{ic} \leqslant -\underline{U}_c$ in C_{nh}. Since $v_{ic} = b_i d_c + b_i m_i p_c + \eta_{ic}$ and $p_c b_i > 0$, it follows that for given $(\eta_{ic}, \eta_{ih}, b_i)$, $m_i \geqslant m_j$ for any $i \in C_{ch}$ and $j \in C_{nh}$. Therefore we can apply Lemma 2 with $t = h$ to obtain

$$\mathbb{E}(Y_i(h)|i \in C_{ch}) \geqslant \mathbb{E}(Y_i(h)|i \in C_{nh}),$$

which is our version of positive selection in the binary case.

B.2 The Ternary Instrument Case

In our setup in Section 4.3, subjects who are assigned $z = 1$ receive a Head Start offer, and those who are assigned $z = 2$ are offered admission in another preschool; those with $z = 0$ receive neither.

First note that under our assumptions,

$$\mathbb{E}(Y_i(n) \mid \eta_{ic}, \eta_{ih}, b_i) = k_n$$

is constant. Therefore, trivially,

$$\mathbb{E}(Y_i(n)|i \in \tilde{C}_{nhc}) \geqslant \mathbb{E}(Y_i(n)|i \in \tilde{C}_{nnn}).$$

Now let us consider the response-groups $\tilde{C}_{n*c}$ and $\tilde{C}_{chc}$. $\tilde{C}_{n*c}$ is defined by the inequalities

$$\bar{U}_c + v_{ic} > 0 > \max(\underline{U}_h + v_{ih}, \underline{U}_c + v_{ic}); \tag{B.1}$$

and $\tilde{C}_{chc}$ is defined by the inequalities

$$\bar{U}_h + v_{ih} > \underline{U}_c + v_{ic} > \max(\underline{U}_h + v_{ih}, 0). \tag{B.2}$$

(B.1) implies that $v_{ic} = b_i d_c + b_i m_i p_c + \eta_{ic} < -\underline{U}_c$, while (B.2) implies the reverse inequality. Here also, applying Lemma 2 with $t = c$ directly gives the conclusion:

$$\mathbb{E}(Y_i(c)|i \in \tilde{C}_{n*c}) \leqslant \mathbb{E}(Y_i(n)|i \in \tilde{C}_{chc}).$$

References

Angrist, J., D. Lang, and P. Oreopoulos (2009): "Incentives and Services for College Achievement: Evidence from a Randomized Trial," *American Economic Journal: Applied Economics*, 1(1), 136–63.

Angrist, J. D., and G. W. Imbens (1995): "Two-stage least squares estimation of average causal effects in models with variable treatment intensity," *Journal of the American Statistical Association*, 90(430), 431–442.

Ao, W., S. Calonico, and Y.-Y. Lee (2021): "Multivalued Treatments and Decomposition Analysis: An Application to the WIA Program," *Journal of Business & Economic Statistics*, 39(1), 358–371.

Attanasio, O., S. Cattan, E. Fitzsimons, C. Meghir, and M. Rubio-Codina (2020): "Estimating the Production Function for Human Capital: Results from a Randomized Controlled Trial in Colombia," *American Economic Review*, 110, 48–85.

Attanasio, O., C. Fernández, E. Fitzsimons, S. Grantham-McGregor, C. Meghir, and M. Rubio-Codina (2014): "Using the Infrastructure of a Conditional Cash Transfer Programme to Deliver a Scalable Integrated Early Child Development Programme in Colombia: A Cluster Randomised Controlled Trial," *British Medical Journal*, 349, g5785.

Bai, Y., S. Huang, S. Moon, A. M. Shaikh, and E. J. Vytlacil (2024): "On the Identifying Power of Monotonicity for Average Treatment Effects," arXiv 2405.14104.

Bai, Y., and M. Tabord-Meehan (2024): "Sharp Testable Implications of Encouragement Designs," arXiv:2411.09808, `https://arxiv.org/abs/2411.09808`.

Balke, A., and J. Pearl (1997): "Bounds on treatment effects from studies with imperfect compliance," *JASA*, 92, 1171–1176.

Bhagwati, J. (1971): "The generalized theory of distortions and welfare," in *Trade, balance of payments and growth*, ed. by J. Bhagwati, R. Jones, R. Mundell, and J. Vanek. North-Holland, Amsterdam.

Bhuller, M., and H. Sigstad (2024): "2SLS with multiple treatments," *Journal of Econometrics*, 242(1), 1057–85.

Blandhol, C., J. Bonney, M. Mogstad, and A. Torgovitsky (2022): "When is TSLS Actually LATE?," Working Paper 29709, National Bureau of Economic Research.

Buchinsky, M., P. Gertler, and R. Pinto (2023): "The Economics of Monotonicity Conditions: Exploring Choice Incentives in IV Models," UCLA, mimeo, `https://www.rodrigopinto.net/`.

Caetano, C., and J. C. Escanciano (2021): "Identifying Multiple Marginal Effects with a Single Instrument," *Econometric Theory*, 37(3), 464–494.

Carneiro, P., J. J. Heckman, and E. J. Vytlacil (2011): "Estimating Marginal Returns to Education," *American Economic Review*, 101(6), 2754–81.

Cattaneo, M. D. (2010): "Efficient semiparametric estimation of multi-valued treatment effects under ignorability," *Journal of Econometrics*, 155(2), 138–154.

D'Haultfoeuille, X., and P. Février (2015): "Identification of Nonseparable Triangular Models With Discrete Instruments," *Econometrica*, 83(3), 1199–1210.

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

Feng, J. (2024): "Matching points: Supplementing instruments with covariates in triangular models," *Journal of Econometrics*, 238(1), 105579.

Goff, L. (2024a): "A vector monotonicity assumption for multiple instruments," *Journal of Econometrics*, 241(1), 105735.

Goff, L. (2024b): "When is IV Identification Agnostic about Outcomes?," arXiv:2406.02835, `https://arxiv.org/abs/2406.02835`.

Goldsmith-Pinkham, P., P. Hull, and M. Kolesár (2022): "Contamination Bias in Linear Regressions," Working Paper 30108, National Bureau of Economic Research.

Heckman, J., and R. Pinto (2018): "Unordered Monotonicity," *Econometrica*, 86(1), 1–35.

Heckman, J. J. (1979): "Sample Selection Bias as a Specification Error," *Econometrica*, 47(1), 153–161.

Heckman, J. J., S. Urzua, and E. Vytlacil (2006): "Understanding instrumental variables in models with essential heterogeneity," *Review of Economics and Statistics*, 88(3), 389–432.

——— (2008): "Instrumental variables in models with multiple outcomes: The general unordered case," *Annales d'économie et de statistique*, 91/92, 151–174.

Heckman, J. J., and E. Vytlacil (2001): "Policy-relevant treatment effects," *American Economic Review*, 91(2), 107–111.

——— (2005): "Structural Equations, Treatment Effects, and Econometric Policy Evaluation," *Econometrica*, 73(3), 669–738.

——— (2007a): "Econometric Evaluation of Social Programs, Part I: Causal Models, Structural Models and Econometric Policy Evaluation," in *Handbook of Econometrics*, ed. by J. J. Heckman, and E. Leamer, vol. 6B, chap. 70, pp. 4779–4874. Elsevier, Amsterdam.

——— (2007b): "Econometric Evaluation of Social Programs, Part II: Using the Marginal Treatment Effect to Organize Alternative Econometric Estimators to Evaluate Social Programs, and to Forecast their Effects in New Environments," in *Handbook of Econometrics*, ed. by J. J. Heckman, and E. Leamer, vol. 6B, chap. 71, pp. 4875–5143. Elsevier, Amsterdam.

Heinesen, E., C. Hvid, L. J. Kirkeboen, E. Leuven, and M. Mogstad (2022): "Instrumental Variables with Unordered Treatments: Theory and Evidence from Returns to Fields of Study," Working Paper 30574, National Bureau of Economic Research.

Huang, L., U. Khalil, and N. Yildiz (2019): "Identification and estimation of a triangular model with multiple endogenous variables and insufficiently many instrumental variables," *Journal of Econometrics*, 208(2), 346–366.

Imbens, G. W. (2000): "The role of the propensity score in estimating dose-response functions," *Biometrika*, 87(3), 706–710.

Imbens, G. W., and J. D. Angrist (1994): "Identification and Estimation of Local Average Treatment Effects," *Econometrica*, 62(2), 467–475.

Kamat, V. (2024): "Identifying the effects of a program offer with an application to Head Start," *Journal of Econometrics*, 240(1), 105679.

Kédagni, D., and I. Mourifié (2020): "Generalized instrumental inequalities: testing the instrumental variable independence assumption," *Biometrika*, 107(3), 661–675.

Kirkeboen, L. J., E. Leuven, and M. Mogstad (2016): "Field of study, earnings, and self-selection," *Quarterly Journal of Economics*, 131(3), 1057–1111.

Kitagawa, T. (2015): "A Test for Instrument Validity," *Econometrica*, 83(5), 2043–2063.

Kline, P., and C. R. Walters (2016): "Evaluating public programs with close substitutes: The case of Head Start," *Quarterly Journal of Economics*, 131(4), 1795–1848.

Lee, S., and B. Salanié (2018): "Identifying effects of multivalued treatments," *Econometrica*, 86(6), 1939–1963.

Mogstad, M., A. Santos, and A. Torgovitsky (2018): "Using Instrumental Variables for Inference About Policy Relevant Treatment Parameters," *Econometrica*, 86(5), 1589–1619.

Mogstad, M., A. Torgovitsky, and C. R. Walters (2020): "Policy Evaluation With Multiple Instrumental Variables," Working Paper 27546, National Bureau of Economic Research.

Mogstad, M., A. Torgovitsky, and C. R. Walters (2021): "The Causal Interpretation of Two-Stage Least Squares with Multiple Instrumental Variables," *American Economic Review*, 111(11), 3663–98.

Mountjoy, J. (2022): "Community Colleges and Upward Mobility," *American Economic Review*, 112, 2580–2630.

Mourifié, I., and Y. Wan (2017): "Testing Local Average Treatment Effect Assumptions," *Review of Economics and Statistics*, 99(2), 305–313.

Muralidharan, K., M. Romero, and K. Wüthrich (2023): "Factorial Designs, Model Selection, and (Incorrect) Inference in Randomized Experiments," *Review of Economics and Statistics*, pp. 1–44, Just Accepted.

Navjeevan, M., and R. Pinto (2022): "Ordered, Unordered and Minimal Monotonicity Criteria," UCLA, mimeo, `https://www.rodrigopinto.net/`.

Nibbering, D., M. Oosterveen, and P. L. Silva (2022): "Clustered local average treatment effects: fields of study and academic student progress.," Discussion Paper No. 15159.

Pinto, R. (2021): "Beyond Intention to Treat: Using the Incentives in Moving to Opportunity to Identify Neighborhood Effects," UCLA, mimeo, `https://www.rodrigopinto.net/`.

Słoczyński, T. (2022): "Interpreting OLS estimands when treatment effects are heterogeneous: Smaller groups get larger weights," *Review of Economics and Statistics*, 104, 1–9.

Sun, Z. (2023): "Instrument validity for heterogeneous causal effect," *Journal of Econometrics*, 237, forthcoming.

Tinbergen, J. (1952): *On the Theory of Economic Policy*. North Holland.

Torgovitsky, A. (2015): "Identification of Nonseparable Models Using Instruments with Small Support," *Econometrica*, 83(3), 1185–1197.

Vytlacil, E. (2002): "Independence, monotonicity, and latent index models: An equivalence result," *Econometrica*, 70(1), 331–341.

Online Appendices to "Treatment Effects with Targeting Instruments"

C Proofs for Section 3.4

Let us first translate Kirkeboen, Leuven, and Mogstad's (2016) assumptions in our notation to show that their assumptions are equivalent to strict one to one targeting.

KLM impose the following in their Assumption 4:

- if $T_i(0) = 1$ then $T_i(1) = 1$
- if $T_i(0) = 2$ then $T_i(2) = 2$.

This can be viewed as a monotonicity assumption. It excludes the twelve response groups $C_{10*}, C_{12*}, C_{2*0}$, and C_{2*1}.

Their Proposition 2 proves point-identification of response-groups when one of three alternative assumptions is added to their Assumption 4. We focus here on the irrelevance assumption in their Proposition 2 (iii), which is the weakest of the three and the one their application relies on. In our notation, it states that:

- if ($T_i(0) \neq 1$ and $T_i(1) \neq 1$), then ($T_i(0) = 2$ iff $T_i(1) = 2$)
- if ($T_i(0) \neq 2$ and $T_i(2) \neq 2$), then ($T_i(0) = 1$ iff $T_i(2) = 1$).

These complicated statements can be simplified. Take the first part. If both $T_i(0)$ and $T_i(1)$ are not 1, then they can only be 0 or 2. Therefore we are requiring $T_i(0) = T_i(1)$. Applying the same argument to the second part, the irrelevance assumption becomes:

- if ($T_i(0) \neq 1$ and $T_i(1) \neq 1$), then $T_i(0) = T_i(1)$
- if ($T_i(0) \neq 2$ and $T_i(2) \neq 2$), then $T_i(0) = T_i(2)$.

It therefore excludes the response-groups $C_{02*}, C_{20*}, C_{0*1}$, and C_{1*0}. The response-group C_{021} appears twice in this list; and four other response-groups were already ruled out by Assumption 4. The reader can easily check that the $3^3 - 12 - (11 - 4) = 8$ response-groups left are exactly the same as in our Figure 4.

Proof of Proposition 9. The moment conditions that define β_0, β_1 and β_2 are

$$\mathbb{E}\left[\left(Y_i - \beta_0 - \beta_1 \mathbb{1}(T_i = 1) - \beta_2 \mathbb{1}(T_i = 2)\right) \mathbb{1}(Z_i = z)\right] = 0 \tag{C.1}$$

for $z = 0, 1, 2$.

Using counterfactual notation, we write

$$Y_i = Y_i(0) + (Y_i(1) - Y_i(0))\mathbb{1}(T_i = 1) + (Y_i(1) - Y_i(0))\mathbb{1}(T_i = 2), \tag{C.2}$$

which allows us to write Equation (C.1) as

$$\mathbb{E}\left[(Y_i(0) - \beta_0 + b_i(1)\mathbb{1}(T_i = 1) + b_i(2)\mathbb{1}(T_i = 2))\,\mathbb{1}(Z_i = z)\right] = 0, \tag{C.3}$$

where $b_i(t) \equiv Y_i(t) - Y_i(0) - \beta_t$ for $t = 1, 2$.

Now since

$$\begin{aligned}\mathbb{1}(T_i = t) = \mathbb{1}(T_i(0) = t) &+ (\mathbb{1}(T_i(1) = t) - \mathbb{1}(T_i(0) = t))\mathbb{1}(Z_i = 1) \\ &+ (\mathbb{1}(T_i(2) = t) - \mathbb{1}(T_i(0) = t))\mathbb{1}(Z_i = 2),\end{aligned}$$

we can expand

$$\begin{aligned}&[Y_i(0) - \beta_0 + b_i(1)\mathbb{1}(T_i = 1) + b_i(2)\mathbb{1}(T_i = 2)] \times \mathbb{1}(Z_i = z) \\ &= [Y_i(0) - \beta_0 + b_i(1)\mathbb{1}(T_i(0) = 1) + b_i(2)\mathbb{1}(T_i(0) = 2) \\ &\quad + b_i(1)(\mathbb{1}(T_i(z) = 1) - \mathbb{1}(T_i(0) = 1)) + b_i(2)(\mathbb{1}(T_i(z) = 2) - \mathbb{1}(T_i(0) = 2))] \times \mathbb{1}(Z_i = z).\end{aligned}$$

Since Z_i is independent of $\{Y_i(t), T_i(z) : t, z = 0, 1, \ldots, T-1\}$, all of the terms that multiply $\mathbb{1}(Z_i = z)$ are independent of it. It follows that for $z = 0, 1, 2$,

$$\begin{aligned}&\mathbb{E}\left[Y_i(0) - \beta_0 + b_i(1)\mathbb{1}(T_i(0) = 1) + b_i(2)\mathbb{1}(T_i(0) = 2)\right. \\ &\left. \quad + b_i(1)(\mathbb{1}(T_i(z) = 1) - \mathbb{1}(T_i(0) = 1)) + b_i(2)(\mathbb{1}(T_i(z) = 2) - \mathbb{1}(T_i(0) = 2))\right] = 0.\end{aligned}$$

When $z = 0$, the second line is zero; therefore

$$\mathbb{E}\left(Y_i(0) - \beta_0 + b_i(1)\mathbb{1}(T_i(0) = 1) + b_i(2)\mathbb{1}(T_i(0) = 2)\right) = 0.$$

The other two equations become

$$\mathbb{E}\left(b_i(1)(\mathbb{1}(T_i(z) = 1) - \mathbb{1}(T_i(0) = 1)) + b_i(2)(\mathbb{1}(T_i(z) = 2) - \mathbb{1}(T_i(0) = 2))\right) = 0$$

for $z = 1, 2$. Remembering that $b_i(t) = Y_i(t) - Y_i(0) - \beta_t$ for $t = 1, 2$, we obtain

$$\mathbb{E}\left[(Y_i(1) - Y_i(0))(\mathbb{1}(T_i(z) = 1) - \mathbb{1}(T_i(0) = 1)) + (Y_i(2) - Y_i(0))(\mathbb{1}(T_i(z) = 2) - \mathbb{1}(T_i(0) = 2))\right]$$
$$= \beta_1 \mathbb{E}(\mathbb{1}(T_i(z) = 1) - \mathbb{1}(T_i(0) = 1)) + \beta_2 \mathbb{E}(\mathbb{1}(T_i(z) = 2) - \mathbb{1}(T_i(0) = 2)).$$

Proposition 9 follows after noting that given Table 4,

- the variable $\mathbb{1}(T_i(z) = 1) - \mathbb{1}(T_i(0) = 1)$ is $\mathbb{1}(i \in \mathcal{C}_1)$ for $z = 1$ and $-\mathbb{1}(i \in C_{112})$ for $z = 2$;
- the variable $\mathbb{1}(T_i(z) = 2) - \mathbb{1}(T_i(0) = 2)$ is $\mathbb{1}(i \in \mathcal{C}_2)$ for $z = 2$ and $-\mathbb{1}(i \in C_{212})$ for $z = 1$.

□

Proof of Corollary 4. Solving the system of equations in Proposition 9 gives, after elementary calculations,

$$\begin{aligned}
\beta_1 \mathcal{D} &= \Pr(i \in C_{212}) \left[\mathbb{E}\left((Y_i(2) - Y_i(0))\mathbb{1}(i \in \mathcal{C}_2)\right) - \mathbb{E}\left((Y_i(1) - Y_i(0))\mathbb{1}(i \in C_{112})\right)\right] \\
&\quad + \Pr(i \in \mathcal{C}_2) \left[\mathbb{E}\left((Y_i(1) - Y_i(0))\mathbb{1}(i \in \mathcal{C}_1)\right) - \mathbb{E}\left((Y_i(2) - Y_i(0))\mathbb{1}(i \in C_{212})\right)\right] \\
&= \Pr(i \in \mathcal{C}_1) \Pr(i \in \mathcal{C}_2) \mathbb{E}\left(Y_i(1) - Y_i(0) | i \in \mathcal{C}_1\right) - \Pr(i \in C_{112}) \Pr(i \in C_{212}) \mathbb{E}\left(Y_i(1) - Y_i(0) | i \in C_{112}\right) \\
&\quad + \Pr(i \in C_{212}) \Pr(i \in \mathcal{C}_2) \left[\mathbb{E}\left(Y_i(2) - Y_i(0) | i \in \mathcal{C}_2\right) - \mathbb{E}\left(Y_i(2) - Y_i(0) | i \in C_{212}\right)\right].
\end{aligned}$$

The difference of treatment effects in the last line is simply $\mathcal{D}_2$; note that it is multiplied by a non-negative term. Suppose for instance that $\mathcal{D}_1, \mathcal{D}_2 \geqslant 0$. Then

(C.4)
$$\begin{aligned}
&\beta_1 \mathcal{D} \\
&\geqslant \Pr(i \in \mathcal{C}_1) \Pr(i \in \mathcal{C}_2) \mathbb{E}\left(Y_i(1) - Y_i(0) | i \in \mathcal{C}_1\right) - \Pr(i \in C_{112}) \Pr(i \in C_{212}) \mathbb{E}\left(Y_i(1) - Y_i(0) | i \in C_{112}\right).
\end{aligned}$$

Moreover, it is easy to prove the following: define $r = (\alpha a - \beta b)/(a - b)$ with $a, b \geqslant 0$ and $a \neq b$. Then

1. if $(\alpha - \beta)$ and $(a - b)$ have the same sign, $r \geqslant \max(\alpha, \beta)$
2. if $(\alpha - \beta)$ and $(a - b)$ have different signs, $r \leqslant \min(\alpha, \beta)$.

Now take

$$
\begin{aligned}
a &= \Pr(i \in \mathcal{C}_1) \Pr(i \in \mathcal{C}_2) \\
b &= \Pr(i \in C_{112}) \Pr(i \in C_{212}) \\
\alpha &= \mathbb{E}\left(Y_i(1) - Y_i(0) | i \in \mathcal{C}_1\right) \\
\beta &= \mathbb{E}\left(Y_i(1) - Y_i(0) | i \in C_{112}\right).
\end{aligned}
$$

Note that a and b are non-negative, and $a - b = \mathcal{D} \neq 0$. Suppose that $\mathcal{D} > 0$ so that Equation (C.4) becomes $\beta_1 \geqslant r$. Since $\alpha - \beta = \mathcal{D}_1 \geqslant 0$, we can apply result 1 and we get

$$
\beta_1 \geqslant \max(\alpha, \beta) = \alpha = \mathbb{E}\left(Y_i(1) - Y_i(0) | i \in \mathcal{C}_1\right).
$$

If on the other hand $\mathcal{D}$ is negative, then we have $\beta_1 \leqslant r$ and since $\mathcal{D}$ and $\mathcal{D}_1$ have different signs result 2 gives

$$
\beta_1 \leqslant \min(\alpha, \beta) = \beta
$$

and a fortiori $\beta_1 \leqslant \alpha$.

Similar arguments apply to β_2, as well as to the the case when $\mathcal{D}_1$ and $\mathcal{D}_2$ are non-positive. □

D Revisiting the 2×3 and 3×3 Models via Heckman and Pinto (2018)

D.1 Notation

We first adapt Heckman and Pinto (2018, HP hereafter)'s notation to our framework. As in the main text, we focus on identifying the probabilities of the various response groups $\Pr(i \in C)$ and the group average outcomes $\mathbb{E}(Y_i(t) | i \in C)$. The following population quantities are directly identified from data for all treatment values t:

$$
\begin{aligned}
\boldsymbol{P}_Z(t) &= \left(P\left(T = t \mid Z = z\right)\right)_{z \in \mathcal{Z}}, \\
\boldsymbol{Q}_Z(t) &= \left(\mathbb{E}\left(Y \, \mathbf{1}(T = t) \mid Z = z\right)\right)_{z \in \mathcal{Z}}.
\end{aligned}
$$

We also define $\boldsymbol{P}_Z = \left(\boldsymbol{P}_Z\left(t\right)\right)_{t \in \mathcal{T}}$.

We choose an arbitrary ordering $(C^1, \ldots, C^S)$ of the S non-empty response groups and we define the S dummy variables $c_i^s = \mathbf{1}(i \in C^s)$. The *response vector* $\boldsymbol{S}$ is $\{c^1, \ldots, c^S\}$.

With this notation, our main objects of interest are

$$\begin{aligned}\boldsymbol{P}_S &= \mathbb{E}\boldsymbol{S} \\ \boldsymbol{Q}_S(t) &= \mathbb{E}\left(Y(t)\boldsymbol{S}\right) \text{ for } t \in \mathcal{T},\end{aligned}$$

from which we obtain $\Pr(i \in C^s) = \boldsymbol{P}_S^s$ and $\mathbb{E}(Y_i(t)|i \in C^s) = \boldsymbol{Q}_S^s(t)/\boldsymbol{P}_S^s$.

As in HP, $\boldsymbol{B}_t$ denotes a binary matrix with dimension $|\mathcal{Z}| \times S$ whose element in row z and column s equals 1 if response group C^s has $T_i = t$ when $Z_i = z$, and zero otherwise. Finally, let $\boldsymbol{\mathcal{B}}$ be the binary matrix of dimension $(|\mathcal{Z}| \cdot |\mathcal{T}|) \times S$ generated by stacking the matrices $\boldsymbol{B}_t$ vertically: $\boldsymbol{\mathcal{B}} = \left[\boldsymbol{B}_0', \ldots, \boldsymbol{B}_{|\mathcal{T}|-1}'\right]'$.

D.2 Theorem T-2 in HP

Let $\boldsymbol{M}^\dagger$ denote the Moore-Penrose pseudo-inverse of a matrix $\boldsymbol{M}$. We define

$$\boldsymbol{K}_t = \boldsymbol{I}_S - \boldsymbol{B}_t^\dagger \boldsymbol{B}_t \text{ and } \boldsymbol{\mathcal{K}} = \boldsymbol{I}_S - \boldsymbol{\mathcal{B}}^\dagger \boldsymbol{\mathcal{B}},$$

where $\boldsymbol{I}_S$ denotes the identity matrix of dimension S. Note that $\boldsymbol{\mathcal{K}}$ and $\boldsymbol{K}_t$ are orthogonal projection matrices in $\mathbb{R}^S$ that only depend on the binary matrices $\boldsymbol{\mathcal{B}}$ and $\boldsymbol{B}_t$. Theorem T-2 in HP shows that

$$\boldsymbol{P}_S = \boldsymbol{\mathcal{B}}^\dagger \boldsymbol{P}_Z + \boldsymbol{\mathcal{K}}\boldsymbol{\lambda}, \tag{D.1}$$

$$\boldsymbol{Q}_S(t) = \boldsymbol{B}_t^\dagger \boldsymbol{Q}_Z(t) + \boldsymbol{K}_t\tilde{\boldsymbol{\lambda}}, \tag{D.2}$$

where $\boldsymbol{\lambda}$ and $\tilde{\boldsymbol{\lambda}}$ are arbitrary S-dimensional vectors.

D.3 Identification in the 2×3 model

We can now re-derive our identification results for the 2 by 3 model using the theorems in HP. To do so, we order the response-types as $\{C_{00}, C_{11}, C_{22}, C_{01}, C_{21}\}$. Then the binary

matrices $\boldsymbol{B}_0, \boldsymbol{B}_1$, and $\boldsymbol{B}_2$ are

$$\boldsymbol{B}_0 = \begin{bmatrix} 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 \end{bmatrix},$$
$$\boldsymbol{B}_1 = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 \end{bmatrix},$$
$$\boldsymbol{B}_2 = \begin{bmatrix} 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix},$$

and

$$\boldsymbol{\mathcal{B}} = \begin{bmatrix} 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix}.$$

It is easy to see that $\boldsymbol{\mathcal{B}}$ has full column rank; it follows that $\boldsymbol{\mathcal{B}}^\dagger \boldsymbol{\mathcal{B}} = \boldsymbol{I}_6$ and $\boldsymbol{\mathcal{K}}$ is the 6 by 6 matrix with all elements zero. Therefore by Theorem T-2 in HP (see equation (D.1) above), $\boldsymbol{P}_S$ is point-identified as $\boldsymbol{P}_S = \boldsymbol{\mathcal{B}}^\dagger \boldsymbol{P}_Z$.

Since

$$\boldsymbol{\mathcal{B}}^\dagger = \frac{1}{6} \begin{bmatrix} 1 & 5 & 1 & -1 & 1 & -1 \\ -1 & 1 & 5 & 1 & -1 & 1 \\ 1 & -1 & 1 & -1 & 1 & 5 \\ 4 & -4 & -2 & 2 & -2 & 2 \\ -2 & 2 & -2 & 2 & 4 & -4 \end{bmatrix},$$

this is not very transparent, however. To derive our identification results, we use (D.2) instead. Note that the equation $\boldsymbol{Q}_S(t) = \boldsymbol{B}_t^\dagger \boldsymbol{Q}_Z(t) + \boldsymbol{K}_t \tilde{\boldsymbol{\lambda}}$ holds for any function of $Y(t)$. If we take it to be a constant function of $Y(t)$, we get $\boldsymbol{Q}_S(t) = \mathbb{E}\boldsymbol{S} = \boldsymbol{P}_S$ and $\boldsymbol{Q}_Z(t) = \boldsymbol{P}_Z(t)$, so that (D.2) boils down to

(D.3) $$\boldsymbol{P}_S = \boldsymbol{B}_t^\dagger \boldsymbol{P}_Z(t) + \boldsymbol{K}_t \tilde{\boldsymbol{\lambda}} \text{ for all values of } t.$$

Now

$$\boldsymbol{B}_0^\dagger = \begin{bmatrix} 0 & 1 \\ 0 & 0 \\ 0 & 0 \\ 1 & -1 \\ 0 & 0 \end{bmatrix} \Longrightarrow \boldsymbol{K}_0 = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix};$$

$$\boldsymbol{B}_1^\dagger = \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 0 \\ -1/2 & 1/2 \\ -1/2 & 1/2 \end{bmatrix} \Longrightarrow \boldsymbol{K}_1 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1/2 & -1/2 \\ 0 & 0 & 0 & -1/2 & 1/2 \end{bmatrix};$$

and

$$\boldsymbol{B}_2^\dagger = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 1 \\ 0 & 0 \\ 1 & -1 \end{bmatrix} \Longrightarrow \boldsymbol{K}_2 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

Let $(\boldsymbol{e}_s)_{s=1,\ldots,5}$ denote the standard basis vectors in $\mathbb{R}^5$. If $\boldsymbol{e}_s'\boldsymbol{K}_t = 0$, then (D.3) point-identifies $\Pr(i \in C^s) = \boldsymbol{e}_s'\boldsymbol{P}_S = \boldsymbol{e}_s'\boldsymbol{B}_t^\dagger \boldsymbol{P}_Z(t)$. Clearly,

$$\boldsymbol{e}_1'\boldsymbol{K}_0 = \boldsymbol{e}_4'\boldsymbol{K}_0 = \boldsymbol{e}_2'\boldsymbol{K}_1 = \boldsymbol{e}_3'\boldsymbol{K}_2 = \boldsymbol{e}_4'\boldsymbol{K}_2 = 0;$$

this reproduces our identification results for $\boldsymbol{P}_S$ in Proposition 5:

$$\boldsymbol{P}_S = \left(\boldsymbol{e}_1'\boldsymbol{B}_0^\dagger \boldsymbol{P}_Z(0), \boldsymbol{e}_2'\boldsymbol{B}_1^\dagger \boldsymbol{P}_Z(1), \boldsymbol{e}_3'\boldsymbol{B}_2^\dagger \boldsymbol{P}_Z(2), \boldsymbol{e}_4'\boldsymbol{B}_0^\dagger \boldsymbol{P}_Z(0), \boldsymbol{e}_5'\boldsymbol{B}_2^\dagger \boldsymbol{P}_Z(2)\right)'.$$

Returning to the counterfactual outcomes $Y(t)$, the same argument results in Proposition 6:

$$\begin{aligned}
\mathbb{E}\left(Y_i(0)\cdot\mathbf{1}\left[i\in C_{00}\right]\right) &= \boldsymbol{e}_1'\boldsymbol{B}_0^{\dagger}\boldsymbol{Q}_Z(0),\\
\mathbb{E}\left(Y_i(1)\cdot\mathbf{1}\left[i\in C_{11}\right]\right) &= \boldsymbol{e}_2'\boldsymbol{B}_1^{\dagger}\boldsymbol{Q}_Z(1),\\
\mathbb{E}\left(Y_i(2)\cdot\mathbf{1}\left[i\in C_{22}\right]\right) &= \boldsymbol{e}_3'\boldsymbol{B}_2^{\dagger}\boldsymbol{Q}_Z(2),\\
\mathbb{E}\left(Y_i(0)\cdot\mathbf{1}\left[i\in C_{01}\right]\right) &= \boldsymbol{e}_4'\boldsymbol{B}_0^{\dagger}\boldsymbol{Q}_Z(0),\\
\mathbb{E}\left(Y_i(2)\cdot\mathbf{1}\left[i\in C_{21}\right]\right) &= \boldsymbol{e}_5'\boldsymbol{B}_2^{\dagger}\boldsymbol{Q}_Z(2).
\end{aligned}$$

We conclude that while the first part of Theorem T-2 in HP (i.e., $\boldsymbol{P}_S = \boldsymbol{\mathcal{B}}^{\dagger}\boldsymbol{P}_Z + \boldsymbol{\mathcal{K}}\boldsymbol{\lambda}$) is useful to determine the degrees of identification by checking the rank of $\boldsymbol{\mathcal{B}}$, it does not yield the most constructive form of identification. To get the objects of interest, it is better to invoke the second part of Theorem T-2 (i.e., $\boldsymbol{Q}_S(t) = \boldsymbol{B}_t^{\dagger}\boldsymbol{Q}_Z(t) + \boldsymbol{K}_t\tilde{\boldsymbol{\lambda}}$). Note that since the 2×3 model satisfies the unordered monotonicity assumption, we could also obtain the same results using Theorem T-6 in HP.

D.4 Identification in the 3×3 model

We now turn to our 3 by 3 model. We sort the response-types as

$$\left\{C_{000}, C_{111}, C_{222}, C_{010}, C_{002}, C_{012}, C_{112}, C_{212}\right\}.$$

Now

$$\boldsymbol{B}_0 = \begin{bmatrix} 1 & 0 & 0 & 1 & 1 & 1 & 0 & 0\\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0\\ 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \end{bmatrix},$$

$$\boldsymbol{B}_1 = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0\\ 0 & 1 & 0 & 1 & 0 & 1 & 1 & 1\\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix},$$

$$\boldsymbol{B}_2 = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1\\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0\\ 0 & 0 & 1 & 0 & 1 & 1 & 1 & 1 \end{bmatrix}.$$

Note that

$$\boldsymbol{B}_0^\dagger = \begin{bmatrix} -0.25 & 0.5 & 0.5 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0.25 & -0.5 & 0.5 \\ 0.25 & 0.5 & -0.5 \\ 0.75 & -0.5 & -0.5 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \Longrightarrow \boldsymbol{K}_0 = \begin{bmatrix} 0.25 & 0 & 0 & -0.25 & -0.25 & 0.25 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ -0.25 & 0 & 0 & 0.25 & 0.25 & -0.25 & 0 & 0 \\ -0.25 & 0 & 0 & 0.25 & 0.25 & -0.25 & 0 & 0 \\ 0.25 & 0 & 0 & -0.25 & -0.25 & 0.25 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

Let $(\boldsymbol{e}_s)_{s=1,\ldots,8}$ denote the standard basis vectors in $\mathbb{R}^8$. Since $\boldsymbol{K}_0$ has no zero column, none of the $\boldsymbol{e}_s'\boldsymbol{K}_0$ is zero and the argument in Section D.3 show that no population share $\Pr(i \in C^s)$ is point-identified by $\boldsymbol{B}_0^\dagger \boldsymbol{P}_Z(0)$. On the other hand,

$$\boldsymbol{B}_1^\dagger = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1/3 & 1/3 & 0 \\ 0 & 0 & 0 \\ -1/3 & 1/3 & 0 \\ 1 & 0 & -1 \\ -1/3 & 1/3 & 0 \end{bmatrix} \Longrightarrow \boldsymbol{K}_1 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2/3 & 0 & -1/3 & 0 & -1/3 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1/3 & 0 & 2/3 & 0 & -1/3 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1/3 & 0 & -1/3 & 0 & 2/3 \end{bmatrix}$$

so that $\boldsymbol{e}_2'\boldsymbol{K}_1 = \boldsymbol{e}_7'\boldsymbol{K}_1 = 0$, which point-identifies the population shares of C_{111} and C_{112}. Similarly,

$$\boldsymbol{B}_2^\dagger = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \\ -1/3 & 0 & 1/3 \\ -1/3 & 0 & 1/3 \\ -1/3 & 0 & 1/3 \\ 1 & -1 & 0 \end{bmatrix} \Longrightarrow \boldsymbol{K}_2 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 2/3 & -1/3 & -1/3 & 0 \\ 0 & 0 & 0 & 0 & -1/3 & 2/3 & -1/3 & 0 \\ 0 & 0 & 0 & 0 & -1/3 & -1/3 & 2/3 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

and $\boldsymbol{e}_3'\boldsymbol{K}_2 = \boldsymbol{e}_8'\boldsymbol{K}_2 = 0$ so that the shares of C_{222} and C_{212} are point-identified. On the other hand, the shares of C_{010}, C_{002}, and C_{012} are not identified. The results in Propositions 7

and 8 follow.

Finally, note that $\boldsymbol{B}_0$ has the following 2×2 sub-matrix:

$$\begin{pmatrix} & [C_{010} & C_{002}] \\ [z=1] & 1 & 0 \\ [z=2] & 0 & 1 \end{pmatrix},$$

where we indicate the relevant columns and rows of matrix $\boldsymbol{B}_0$. Given this pattern, Theorem T-3 and Remark 6.3 in Heckman and Pinto (2018) imply that the unordered monotonicity assumption is not satisfied for the 3×3 model. As mentioned in the main text, this , switching from instrument value from 1 to 2 causes observations in C_{010} to move to treatment 0, while those in C_{002} move out of treatment 0. Recall that the ARUM structure rules out "direct two-way flows" (that is, instrument values 1 and 2, respectively, make treatments 1 and 2 more favorable for everyone). However, the 3×3 model allows for "indirect two-way flows", where treatment 0 is not targeted by either $z = 1$ or $z = 2$. Unordered monotonicity is more restrictive than ARUM in that it rules out both direct and indirect two-way flows.

E The 3×3 Model of Pinto (2021)

Pinto (2021) has proposed a 3×3 model of the Moving to Opportunity (MTO) experiment. Here we use our framework to identify response-group probabilities and several counterfactual averages.

We follow the notation in Pinto (2021). Let $\mathcal{Z} = \{z_c, z_e, z_8\}$ and $\mathcal{T} = \{t_h, t_l, t_m\}$, where

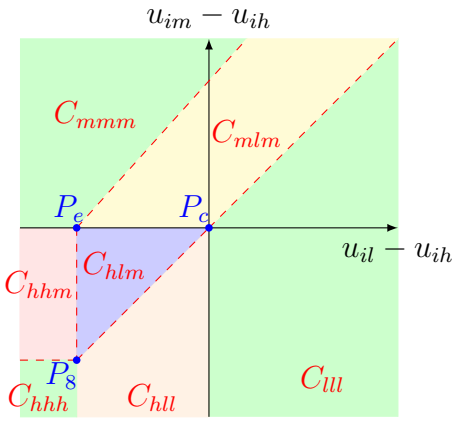
- z_c refers to control families, z_e those who received the experimental voucher, and z_8 those who received Section 8 voucher;
- t_h refers to families who did not move and chose high-poverty neighborhoods, t_l those who moved to low-poverty neighborhoods, and t_m those who moved to medium-poverty neighborhoods.

There are 7 response types in Pinto (2021): the three always-taker groups C_{hhh}, C_{lll}, and C_{mmm}, and four complier groups:

- C_{hlm}: families who choose high-poverty without vouchers, low-poverty with the experimental voucher, and medium-poverty with Section 8 vouchers (Pinto calls this group full-compliers);
- C_{hll}: families who choose high-poverty without vouchers, low-poverty with either voucher;

- C_{mlm}: families who choose medium-poverty without the experimental voucher, low-poverty with it;
- C_{hhm}: families who choose high-poverty without Section 8 voucher, medium-poverty with it.

Figure 8: MTO



The seven response groups are illustrated in Figure 8 and in Table 6.

Table 6: Response Groups in MTO

	$T_i(z) = t_h$	$T_i(z) = t_l$	$T_i(z) = t_m$
$z = z_c$	$C_{hhh} \bigcup C_{hhm} \bigcup C_{hlm} \bigcup C_{hll}$	C_{lll}	$C_{mmm} \bigcup C_{mlm}$
$z = z_e$	$C_{hhh} \bigcup C_{hhm}$	$C_{lll} \bigcup C_{mlm} \bigcup C_{hlm} \bigcup C_{hll}$	C_{mmm}
$z = z_8$	C_{hhh}	$C_{lll} \bigcup C_{hll}$	$C_{mmm} \bigcup C_{hhm} \bigcup C_{hlm} \bigcup C_{mlm}$

Proposition 10 (Response-group probabilities in MTO)**.** *The following probabilities are iden-*

tified:

$$
\begin{aligned}
\Pr(C_{hhh}) &= P(t_h|z_8), \\
\Pr(C_{lll}) &= P(t_l|z_c), \\
\Pr(C_{mmm}) &= P(t_m|z_e), \\
\Pr(C_{hhm}) &= P(t_h|z_e) - P(t_h|z_8), \\
\Pr(C_{hll}) &= P(t_l|z_8) - P(t_l|z_c), \\
\Pr(C_{mlm}) &= P(t_m|z_c) - P(t_m|z_e), \\
\Pr(C_{hlm}) &= 1 - P(t_h|z_e) - P(t_l|z_8) - P(t_m|z_c).
\end{aligned} \tag{E.1}
$$

The model has the following testable implications:

$$
\begin{aligned}
P(t_h|z_e) &\geqslant P(t_h|z_8), \\
P(t_l|z_8) &\geqslant P(t_l|z_c), \\
P(t_m|z_c) &\geqslant P(t_m|z_e), \\
1 &\geqslant P(t_h|z_e) + P(t_l|z_8) + P(t_m|z_c).
\end{aligned} \tag{E.2}
$$

The following proposition identifies a number of group average outcomes.

Proposition 11 (Identification in MTO)**.** *The following group average outcomes are point-*

identified:

$$\begin{aligned}
\mathbb{E}\left[Y_i(t_h)|i \in C_{hhh}\right] &= \frac{\bar{E}_{z_8}(t_h)}{P(t_h|z_8)}, \\
\mathbb{E}\left[Y_i(t_l)|i \in C_{hhh}\right] &= \frac{\bar{E}_{z_c}(t_l)}{P(t_l|z_c)}, \\
\mathbb{E}\left[Y_i(t_m)|i \in C_{mmm}\right] &= \frac{\bar{E}_{z_e}(t_m)}{P(t_m|z_e)}, \\
\mathbb{E}\left[Y_i(t_h)|i \in C_{hhm}\right] &= \frac{\bar{E}_{z_e}(t_h) - \bar{E}_{z_8}(t_h)}{P(t_h|z_e) - P(t_h|z_8)}, \\
\mathbb{E}\left[Y_i(t_l)|i \in C_{hll}\right] &= \frac{\bar{E}_{z_8}(t_l) - \bar{E}_{z_c}(t_l)}{P(t_l|z_8) - P(t_l|z_c)}, \\
\mathbb{E}\left[Y_i(t_m)|i \in C_{mlm}\right] &= \frac{\bar{E}_{z_c}(t_m) - \bar{E}_{z_e}(t_m)}{P(t_m|z_c) - P(t_m|z_e)}, \\
\mathbb{E}\left[Y_i(t_h)|i \in C_{hll} \bigcup C_{hlm}\right] &= \frac{\bar{E}_{z_c}(t_h) - \bar{E}_{z_e}(t_h)}{P(t_h|z_c) - P(t_h|z_e)}, \\
\mathbb{E}\left[Y_i(t_l)|i \in C_{mlm} \bigcup C_{hlm}\right] &= \frac{\bar{E}_{z_e}(t_l) - \bar{E}_{z_8}(t_l)}{P(t_l|z_e) - P(t_l|z_8)}, \\
\mathbb{E}\left[Y_i(t_m)|i \in C_{hhm} \bigcup C_{hlm}\right] &= \frac{\bar{E}_{z_8}(t_m) - \bar{E}_{z_c}(t_m)}{P(t_m|z_e) - P(t_m|z_c)}.
\end{aligned}$$

The proofs of Propositions 10 and 11 are straightforward; we omit the details.

F The MVPF of Extending Head Start

Recall our ternary instrument setting:

- $Z = 0$ means no offer of admission to Head Start or to another preschool;
- $Z = 1$ means an offer of admission in Head Start only;
- $Z = 2$ means an offer of admission in another preschool only.

$Z = 0$ does not preclude other ways to get into h or c, $Z = 1$ does not preclude other ways to get into c, and $Z = 2$ does not preclude other ways to get into h.

We denote $p(z)$ the probability that $Z = z$. We are considering an increase in $p(1)$: more offers of admission to Head Start. As $p(1)$ increases, we also increase $p(2)$ to maintain the number of slots in alternative preschools constant. Like Kline and Walters (2016), we assume that this increase in $p(2)$ only brings into alternative preschools children that would otherwise not attend preschools.

The MVPF is the ratio of the benefits dB of increasing $p(1)$ by $dp(1)$ to its budgetary costs dC. We have $B = (1-\tau)p\mathbb{E}Y$, where p is the pre-tax return to expected scores, and τ the tax rate. Hence

$$dB = (1-\tau)pd\mathbb{E}Y.$$

The budget costs are the subsidies (ϕ_j per student) to Head Start and other preschools, minus the tax receipts:

$$C = \phi_h \Pr(D = h) + \phi_c \Pr(D = c) - \tau p\mathbb{E}Y.$$

Therefore

$$\text{MVPF} = \frac{(1-\tau)pd\mathbb{E}Y/dp(1)}{\phi_h d\Pr(D = h)/dp(1) - \tau pd\mathbb{E}Y/dp(1)}.$$

In order to compute the MVPF, we start by evaluating the marginal return in expected outcomes $d\mathbb{E}Y/dp(1)$.

F.1 The Expected Change in Outcomes

Since

$$\Pr(D = c) = \Pr(D(0) = c) + \sum_{z=1,2} p(z)(\Pr(D(z) = c) - \Pr(D(0) = c))$$

to keep it constant we must have

$$\frac{dp(2)}{dp(1)} = \frac{\Pr(D(0) = c) - \Pr(D(1) = c)}{\Pr(D(2) = c) - \Pr(D(0) = c)}.$$

$D(0) = c$ implies $D(2) = c$ since $Z = 2$ targets c. Therefore $\Pr(D(2) = c) - \Pr(D(0) = c) = \Pr(D(2) = c, D(0) \neq c)$. Since $Z = 1$ targets h, $D(1) = c$ implies $D(0) = c$; and $D(0) = c$ implies that $D(1)$ can only be c or h. This gives us

$$\Pr(D(0) = c) - \Pr(D(1) = c) = \Pr(D(0) = c, D(1) \neq c) = \Pr(D(0) = c, D(1) = h),$$

which is the proportion of the group C_{ch} in the 2×3 model. Therefore

$$\frac{dp(2)}{dp(1)} = \frac{\Pr(i \in C_{ch})}{\Pr(D(2) = c, D(0) \neq c)}.$$

The resulting change in expected scores is

$$d\mathbb{E}Y = dp(1)\mathbb{E}(Y(D(1)) - Y(D(0)) + dp(2)\mathbb{E}(Y(D(2)) - Y(D(0)).$$

Now an offer of Head Start ($Z = 1$) can only move children to Head Start: $D(1) \neq D(0)$ implies that $D(1) = h$. As a consequence,

$$\mathbb{E}(Y(D(1)) - Y(D(0)) = \mathbb{E}(Y(h) - Y(D(0))|D(1) = h, D(0) \neq h) \times \Pr(D(1) = h, D(0) \neq h)$$

and by the same argument,

$$\mathbb{E}(Y(D(2)) - Y(D(0)) = \mathbb{E}(Y(c) - Y(D(0))|D(1) = c, D(0) \neq c) \times \Pr(D(2) = c, D(0) \neq c).$$

Putting things together gives

$$\begin{aligned}\frac{d\mathbb{E}Y}{dp(1)} &= \Pr(D(1) = h, D(0) \neq h)\times \\ &\quad (\mathbb{E}(Y(h) - Y(D(0))|D(1) = h, D(0) \neq h) + \mathbb{E}(Y(c) - Y(D(0))|D(1) = c, D(0) \neq c) \times S_c) \\ &= \Pr(D(1) = h, D(0) \neq h) \times (\text{LATE}_h + S_c\text{LATE}_c)\,,\end{aligned}$$

where

$$S_c = \frac{\Pr(i \in C_{ch})}{\Pr(D(1) = h, D(0) \neq h)}$$

is, as in the text of the paper, the proportion of the h-compliers that come from c.

F.2 The MVPF

We still need to compute the denominator $d\Pr(D = h)/dp(1)$. It is

$$(\Pr(D(1) = h) - \Pr(D(0) = h)) - \frac{dp(2)}{dp(1)}(\Pr(D(0) = h) - \Pr(D(2) = h)).$$

The first term in the difference is $\Pr(D(1) = h, D(0) \neq h)$, the proportion of h-compliers. The second term equals

$$\frac{\Pr(i \in C_{ch})}{\Pr(D(2) = c, D(0) \neq c)}(\Pr(D(0) = h) - \Pr(D(2) = h)).$$

Since $Z = 2$ targets c, the difference $\Pr(D(0) = h) - \Pr(D(2) = h)$ represents the proportion of children who would get to Head Start under $Z = 0$ and leave it when offered admission to another preschool ($Z = 2$) as $p(1)$ increases. Since these children can only have $D(2) = c$, our assumption rules out this group and the second term of the difference is zero.

As in Kline and Walters (2016), $\text{LATE}_c = \text{LATE}_{nc}$; we end up with

$$\text{MVPF} = \frac{(1-\tau)p\left(\text{LATE}_h + S_c\text{LATE}_{nc}\right)}{\phi_h - \tau p\left(\text{LATE}_h + S_c\text{LATE}_{nc}\right)},$$

which happens to coincide with the formula used by Kline and Walters (2016).

G An Example of Positive Codependence

In this section, we provide details for the example that satisfies the positive codependence assumption in Assumption 7. Recall that we have considered the example:

$$\mathbb{E}(Y_i(2)|u_{i0}, u_{i1}, u_{i2}) - \mathbb{E}Y_i(2) = a_0 u_{i0} + a_1 u_{i1} + a_2 u_{i2},$$

where a_0, a_1, and a_2 are some constants. Rewrite

$$\begin{aligned}
\mathbb{E}(Y_i(2)|u_{i0}, u_{i1}, u_{i2}) - \mathbb{E}Y_i(2) &= (a_0 + a_1 + a_2)u_{i0} - a_1(u_{i2} - u_{i1}) + (a_1 + a_2)(u_{i2} - u_{i0}) \\
&= (a_0 + a_1 + a_2)u_{i0} - a_1\zeta_i + (a_1 + a_2)\xi_i \\
&= \mathbb{E}(Y_i(2)|u_{i0}, \zeta_i, \xi_i) - \mathbb{E}Y_i(2).
\end{aligned}$$

Hence,

$$\mathbb{E}(Y_i(2)|\zeta_i, \xi_i) - \mathbb{E}Y_i(2) = (a_0 + a_1 + a_2)\mathbb{E}(u_{i0}|\zeta_i, \xi_i) - a_1\zeta_i + (a_1 + a_2)\xi_i.$$

Also, recall that we have assumed that (u_{i0}, u_{i1}, u_{i2}) are jointly normal and mutually uncorrelated with the common mean 0 and the common variance 1. Thus,

$$\begin{pmatrix} u_{i0} \\ \zeta_i \\ \xi_i \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 1 \\ -1 & 0 & 0 \end{pmatrix} \begin{pmatrix} u_{i0} \\ u_{i1} \\ u_{i2} \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 & -1 \\ 0 & 2 & 1 \\ -1 & 1 & 2 \end{pmatrix} \right],$$

which implies that

$$\mathbb{E}(u_{i0}|\zeta_i, \xi_i) = \begin{pmatrix} 0 & -1 \end{pmatrix} \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}^{-1} \begin{pmatrix} \zeta_i \\ \xi_i \end{pmatrix} = \frac{1}{3}\zeta_i - \frac{2}{3}\xi_i.$$

Combining all together yields

$$\begin{aligned}\mathbb{E}(Y_i(2)|\zeta_i,\xi_i) - \mathbb{E}Y_i(2) &= (a_0+a_1+a_2)\left(\frac{1}{3}\zeta_i - \frac{2}{3}\xi_i\right) - a_1\zeta_i + (a_1+a_2)\xi_i \\ &= \frac{a_2+a_0-2a_1}{3}\zeta_i + \frac{a_2+a_1-2a_0}{3}\xi_i\end{aligned}$$

Thus, in this example, Assumption 7 holds if and only if

$$a_2 + a_0 \geqslant 2a_1 \text{ and } a_2 + a_1 \geqslant 2a_0.$$