Filtered and Unfiltered Treatment Effects with Targeting Instruments*

Sokbae Lee† **Bernard Salanié**‡

December 15, 2020

Abstract

Multivalued treatments are commonplace in applications. We explore the use of discrete-valued instruments to control for selection bias in this setting. We establish conditions under which counterfactual averages and treatment effects are identified for heterogeneous complier groups. These conditions restrict (i) the unobserved heterogeneity in treatment assignment, (ii) how the instruments target the treatments, and optionally (iii) the extent to which counterfactual averages are heterogeneous. We allow for limitations in the analyst's information via the concept of a filtered treatment. Finally, we illustrate the usefulness of our framework by applying it to data from the Student Achievement and Retention Project and the Head Start Impact Study.

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

*We thank Josh Angrist, Junlong Feng and Len Goff 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, 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; and multivalued instruments are just as frequent. 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].

As the training, education choice, and tax-benefit examples illustrate, multivalued treatments are often also subject to selection on unobservables. We explore in this paper the use of discrete-valued instruments in order to control for selection bias when evaluating discrete-valued treatments. We establish conditions under which counterfactual averages and treatment effects are identified for various (sometimes composite) complier groups. These conditions require a combination of assumptions that restrict both the unobserved heterogeneity in treatment assignment and the configuration of the instruments themselves. Some of our results also restrict the heterogeneity of counterfactual averages.

In addition, when treatments can take multiple values the analyst often only observes a partition of treatment choices. She might for instance only know whether an unemployed individual went through a training program, without knowing exactly which program it was. More generally, the analyst only observes a *filtered treatment*. Treatment effects are of course harder to identify in the filtered model. As filtering often occurs in applications, we examine identification both in unfiltered and in filtered treatment.

Existing work on multivalued treatments under selection on observables includes Imbens (2000), Cattaneo (2010), and Ao, Calonico, and Lee (2019) among others. The literature that uses discrete-valued instruments to evaluate treatment effects under selection on unobservables is more sparse. On the theoretical side, Angrist and Imbens (1995) analyzed two-stage least squares (TSLS) estimation when the treatment takes a finite number of ordered values. Heckman, Urzua, and Vytlacil (2006); Heckman and Vytlacil (2007b); Heckman, Urzua, and Vytlacil (2008) showed how treatment effects can be identified in discrete choice models for the ordered and unordered cases, respectively. More recently, Heckman and Pinto (2018) focused on unordered treatments and introduced the notion of "unordered monotonicity" under which treatment assignment is formally analogous to an additively separable discrete

[1]Muralidharan, Romero, and Wüthrich (2019) review recent applications of factorial designs.

choice model. Several recent papers have studied the case of binary treatments with multiple instruments, as well as binary instruments with multivalued or continuous treatments. For the former, Mogstad, Torgovitsky, and Walters (2020a) and Goff (2020) analyzed the identifying power of different monotonicity assumptions[2]. For the latter, Torgovitsky (2015), D'Haultfoeuille and Février (2015), Huang, Khalil, and Yildiz (2019), Caetano and Escanciano (2020), and Feng (2020) developed identification results for different models. On the applied side, Kirkeboen, Leuven, and Mogstad (2016) used discrete instruments to obtain TSLS estimates of returns to different fields of study. Kline and Walters (2016) revisited the Head Start Impact Study (HSIS) and accounted for the presence of a substitute treatment (alternative preschools in this case). Kamat (2020) also analyzed the average effects of Head Start preschool access using the HSIS dataset. Pinto (2015, 2019) applied unordered monotonicity to the Moving to Opportunity program.

Our work is substantially different from any of the aforementioned papers. Rather than focusing on specific cases, we seek a parsimonious framework within which many useful models with multiple treatments and multiple instruments can be analyzed. Both filtering and the multiplicity of treatments and instruments may give rise to a bewildering number of cases. In the binary/binary model, the analyst can usually take for granted that switching on the binary instrument makes treatment (weakly) more likely for any observation[3].

With multiple instrument values and multiple treatments, the correspondence is much less direct. In order to reduce the complexity of the problem, we start by imposing an additive random-utility model (ARUM) structure on the unfiltered treatment model. While we do this mostly for practicality, we should note here that it is related to the unordered monotonicity property of Heckman and Pinto (2018). Under ARUM, it is natural to speak of an instrument *targeting* an unfiltered treatment by increasing its relative "mean value". Most of our paper relies on the assumption of *strict targeting*, which obtains when each instrument only promotes the treatments it targets.

To illustrate, consider the effect of various programs T on some outcomes Y. Let each instrument value z stand for a policy regime, under which the access to some programs is made easier or harder than in a control group. Under ARUM, this translates into a profile of relative mean values of any treatment t under the policy regimes $z \in \mathcal{Z}$. We say that an instrument value z targets a treatment t when it maximizes its relative mean value. For simplicity, we will use the term "subsidy" to refer to any exogenous shift in the mean values;

[2]Mogstad, Torgovitsky, and Walters (2020b) 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).

[3]This is satisfied under the LATE-monotonicity assumption (e.g., Imbens and Angrist, 1994; Vytlacil, 2002; Heckman and Vytlacil, 2007a).

in the other examples cited at the beginning of the introduction, costs or eligibility conditions could be used instead. Suppose, then, that each policy regime consists of values of subsidies for a subset of the programs, and that these subsidies enter mean values additively. Then a policy regime z targets a treatment t if it has the highest subsidy for this program among all policy regimes. Strict targeting requires that all policy regimes z' that do not target t have the same (lower) subsidy for t.

With complete treatment data (the unfiltered treatment), combining ARUM and strict targeting allows us to point-identify the size of some complier groups and the corresponding treatment effects, and to partially identify others. We use two examples to demonstrate the identification power and implications of ARUM and strict targeting. In our first example, three unordered treatment values target three instrument values. This 3×3 model was also studied by Kirkeboen, Leuven, and Mogstad (2016). Our second 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). We obtain novel identification results for both examples; they lead to new estimands, which provide more information than the TSLS estimators.

Our second contribution is to explore the impact of filtering on identifiability. Filtering breaks down some of the structure of the unfiltered treatment model. Even if the latter has strict, one-to-one targeting, instruments will typically not target filtered treatments D. Furthermore, ARUMs are usually not additive after filtering. As a consequence, unordered monotonicity may not carry over to the filtered treatment (an observation already in Lee and Salanié (2018)). We give a simple example in which the unfiltered treatment model has strict one-to-one targeting, and yet the filtered treatment model allows for two-way flows. We show that our assumptions on unfiltered treatment and on the filtering process nevertheless allow us to identify various parameters of interest in the filtered treatment model. We demonstrate this in our two leading examples, and also within a factorial design with a vector of two binary instruments and three treatment values.

Finally, we illustrate the usefulness of our framework by applying it to data from the Student Achievement and Retention (STAR) Project (Angrist, Lang, and Oreopoulos, 2009) and to Kline and Walters's (2016) analysis of the Head Start Impact Study. We find that the large intent-to-treat (ITT) effect of the STAR for female college students results from the aggregation of two very different treatment effects; this highlights the value of unbundling the heterogeneous compliers. We also 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.

The remainder of the paper is organized as follows. Section 1 defines our framework and introduces filtered and unfiltered treatments. In Section 2, we study identification in

the unfiltered treatment model. We define the concepts of targeting, one-to-one targeting, and strict targeting; we derive their implications for the identification of the probabilities, the counterfactual averages, and the treatment effects of various complier groups. Section 3 turns to filtered models. We obtain new identification results in several leading classes of applications. Finally, we present estimation results for the two aforementioned empirical studies in Section 4. The Appendices contain the proofs of all propositions and lemmata, along with some additional material.

1 Filtered and Unfiltered Treatment

In all of the paper, we denote observations as $i = 1, \ldots, n$. We focus throughout on a treatment that takes discrete values, which we label $d \in \mathcal{D}$. For simplicity, we will call $D = d$ "treatment d". These values are unordered: e.g. $d = 2$, when available, is not "more treatment" than $d = 1$. In most of our examples, there is a well-defined control group, which is denoted by $d = 0$. We assume that discrete-valued instruments $Z_i \in \mathcal{Z}$ are available. We condition on all other exogenous covariates X_i throughout, and we omit them from the notation. We will use the standard counterfactual notation: $D_i(z)$ and $Y_i(d, z)$ denote respectively potential treatments and outcomes. $\mathbb{1}(A)$ denotes the indicator of set A.

The validity of the instruments requires the usual exclusion restrictions:

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

(ii) $Y_i(d)$ *and* $D_i(z)$ *are independent of* Z_i *for all* (d, z) *in* $\mathcal{D} \times \mathcal{Z}$.

Under Assumption 1, we define $D_i := D_i(Z_i)$ and $Y_i := Y_i(D_i)$. Throughout the paper, we assume that we observe (Y_i, D_i, Z_i) for each i. In addition, the instruments must be relevant. In the usual binary instrument/binary treatment case (hereafter "binary/binary"), this translates into a requirement that the propensity score vary with the instruments. In our more general setting, we impose:

Assumption 2 (Relevant Instruments)**.** *Let* $\mathbf{Z}_i$ *denote a column vector whose elements are* $\mathbb{1}(Z_i = z)$ *for* $z \in \mathcal{Z}$*, and* $\mathbf{D}_i$ *denote a column vector whose elements are* $\mathbb{1}(D_i = d)$ *for* $d \in \mathcal{D}$*. Then* $\mathbb{E}[\mathbf{Z}_i \mathbf{D}_i^\top]$ *has full rank.*

1.1 Filtering

We now move beyond these standard assumptions. With multivalued treatments, it is often the case that the analyst only observes a coarse version of treatment assignment. She

may only know, for instance, that an individual received training—but not which of several training programs was used. Similarly, she may only know that a subsidy was granted, but not the value of the subsidy.

To allow for this possibility, we let our observed treatment D to be a "filtered" version of an underlying "unfiltered" treatment variable T: $D = M(T)$, where M is a *filtering map* that is surjective but may not be injective. In the two examples above, M would map several training programs into a "received training" category, and several subsidy values into a "received subsidy" category.

1.2 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 (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 are not ordered in any meaningful way, we cannot apply the results in Angrist and Imbens (1995) for instance. Mogstad, Torgovitsky, and Walters (2020a) state several versions of monotonicity for a binary treatment model with $|\mathcal{Z}| > 2$. They propose an assumption PM (partial monotonicity) which applies binary LATE-monotonicity component by component. This requires that the instruments be interpretable as vectors, 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}, or their support. The usual models of multinomial choice belong to this family. ARUM also includes ordered treatments, for which $u_{it} = \sigma(t)u_i$ for some increasing positive function σ.

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

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

Adding these two restrictions (and assuming that one of these inequalities is strict) implies the "increasing differences" property:

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

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 "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 in a set $\mathcal{Z}$ can be represented as an ARUM.

1.3 Assignment to Treatment

Filtering and ARUM define the class of models of assignment to treatment that we analyze in this paper. To pursue the discrete choice analogy: in the unfiltered model, each observation chooses a treatment within $\mathcal{T}$ and the analyst observes this choice. In a filtered model, choices are aggregated into groups; the analyst only observes which group the treatment chosen belongs to. The aggregation occurs via a filtering map from $\mathcal{T}$ to $\mathcal{D}$.

Assumption 3 (ARUM)**.** *The unfiltered treatment assignment model consists of:*

1. *a finite set $\mathcal{T}$;*
2. *a finite set of instrument values $\mathcal{Z}$;*
3. *an ARUM model of unfiltered 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.

Assumption 4 (Filtered Treatment)**.** *A set $\mathcal{D}$ is a partition of $\mathcal{T}$; or equivalently, there exists a surjective* filtering map $M : \mathcal{T} \to \mathcal{D}$.

We call the model that generates $D_i = M(T_i)$ the *filtered treatment model* and the model for T_i the *underlying unfiltered treatment model*. 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 3 is just the standard monotonicity assumption, with a threshold-crossing rule

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

and the only possible filtering maps are trivial.

If we add a third unfiltered treatment value so that $\mathcal{T} = \{0, 1, 2\}$, the ARUM assumption starts to bite as it excludes two-way flows in the unfiltered treatment model. However, it does allow for two-way flows on the observed, filtered treatment D. Suppose for instance that $\mathcal{D} = \{0, 1\}$ and the filtering map has $M(0) = 0$ and $M(1) = M(2) = 1$: the analyst only observes whether $T_i > 0$. Take one observation i with $T_i(z) = 0$ and $T_i(z') = 1$, so that $D_i(z) = 0$ and $D_i(z') = 1$. Under an ARUM for T, there may exist other observations j with $T_j(z) = 2$ and $T_j(z') = 0$. Such observations would have $D_j(z) = 1$ and $D_j(z') = 0$, invalidating LATE-monotonicity for the filtered treatment model[4].

In Lee and Salanié (2018), we studied models with multivalued treatments and *continuous* instruments. We allowed for treatment assignment to be determined by any logical combination (using AND, NOT and OR) of a finite set of threshold-crossing rules of the form $u_{ij} \leqslant Q_j(z)$. This class of models can be generated by

1. taking an ARUM of assignment to treatment values $T \in \mathcal{T}$,
2. generating the observed treatment $D \in \mathcal{D}$ from a partition of the set $\mathcal{T}$.

Aa a consequence, all examples with continuous instruments in Lee and Salanié (2018) translate directly to the discrete instruments setting in this paper. Consider for instance the following double hurdle model; it has $|\mathcal{D}| = 2$, and an underlying unfiltered treatment model with $|\mathcal{T}| = 3$.

Example 1 (Double Hurdle Treatment)**.** The unfiltered treatment has $\mathcal{T} = \{0, 1, 2\}$ and

$$T_i(z) = \arg\max_{t=0,1,2} \left(U_z(t) + u_{it}\right),$$

[4]For a numerical example, take $U_t(z) = t$ and $U_0(z'), U_1(z'), U_2(z') = 0, 3, 0$.

where the vector (u_{i0}, u_{i1}, u_{i2}) is distributed independently of Z_i.

Suppose that the filtered treatment is generated by $D = \mathbb{1}(T = 0)$, which corresponds to the filtering map $M(0) = 1, M(1) = M(2) = 0$; that is,

$$\begin{cases} D_i(z) = 1 & \text{iff } \max(U_z(1) + u_{i1}, U_z(2) + u_{i2}) < U_z(0) + u_{i0} \\ D_i(z) = 0 & \text{otherwise.} \end{cases} \tag{1.1}$$

Here our logical combination of threshold rules is simply an "AND" over the two inequalities: $u_{i0} - u_{i1} > U_z(1) - U_z(0)$ and $u_{i0} - u_{i2} > U_z(2) - U_z(0)$. □

Lee and Salanié (2018) gave a set of assumptions under which the marginal treatment effect can be identified in a filtered treatment model, provided that enough continuous instruments are available. In Example 1, we would need two continuous instruments, and some additional restrictions. The current paper is exploring identification with discrete-valued instruments. In these settings, the combination of Assumptions 1, 2, and 3 is far from sufficient to identify interesting treatment effects in filtered and unfiltered treatment models 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 unfiltered treatments t. We first state a general definition[5].

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

- Observation i has (elemental) *response-vector* $R_{\tilde{t}}$ if and only if for all $z \in \mathcal{Z}$, $T_i(z) = \tilde{t}(z)$. The set $C_{\tilde{t}}$ denotes the set of observations with response-vector $R_{\tilde{t}}$ and we call it a *response-group*.

- We extend the definition in the natural way to incompletely specified mappings, where $\tilde{t}$ is a correspondence from $\mathcal{Z}$ to $\mathcal{T}$. We call the corresponding response-vectors and response-groups *composite*.

2 Identifying the Unfiltered Treatment Model

We start by introducing additional assumptions on the underlying unfiltered treatment model. We will illustrate these assumptions on two examples which we call the "binary instrument model" or the "$2 \times T$" model; and the "3×3 model". We first define them briefly.

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

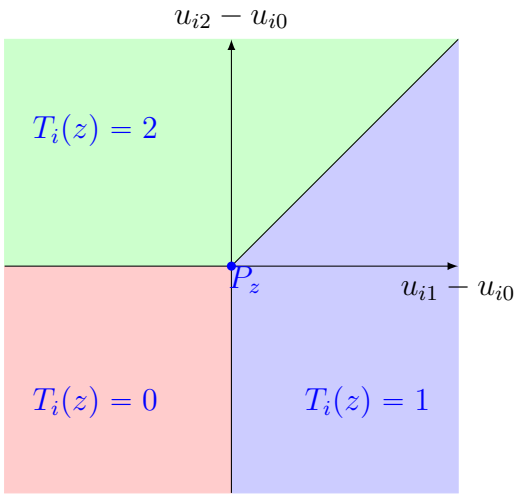
Example 2 (The binary instrument $(2 \times T)$ model)**.** $\mathcal{T} = \{0, 1, \ldots, T-1\}$ and $\mathcal{Z} = \{0, 1\}$. □

Example 3 (3×3 unfiltered model)**.** Assume that $\mathcal{Z} = \{0, 1, 2\}$ and $\mathcal{T} = \{0, 1, 2\}$. In the $(u_{i1} - u_{i0}, u_{i2} - u_{i0})$ plane, the points of coordinates $P_z = (U_z(0) - U_z(1), U_z(0) - U_z(2))$ for $z = 0, 1, 2$ are important; 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.

This is shown in Figure 1 for a given z, where the origin is in P_z. Note that the u_{id} need not have full support; it would be easy to accommodate restrictions. □

Figure 1: Unfiltered treatment assignment in the 3×3 model for given z



When we discuss filtering in section 3, we will use a factorial design with imperfect compliance:

Example 4 ($4 \times T$ factorial design)**.** $\mathcal{Z} = \{0 \times 0, 0 \times 1, 1 \times 0, 1 \times 1\}$, where the two digits indicate the values of two binary instruments z_1 and z_2; and $\mathcal{T} = \{0, 1, \ldots, T-1\}$ with $2 \leqslant T \leqslant 4$. □

2.1 Targeted Treatments

"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.

Under Assumption 3, 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. Intuitively, an instrument z targets a treatment t if it makes the difference $(U_z(t) - U_z(\tau))$ as large as possible for given τ. Instead of requiring this for any τ, we will choose a *reference treatment* $t_0 \in \mathcal{T}$ and require that z maximize $(U_z(t) - U_z(t_0))$ for this particular t_0. In many applications, the control group is a natural choice for a reference treatment. Since the control group is usually denoted by $t = 0$, we will extend the notation and denote the reference treatment by $t_0 = 0$.

The following definition makes this more precise.

Definition 2 (Targeted Treatments and Targeting Instruments)**.** Let $t = 0$ denote a reference treatment value. For any $z \in \mathcal{Z}$ and $t \in \mathcal{T}$, let

$$\Delta_z(t) \equiv U_z(t) - U_z(0)$$

denote the *relative mean* of treatment t given instrument z. Moreover, let

$$\bar{\Delta}_t \equiv \max_{z \in \mathcal{Z}} \Delta_z(t) \quad \text{and} \quad \bar{Z}(t) \equiv \arg\max_{z \in \mathcal{Z}} \Delta_z(t)$$

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

Definition 2 calls for several remarks. First, by construction $\Delta_z(0) \equiv 0$ and $\bar{Z}(0) = \mathcal{Z}$. Therefore $t = 0$ is not in $\mathcal{T}^*$; the set $\mathcal{T}^*$ may exclude other treatment values, however.

Instruments that do not target any treatment ($z \notin \mathcal{Z}^*$) yield dominated relative mean values in the following sense: for every $t \in \mathcal{T}^*$, $\Delta_z(t) < \bar{\Delta}_t$. If a treatment value t is not targeted ($t \notin \mathcal{T}^*$), by definition the function $z \to \Delta_z(t)$ is constant over $z \in \mathcal{Z}$, with value $\bar{\Delta}_t$. While treatment values 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 $\bar{Z}$ and $\bar{T}$ are not observable; any assumption on targeting instruments and targeted treatments must be a priori and will be context-dependent. As we will see, these prior assumptions sometimes have consequences that can be tested.

Let us return to the illustration that we used in the introduction. A policy regime 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 relative mean values are $\Delta_z(t) = \Delta_0(t) + S_z(t)$ and for any treatment t, the set $\bar{Z}(t)$ consists of the instrument values z that subsidize t most heavily. As this illustration suggests, the sets $\bar{Z}(t)$ may not be singletons, and they may well intersect. We will show this on several examples.

2.1.1 Targeting Examples

Example 5 (4×3 design)**.** Suppose that $\mathcal{T} = \{0, 1, 2\}$, and we have two binary instruments. $z_1 = 1$ is intended to promote treatment 1 and $z_2 = 1$ is intended to promote treatment 2. Using similar notation to Example 4 :

$$\Delta_{1\times 0}(1) = \bar{\Delta}_1 > \max(\Delta_{0\times 0}(1), \Delta_{0\times 1}(1)),$$
$$\Delta_{0\times 1}(2) = \bar{\Delta}_2 > \max(\Delta_{0\times 0}(2), \Delta_{1\times 0}(2)).$$

Depending on the context, it may be reasonable to assume that $\Delta_{1\times 1}(1) = \Delta_{1\times 0}(1)$ and $\Delta_{1\times 1}(2) = \Delta_{0\times 1}(2)$: turning on the two instruments increases the appeal of $t = 1$ (resp. $t = 2$) just as much as if only z_1 (resp. z_2) had been turned on. This would be quite natural if $z_1 = 1$ subsidizes treatment 1 and $z_2 = 1$ subsidizes treatment 2: then 1×1 is the policy regime that subsidizes both. Then we have $\bar{Z}(1) = \{1 \times 0, 1 \times 1\}$ and $\bar{Z}(2) = \{0 \times 1, 1 \times 1\}$; instrument $z = 1 \times 1$ targets both $t = 1$ and $t = 2$, so that $\bar{T}(1 \times 1) = \{1, 2\}$. That is, each non-zero treatment value is targeted by two instrument values, and one instrument value targets both non-zero treatments. □

Example 6 (Two Instruments Target the Same Treatment)**.** Let us now modify Example 5 slightly: the instrument can only take values $0 \times 0, 1 \times 0$, and 1×1. Then $z = 1 \times 0$ and $z = 1 \times 1$ both target treatment $t = 1$: $\bar{Z}(1) = \{1 \times 0, 1 \times 1\}$. □

The following special case of Example 2 may also be helpful.

Example 7 (A Binary Instrument Targets Two Treatments)**.** In this example, $\mathcal{Z} = \{0, 1\}$ and $\mathcal{T} = \{0, 1, 2\}$. A fraction of individuals in the sample receives a subsidy $z = 1$ that can be used for both treatments $t = 1$ and $t = 2$; under $z = 0$, no treatment is subsidized. We would expect that $\Delta_1(1) > \Delta_0(1)$ and $\Delta_1(2) > \Delta_0(2)$, so that $\bar{Z}(1) = \bar{Z}(2) = \{1\}$; then we have $\mathcal{T}^* = \{1, 2\}$ and $\mathcal{Z}^* = \{1\}$. □

2.1.2 One-to-one Targeting

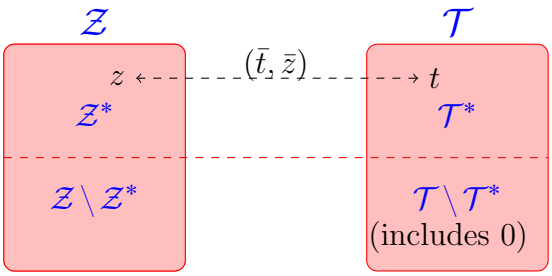
Sometimes we will impose the much stronger Assumption 5, or only one of its two parts. The first part says that a targeted treatment can only have one targeting instrument; the second part stipulates that a targeting instrument may target only one treatment.

Assumption 5 (One-to-one Targeting)**.** *(i) For any $t \in \mathcal{T}^*$, the set $\bar{Z}(t)$ is a singleton $\{\bar{z}(t)\}$.*

(ii) For any $z \in \mathcal{Z}^*$, $\bar{T}(z)$ is a singleton $\{\bar{t}(z)\}$.

Example 5 violates both parts of Assumption 5. Example 6 violates its first part only, and Example 7 only violates its second part. Note that if both parts of Assumption 5 hold, the function $\bar{z}$ and its inverse $\bar{t}$ are bijective between $\mathcal{Z}^*$ and $\mathcal{T}^*$, and these two sets the same number of elements. Figure 2 illustrates one-to-one targeting.

Figure 2: One-to-one Targeting



To illustrate, we now impose Assumption 5 on Examples 2 and 3.

Example 2 with One-to-one Targeting. Take our binary instrument model (Example 2). Suppose that the observations with $z = 1$ receive a subsidy for the treatment $t = 1$. Other treatment values $t \neq 1$ are not subsidized. Then $\Delta_1(t) = \Delta_0(t)$ for all $t \neq 1$, so that $\mathcal{Z}^* = \{1\}$. □

Example 3 with One-to-one Targeting. In the 3×3 model of Example 3, each $z > 0$ instrument could be a subsidy that targets the corresponding treatment $\bar{t}(z)$ in the sense that the subsidy $S_{z'}(t)$ is highest for $z' = z$. □

Assumption 3, conjoined with Assumption 5, imposes some useful restrictions on response groups.

Proposition 1 (Unfiltered response groups (1))**.** *Under Assumptions 3 and 5, for any $t \in \mathcal{T}^*$:*

- *if $T_i(\bar{z}(t)) = 0$, then $T_i(z) \neq t$ for all $z \in \mathcal{Z}$;*

- *as a consequence, all response-groups $C_{\tilde{t}}$ with $\tilde{t}(\bar{z}(t)) = 0$ and $\tilde{t}(z) = t$ for some $z \neq t$ are empty.*

Example 3 (continued) Return to the 3×3 model and assume that the targeted set of treatments $\mathcal{T}^* = \{1, 2\}$ and that Assumptions 3 and 5 hold. This imposes

$$\Delta_1(1) > \max(\Delta_2(1), \Delta_0(1)) \text{ and } \Delta_2(2) > \max(\Delta_1(2), \Delta_0(2)).$$

A possible interpretation is that policy regime $z = 1$ (resp. $z = 2$) subsidizes treatment $t = 1$ (resp. $t = 2$) more than policy regimes $z = 0$ and $z = 2$ (resp. $z = 1$) do.

Since P_z has coordinates $(-\Delta_z(1), -\Delta_z(2))$,

- P_1 must lie to the left of P_0 and P_2,
- 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 $3^3 = 27$ response-vectors, R_{000} to R_{222}, with corresponding response-groups C_{000} to C_{222}. Groups C_{ddd} are "always-takers"[6] of treatment 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, R_{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,

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

Assumption 3 asserts the emptiness of four composite groups out of the 27 possible: C_{10*}, C_{*01}, C_{*20}, and C_{2*0} by Proposition 1. They correspond to 10 elemental groups.[7]

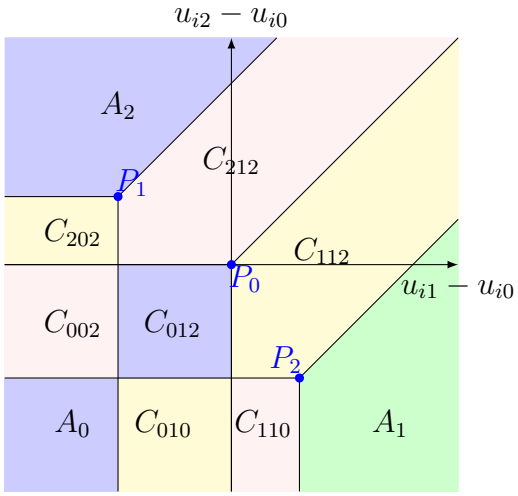
This still leaves us with 17 elemental groups, and potentially complex assignment patterns. Consider for instance Figure 3. It shows one possible configuration for the 3×3 model; the positions for P_0, P_1 and P_2 are consistent with Assumptions 3 and 5.

The number of distinct response-groups (ten) and the contorted shape of the C_{212} and C_{112} groups in Figure 3 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.

[6] 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.

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

Figure 3: A 3×3 example



Heckman and Pinto (2018, pp. 16–20) and Kirkeboen, Leuven, and Mogstad (2016) also studied the 3×3 model; they proposed sets of assumptions that identify some treatment effects. While Heckman and Pinto's example is rather specific, the framework in Kirkeboen, Leuven, and Mogstad (2016) is similar to ours; we will return to the differences between our approaches in Section 2.4. □

2.2 Strict Targeting

Figure 3 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. But these are assumptions on the dependence of the $U_z(d)$ on instruments. The first one imposes $\Delta_1(2) = \Delta_0(2)$ and the second one imposes $\Delta_2(1) = \Delta_0(1)$. This can be interpreted as policy regime $z = 1$ (resp. $z = 2$) subsidizing treatment $t = 1$ (resp. $z = 2$) only.

To put it differently, we are now requiring that the instruments $z \in \bar{Z}(t)$, which maximize $\Delta_z(t) = U_z(t) - U_z(0)$, should not shift assignment between the other values of the treatment. The following assumption is a direct extension of the discussion above to our general discrete model.

Assumption 6 (Strict Targeting)**.** *Take any targeted treatment value $t \in \mathcal{T}^*$. Then the function $z \in \mathcal{Z} \to \Delta_z(t)$ takes the same value for all $z \notin \bar{Z}(t)$. We denote this common value by $\underline{\Delta}_t$, and we will say of the instrument values $z \in \bar{Z}(t)$ that they* strictly target *t.*

Under Assumption 6, turning on instrument $z \in \bar{Z}(t)$ promotes treatment t without affecting the relative mean values $\Delta_z(t')$ of other treatment values t'. This explains our use

of the term "strict targeting". In this ARUM specification, an instrument in $\bar{Z}(t)$ 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 $z \in \bar{Z}(t)$ subsidize t at the same high rate, and all other instrument values offer the same, lower subsidy.[8]

Note that while we only state the assumption for $t \in \mathcal{T}^*$, it holds by definition for all $t \in \mathcal{T} \backslash \mathcal{T}^*$. Since $\bar{Z}(t) = \mathcal{Z}$ for these treatment values, $\underline{\Delta}_t = \bar{\Delta}_t$ is the common value of $\Delta_z(t)$ over all of $\mathcal{Z}$ when $t \notin \mathcal{T}^*$.

Moreover, Assumption 6 only bites for a given $t \in \mathcal{T}^*$ if $\mathcal{Z} \backslash \bar{Z}(t)$ has at least two values. Since $\bar{Z}(t)$ is never empty, this shows that Assumption 6 automatically holds if $|\mathcal{Z}| = 2$ (one binary instrument), as in our Example 2. It only holds in the 4×3 design of Example 5 if $\Delta_{0\times1}(1) = \Delta_{0\times0}(1)$ and $\Delta_{1\times0}(2) = \Delta_{0\times0}(2)$ (so that $z = 0 \times 1$ does not subsidize $t = 1$ and $z = 1 \times 0$ does not subsidize $t = 2$).

Note that one-to-one targeting and strict targeting are logically independent assumptions: neither one implies the other. As we just saw, the design in Example 5 may exhibit strict targeting; but it never satisfies one-to-one targeting. The converse may also hold, for instance if $z = 1$ and $z = 2$ both subsidize $t = 1$, and $z = 2$ is a more generous subsidy. Then we would expect $\bar{Z}(1) = \{2\}$ yet $\Delta_0(1) < \Delta_1(1)$.

Example 8 (School Vouchers)**.** To shed light on Assumption 6, consider two possible policies for allocating school vouchers. A first policy consists of school-specific vouchers. Each individual i is offered randomly a choice of $m_i \geqslant 0$ vouchers for a subset Z_i of schools. If $m_i \geqslant 1$, the individual may choose to use a voucher to enroll in a school in Z_i, to enroll in another school, or to drop out altogether. Let T_i denote this choice, with $T_i = 0$ denoting dropping out. For any given school $t \neq 0$, the value of $\Delta_z(t)$ is highest when $t \in z$, since then a voucher can be used. Therefore $\bar{Z}(t)$ is the set of menus of vouchers that include school t; and $\mathcal{T}^*$ is the set of schools for which a voucher is sometimes, but not always offered. Whether $\bar{Z}(t)$ is a single menu or not, all other menus of vouchers yield the same $\Delta_z(t)$: school t is strictly targeted[9].

Another possible policy consists in "universal vouchers." These would subsidize tuition for every year of study in all schools, in the hope of increasing the number of years of education. Now z is a subsidy rate, and t the number of years of education. Since a higher subsidy rate reduces the cost of education, for any t the function $\Delta_z(t)$ achieves its maximum $\bar{\Delta}_t$ for the highest subsidy rate $z_{\max}$ on offer: for each t, $\bar{Z}(t) = \{z_{\max}\}$ and Assumption 6 fails. More

[8] Again, any of these "subsidies" could be zero or negative.

[9] Note that iff $m_i \leqslant 1$ for each individual, targeting is one-to-one. If not, either part of Assumption 5 could fail. If the schools are (A, B, C, D), a set of two menus $z_1 = \{A, C\}$ and $z_2 = \{B, D\}$ fails the second part of Assumption 5; a set $z_1 = \{A, C\}$ and $z_2 = \{B, C\}$ fails both parts.

importantly, if $|\mathcal{Z}| > 2$ then for any $t > 0$, the value of $\Delta_z(t)$ increases with the subsidy rate z. Strict targeting would clearly not be an appropriate assumption in this setting. $\square$

Extending our geometric illustration of Example 3, let P_z be the point in $\mathbb{R}^{|\mathcal{T}^*|}$ with coordinates $(-\Delta_z(t))_{t\in\mathcal{T}^*}$. Under Assumption 6, the point P_z has its t coordinate equal to $-\overline{\Delta}_t$ on any axis t which it targets ($t \in \bar{T}(z)$), and $-\underline{\Delta}_t$ on any other axis. Since $-\underline{\Delta}_t > -\overline{\Delta}_t$, two points P_z and $P_{z'}$ have the same coordinate on any axis $t \notin \bar{T}(z) \bigcup \bar{T}(z')$; and P_z is below $P_{z'}$ on axis t if $t \in \bar{T}(z)\backslash\bar{T}(z')$.

Now suppose that $\mathcal{Z} \setminus \mathcal{Z}^*$ contains at least two values z_0 and z_1. Since neither targets any treatment, under Assumption 6 $\Delta_{z_1}(t) = \Delta_{z_0}(t)$ for any $t \in \mathcal{T}^*$. Moreover, $\Delta_z(t)$ equals $\underline{\Delta}_t$ for all $z \in \mathcal{Z}$ if $t \notin \mathcal{T}^*$. Therefore the functions Δ_{z_0} and Δ_{z_1} coincide on all of $\mathcal{T}$. By the previous paragraph, if $z \in \mathcal{Z}^*$ then the point $P_{z_0} = P_{z_1}$ is above the point P_z on any axis $t \in \bar{T}(z)$. Moreover, the counterfactual treatments $T_i(z_0)$ and $T_i(z_1)$ must be equal for any observation i. In that sense, all instrument values in $\mathcal{Z} \setminus \mathcal{Z}^*$ are equivalent under strict targeting.

We summarize this in Lemma 1.

Lemma 1 (Some consequences of strict targeting)**.** *Under Assumption 3 and Assumption 6,*

(i) The coordinates of two points P_z and $P_{z'}$ in $\mathbb{R}^{|\mathcal{T}^*|}$ coincide on any axis t' that is not in the symmetric difference $\bar{T}(z) \triangle \bar{T}(z')$.

(ii) If $z \in \bar{Z}(t)$ and $z' \notin \bar{Z}(t)$, the point $P_{z'}$ is above the point P_z on the axis t.

(iii) The set of instrument values $\mathcal{Z}$ is either $\mathcal{Z}^*$, or the union of $\mathcal{Z}^*$ and of a non-empty set that we denote $\mathcal{Z}_0$. In the latter case, the point P_{z_0} is the same for all $z_0 \in \mathcal{Z}_0$; for any $z \in \mathcal{Z}^*$, it is above the point P_z on any axis $t \in \bar{T}(z)$, and it has the same coordinates on all other axes.

In many applications there is a "control group", which receives no program subsidy or any other intervention. This group provides us with a natural reference instrument z_0. To avoid multiplying subcases, we assume from now on that $\mathcal{Z}_0$ is non-empty.

Assumption 7 (Reference Instruments)**.** *The set of reference instruments $\mathcal{Z}_0 = \mathcal{Z}\backslash\mathcal{Z}^*$ in Lemma 1 (iii) is non-empty.*

Strict targeting imposes a lot of structure on the mapping from instruments to treatments. To make this clear, we first state a definition.

Definition 3 (Top targeted and top alternative treatments)**.** Take any observation i in the population.

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

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

and $T_i^*(z) \subset \bar{T}(z)$ denote the set of maximizers. We call the elements of T_i^* the *top targeted treatments*.

(ii) Also define

$$\Delta_i^* = \max_{t \in \mathcal{T}} (\underline{\Delta}_t + u_{it})$$

and let $\tau_i^* \subset \mathcal{T}$ denote the set of maximizers. We call the elements of τ_i^* the *top alternative treatments*.

Under strict targeting, an observation i can react to being assigned an instrument z in two ways. If z is in $\mathcal{Z}^*$, then i can choose among the treatments that z targets. Alternatively, it may choose as if no treatment was targeted (as it must if z is not in $\mathcal{Z}^*$). We now make this more rigorous by proving that observations can only opt for one of their top targeted treatments, if any, or for one of their top alternative treatments.

We now state our main result on response-groups.

Proposition 2 (Unfiltered response groups under strict targeting)**.** *Let Assumptions 3, 6 and 7 hold. Then for every observation i,*

(i) if $z \in \mathcal{Z}^*$, then $T_i(z)$ can only be in $T_i^*(z)$ or in τ_i^*.

(ii) $T_i(\mathcal{Z}_0) \subset \tau_i^*$.

For simplicity, we work from now on under the assumption that the distribution of the error terms in the ARUM has no mass points.

Assumption 8 (Absolutely continuous errors)**.** *The distribution of the random vector $(u_{it})_{t \in \mathcal{T}}$ is absolutely continuous.*

Under Assumption 8, the sets τ_i^* and $T_i^*(z)$ are singletons[10] with probability 1; with a minor abuse of notation, we let τ_i^* and $T_i^*(z)$ denote their elements.

Proposition 3 (Unfiltered classes under strict targeting)**.** *Under Assumptions 3, 6, 7, and 8, the population consists of classes denoted by $c(A, \tau)$, where A is a possibly empty subset of $\mathcal{Z}^*$ and τ is a treatment value. If observation i is in $c(A, \tau)$, then the following holds.*

- *$T_i(z) = T_i^*(z)$ for all $z \in A$.*
- *$\tau_i^* = \tau$, and for all $z \in \mathcal{Z} \setminus A$, $T_i(z) = \tau$.*
- *If $\tau \in \mathcal{T}^*$, then $\bar{Z}(\tau) \subset A$.*

[10]Note that this does not extend to the sets $\bar{Z}(t)$ and $\bar{T}(z)$, which can still have several elements.

2.2.1 Strict one-to-one targeting

Proposition 3 has a straightforward corollary under one-to-one targeting (Assumption 5). Recall that under one-to-one targeting, the sets $\bar{Z}(t)$ and $\bar{T}(z)$ are singletons; then $T_i^*(z) = z$ for each z in $\mathcal{Z}^*$. This simplifies the statement of our characterization result as we only need to distinguish $\tau = 0$ and $\tau \in A \subset \mathcal{Z}^*$.

Corollary 1 (Unfiltered classes under strict, one-to-one targeting)**.** *Under Assumptions 3, 5, 6, 7, and 8, the population consists of classes denoted by $c(A, \tau)$, where A is a possibly empty subset of $\mathcal{Z}^*$ and*

1. *$\tau \in \bar{t}(A)$ or $\tau \in \mathcal{T} \setminus \mathcal{T}^*$;*
2. *if observation i is in $c(A, \tau)$, then $\tau_i^* = \tau$ and*
   (a) *$T_i(z) = \bar{t}(z)$ for all $z \in A$,*
   (b) *$T_i(z) = \tau$ for all $z \in \mathcal{Z} \setminus A$.*

Given any (possibly empty) subset A of $\mathcal{Z}^*$ and a treatment value τ, an observation i belongs to $c(A, \tau)$ if and only if for all $z \in A$ and $z' \in \mathcal{Z}^* \setminus A$,

$$V_i^*(z') < \Delta_i^* = \underline{\Delta}_\tau + u_{i\tau} < V_i^*(z).$$

The following examples may be useful. When A is empty, part 1 of the Corollary imposes that $\tau \in \mathcal{T} \setminus \mathcal{T}^*$. Whatever the value of the instrument z is, an observation i in $c(\varnothing, \tau)$ will take up a non-targeted treatment. If now A is the singleton $\{z\}$ for some $z \neq 0$, the class $c(\{z\}, \bar{t}(z))$ is the corresponding group of always-takers $A_{\bar{t}(z)}$. In the polar case $A = \mathcal{Z}^*$, when it is assigned a targeting instrument value ($z \in \mathcal{Z}^*$), the observation complies by picking one of the treatments it targets ($T_i(z) = T_i^*(z)$, which is $\bar{t}(z)$ under one-to-one targeting). When both A and $\mathcal{Z} \setminus A$ are non-empty, the observation complies when the instrument z is in A, and it does not respond to changes in the value of z when it is in $\mathcal{Z} \setminus A$.

Figure 4 represents the mapping of instruments to treatments for an observation i in class $c(A, \tau)$ under strict, one-to-one targeting when $\tau \in A$. Figure 5 shows a class $c(A, \tau)$ for $\tau \in \mathcal{T} \setminus \mathcal{T}^*$. The white areas show that treatment values in $\mathcal{T}^* \setminus \bar{t}(A)$ are never assigned.

2.2.2 Applications

Example 3 (continued) To illustrate Corollary 1, we return to the 3×3 model of Example 3, where $\mathcal{Z}^* = \mathcal{T}^* = \{1, 2\}$ and $\mathcal{Z} = \mathcal{T} = \{0, 1, 2\}$. A can be $\varnothing, \{1\}, \{2\}$, or $\{1, 2\}$, with corresponding values of τ in $\{0\}, \{0, 1\}, \{0, 2\}$ or $\{0, 1, 2\}$ respectively. The class $c(\varnothing, 0)$

Figure 4: An unfiltered class $c(A, \tau \in \bar{t}(A))$ under strict, one-to-one targeting

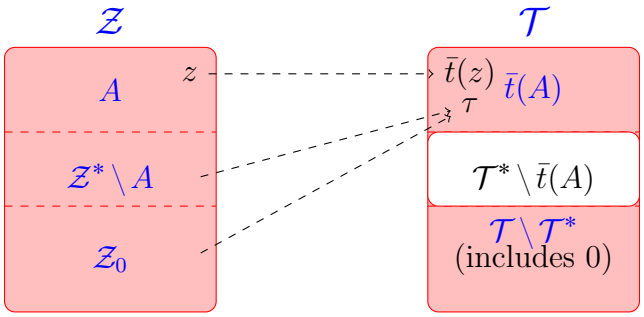
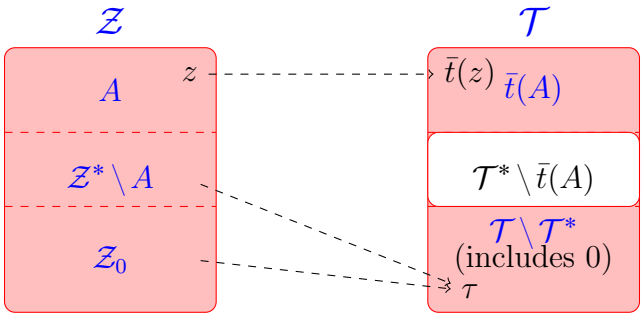


Figure 5: An unfiltered class $c(A, \tau \in \mathcal{T} \setminus \mathcal{T}^*)$ under strict, one-to-one targeting



corresponds to the always-takers of 0, $A_0 = C_{000}$. For $A = \{1\}$ we get C_{010} and A_1, and for $A = \{2\}$ we get C_{002} and A_2. Finally, with $A = \{1, 2\}$ we obtain the composite response group $C_{*12} = C_{012} \bigcup C_{112} \bigcup C_{212}$.

The eight elemental response groups are illustrated in Figure 6, again with the origin in P_0. Comparing Figure 6 with Figure 3 shows the identifying power of Assumption 6. □

Sometimes one can obtain the characterization in Corollary 1 without imposing one-to-one targeting. To see this, consider the following variant of the targeted binary instrument model.

Example 2 with Strict Targeting and $T = 3$. Assume that $\mathcal{T} = \{0, 1, 2\}$ and $\mathcal{Z} = \{0, 1\}$. We interpret $z = 1$ as offering a subsidy for $t = 1$, and $z = 0$ as the absence of subsidy; treatment $t = 2$ is never subsidized. Therefore $\Delta_1(1) > \Delta_0(1)$ and $\Delta_1(2) = \Delta_0(2)$; we have $\bar{Z}(1) = \{1\}, \bar{Z}(2) = \{0, 1\} = \mathcal{Z}$, and $\mathcal{T}^* = \mathcal{Z}^* = \{1\}$. Since we only have a binary instrument, strict targeting holds in this example.

We can have classes $A = \varnothing$ with $\tau \in \{0, 2\}$, and $A = \{1\}$ with $\tau \in \mathcal{T}$. The former generates the always-takers groups $A_0 = C_{00}$ and $A_2 = C_{22}$, and the latter has the two groups of compliers C_{01} and C_{21} and the always-taker group $A_1 = C_{11}$. These five elemental response-groups are illustrated in Figure 7.

If we had not imposed $\Delta_0(2) = \Delta_1(2)$, Assumption 6 would still hold but $t = 2$ would

Figure 6: Unfiltered, strictly one-to-one targeted treatment: 3×3 model

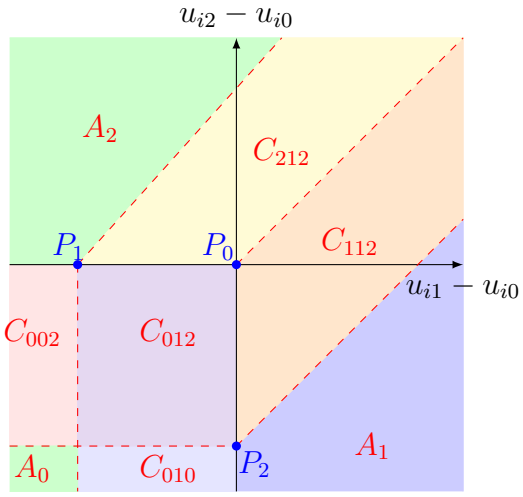
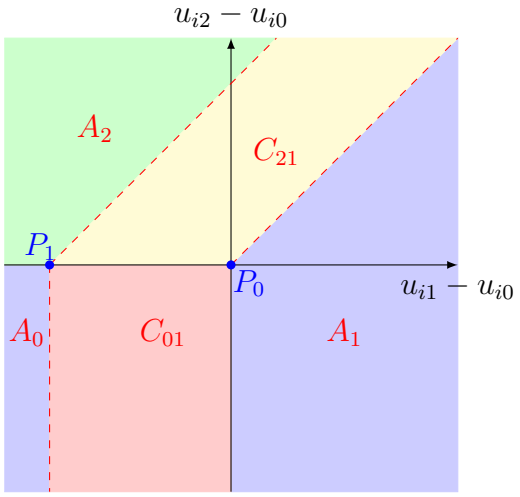


Figure 7: 2×3 model with one targeted treatment



belong to $\mathcal{T}^*$. If for instance $t = 1$ and $t = 2$ are both training programs, being offered a subsidy for $t = 1$ may also make the recipient more aware of the value of training in general. In that case we would have $\Delta_1(2) > \Delta_0(2)$ and $\bar{Z}(2) = \{1\}$, so that $\mathcal{T}^* = \{1, 2\}$. We would not have one-to-one targeting anymore since $z = 1$ would target both $t = 1$ and $t = 2$. Therefore, Corollary 1 would not apply. Proposition 3 would apply, however, allowing for a sixth response group C_{12}, with $A = \{1\}$ and $\tau = 2$.

Still, it seems likely that a subsidy for $t = 1$ would increase the appeal of $t = 1$ more than that of $t = 2$:

$$(U_1(1) - U_0(1)) - (U_1(2) - U_0(2)) = (\Delta_1(1) - \Delta_0(1)) - (\Delta_1(2) - \Delta_0(2)) > 0.$$

This is enough to rule out the possibility of the response group C_{12}. To see this, assume that $T_i(0) = 1$. This implies $U_0(1) + u_{i1} > U_0(2) + u_{i2}$, so that

$$U_1(1) + u_{i1} > U_0(2) + (U_1(1) - U_0(1)) + u_{i2} > U_1(2) + u_{i2}$$

and $T_i(1)$ cannot be 2. □

2.3 Identifying Group Probabilities

Now that we have characterized response-groups, we seek to identify the probabilities of the corresponding response-groups in the unfiltered treatment model.

Definition 4 (Generalized propensity scores)**.** We write $P(t|z)$ for the generalized propensity score $\Pr(T_i = t|Z_i = z)$.

2.3.1 Strict, One-to-one Targeting

Under strict, one-to-one targeting, the response-groups are easily enumerated.

Proposition 4 (Counting response-groups under strict, one-to-one targeting)**.** *Under Assumptions 3, 5, 6, 7, and 8, the number of response-groups is*

$$N = (2|\mathcal{T}| - |\mathcal{Z}| + 1) \times 2^{|\mathcal{Z}|-2}.$$

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 $(|\mathcal{T}| - 1) \times |\mathcal{Z}|$. As the probabilities of the response-groups must sum to one, we have $(N - 1)$ unknowns.

Table 1: Number of required identifying restrictions: unfiltered treatment under strict, one-to-one targeting

Row	$\mathcal{T}$	$\mathcal{Z}$	$N-1$	$\lvert\mathcal{Z}\rvert \times (\lvert\mathcal{T}\rvert - 1)$	Required	Example
(1)	{0,1}	{0,1}	2	2	0	*LATE*
(2)	{0,1,...,$\lvert\mathcal{T}\rvert-1$}	{0,1}	$2(\lvert\mathcal{T}\rvert-1)$	$2(\lvert\mathcal{T}\rvert-1)$	0	Example 2
(3)	{0,1,2}	{0,1,2}	7	6	1	Example 3

Table 1 shows some values of the number of equations $(|\mathcal{T}| - 1) \times |\mathcal{Z}|$ and the number of unknowns $(N - 1)$ for three examples. We focus there on the leading case in which $\mathcal{Z} = \{0\} \bigcup \mathcal{Z}^*$. The first row of $|\mathcal{T}| = 2$ and $|\mathcal{Z}| = 2$ is the standard LATE case: the response

group consists of never-takers (A_0), compliers (C_{01}), and always-takers (A_1). Row (2) shows another case of exact identification, and row (3) reveals that one restriction is required for the 3×3 model. More generally, the degree of underidentification in a $T \times T$ model increases as T gets larger.

It is not difficult to write down the equations that link observed propensity scores and group probabilities.

Proposition 5 (Identifying equations for response-groups: unfiltered treatment under strict, one-to-one targeting)**.** *Under Assumptions 1, 2, 3, 5, 6, 7, and 8, the empirical content of the generalized propensity scores of the unfiltered treatment model is the following system of equations, for all $z \in \mathcal{Z}$ and $t \in \mathcal{T}$:*

$$
\begin{aligned}
P(t|z) = & \sum_{A \subset \mathcal{Z}^* \setminus \{z\}} \mathbf{1}(t \notin \mathcal{T}^* \setminus \bar{t}(A)) \Pr(c(A,t)) \\
& + \mathbf{1}(z \in \mathcal{Z}^*, t = \bar{t}(z)) \sum_{\substack{A \subset \mathcal{Z}^* \\ z \in A}} \sum_{\tau \notin \mathcal{T}^* \setminus \bar{t}(A)} \Pr(c(A,\tau)).
\end{aligned} \tag{2.1}
$$

While this may look cryptic, it it easy enough to apply in specific cases.

2.3.2 Applications

Proposition 5 can be applied directly to some of the rows of Table 1. According to the table, our Example 2 is just identified under strict, one-to-one targeting. Proposition 6 confirms it and gives explicit formulæ, along with simple testable predictions. To avoid repetitions, in the remainder of Section 2, we assume that Assumptions 1, 2, 3, 5, 6, 7, and 8 hold.

Proposition 6 (Response-group probabilities in Example 2)**.** *The following probabilities are identified:*

$$
\begin{aligned}
\Pr(A_1) &= P(1|0), \\
\Pr(A_t) &= P(t|1) \text{ for } t \neq 1, \\
\Pr(C_{t1}) &= P(t|0) - P(t|1) \text{ for } t \neq 1.
\end{aligned} \tag{2.2}
$$

The model has $(|T| - 1)$ testable predictions:

$$
P(t|0) \geqslant P(t|1) \text{ for } t \neq 1.
$$

Row (3) of Table 1 is the 3×3 model of Example 3, in which eight elemental groups are non-empty. One restriction is missing to point-identify the probabilities of all eight

response-groups. Table 2 shows which groups take $T_i = t$ when $Z_i = z$.

Table 2: Response Groups of Example 3

	$T_i(z) = 0$	$T_i(z) = 1$	$T_i(z) = 2$
$z = 0$	$A_0 \bigcup C_{010} \bigcup C_{002} \bigcup C_{012}$	$A_1 \bigcup C_{112}$	$A_2 \bigcup C_{212}$
$z = 1$	$A_0 \bigcup C_{002}$	$A_1 \bigcup C_{010} \bigcup C_{012} \bigcup C_{112} \bigcup C_{212}$	A_2
$z = 2$	$A_0 \bigcup C_{010}$	A_1	$A_2 \bigcup C_{002} \bigcup C_{012} \bigcup C_{112} \bigcup C_{212}$

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. In addition, the probabilities of two composite groups of compliers are point-identified. The other four probabilities are constrained by three adding-up constraints.

Proposition 7 (Response-group probabilities in the 3×3 model of Example 3)**.** *The following probabilities are identified:*

$$\begin{aligned}
\Pr(A_1) &= P(1|2), \\
\Pr(A_2) &= P(2|1), \\
\Pr(C_{112}) &= P(1|0) - P(1|2), \\
\Pr(C_{212}) &= P(2|0) - P(2|1), \\
\Pr(C_{010} \bigcup C_{012}) &= P(0|0) - P(0|1), \\
\Pr(C_{002} \bigcup C_{012}) &= P(0|0) - P(0|2), \\
\Pr(C_{010} \bigcup A_0) &= P(0|2).
\end{aligned} \tag{2.3}$$

The model has the following testable implications:

$$\begin{aligned}
&P(1|0) \geqslant P(1|2), \\
&P(2|0) \geqslant P(2|1), \\
&P(0|0) \geqslant \max(P(0|1), P(0|2)).
\end{aligned} \tag{2.4}$$

The four partially-identified group probabilities can be parameterized as

$$\begin{aligned}
\Pr(A_0) &= p, \\
\Pr(C_{002}) &= P(0|1) - p, \\
\Pr(C_{010}) &= P(0|2) - p, \\
\Pr(C_{012}) &= P(0|0) - P(0|1) - P(0|2) + p,
\end{aligned}$$

where the unknown p satisfies

$$\max\{0, P(0|1) + P(0|2) - P(0|0)\} \leqslant p \leqslant \min\{1, P(0|1), P(0|2)\}.$$

2.4 Identifying Effects of Unfiltered Treatment

We now establish identification of treatment effects for the complier groups whose probabilities are identified. To simplify the exposition, we introduce one more element of notation.

Definition 5 (Conditional average outcomes)**.** For any $z \in \mathcal{Z}$ and $t \in \mathcal{T}$, we define the conditional average outcome by

$$\bar{E}_z(t) = \mathbb{E}(Y_i \mathbb{1}(T_i = t) | Z_i = z).$$

To give a trivial example, the LATE formula (row (1) of Table 1) is

$$\mathbb{E}(Y_i(1)|i \in C_{01}) = \frac{\bar{E}_1(1) - \bar{E}_0(1)}{P(1|1) - P(1|0)} \quad \text{and} \quad \mathbb{E}(Y_i(0)|i \in C_{01}) = \frac{\bar{E}_0(0) - \bar{E}_1(0)}{P(1|1) - P(1|0)},$$

yielding the familiar form[11]:

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

While the $\bar{E}_z(t)$ are directly identified from the data, the conditional average group 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 observed conditional average outcomes. We will use the following identity repeatedly:

[11] Throughout the remainder of the paper, we assume, as is standard, that probability differences appearing in the denominator of estimands are always nonzero.

Lemma 2 (Decomposing conditional average outcomes)**.** *Let $z \in \mathcal{Z}$ and $t \in \mathcal{T}$. Then*

$$\bar{E}_z(t) = \sum_{C_{(z)}=t} \mathbb{E}(Y_i(t)|i \in C) \Pr(i \in C),$$

where $C_{(z)} = t$ means that response group C has treatment t when assigned instrument z. In addition,

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

First consider Example 2, where the probabilities of all $(2|\mathcal{T}| - 1)$ response groups are identified (Proposition 6).

Proposition 8 (Identification in the $2 \times T$ model under strict, one-to-one targeting)**.** *The following quantities are point-identified:*

$$\begin{aligned}
\mathbb{E}\left[Y_i(1)|i \in A_1\right] &= \frac{\bar{E}_0(1)}{P(1|0)}, \\
\mathbb{E}\left[Y_i(t)|i \in A_t\right] &= \frac{\bar{E}_1(t)}{P(t|1)} \text{ for } t \neq 1, \\
\mathbb{E}\left[Y_i(t)|i \in C_{t1}\right] &= \frac{\bar{E}_0(t) - \bar{E}_1(t)}{P(t|0) - P(t|1)} \text{ for } t \neq 1.
\end{aligned}$$

However, the standard Wald estimator only partially identifies the average treatment effects on the complier groups C_{t1}:

$$\begin{aligned}
\frac{\mathbb{E}(Y_i|Z_i = 1) - \mathbb{E}(Y_i|Z_i = 0)}{\Pr(T_i = 1|Z_i = 1) - \Pr(T_i = 1|Z_i = 0)} &= \frac{(\bar{E}_1(1) - \bar{E}_0(1)) - \sum_{t \neq 1}(\bar{E}_0(t) - \bar{E}_1(t))}{P(1|1) - P(1|0)} \\
&= \sum_{t \neq 1} \alpha_t \mathbb{E}\left[Y_i(1) - Y_i(t)|i \in C_{t1}\right], \qquad (2.5)
\end{aligned}$$

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 positive and sum to 1.

Proposition 8 shows that we only identify a convex combination (with point-identified weights) of the ATEs on the $|\mathcal{T}^*|$ complier groups. It is possible to bound the average treatment effects in a straightforward manner if we assume that the support of Y_i is known and finite. Alternatively, we may add conditions 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 the homogeneity of the average outcomes under treatment.

Corollary 2 (Point-identifying treatment effects in the $2 \times T$ model)**.** *Suppose that the average counterfactual outcomes under treatment* 1 *are identical for all complier groups:*

$$\mathbb{E}\left[Y_i(1)|i \in C_{t1}\right] \text{ does not depend on } t \neq 1. \tag{2.6}$$

Then the average treatment effects for all complier groups C_{t1} *are point-identified:*

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

To interpret the homogeneity condition in (2.6), suppose that we are concerned with the effect of one subsidized program ($t = 1$) when other, unsubsidized programs ($t > 1$) are also available. Then (2.6) imposes that outcomes for compliers (who switch to the subsidized program when offered a subsidy) are on average the same regardless where the compliers switched from.

We now move to the 3×3 model of Example 3. The following proposition identifies a number of mean counterfactual outcomes.

Proposition 9 (Identification in the 3×3 model under strict, one-to-one targeting)**.** *The following quantities are point-identified:*

$$\begin{aligned}\mathbb{E}\left[Y_i(1)|i \in A_1\right] &= \frac{\bar{E}_2(1)}{P(1|2)}, \\ \mathbb{E}\left[Y_i(2)|i \in A_2\right] &= \frac{\bar{E}_1(2)}{P(2|1)}, \\ \mathbb{E}\left[Y_i(0)|i \in C_{010} \bigcup C_{012}\right] &= \frac{\bar{E}_0(0) - \bar{E}_1(0)}{P(0|0) - P(0|1)}, \\ \mathbb{E}\left[Y_i(0)|i \in C_{002} \bigcup C_{012}\right] &= \frac{\bar{E}_0(0) - \bar{E}_2(0)}{P(0|0) - P(0|2)}, \\ \mathbb{E}\left[Y_i(1)|i \in C_{010} \bigcup C_{012} \bigcup C_{212}\right] &= \frac{\bar{E}_1(1) - \bar{E}_0(1)}{P(1|1) - P(1|0)}, \\ \mathbb{E}\left[Y_i(1)|i \in C_{112}\right] &= \frac{\bar{E}_0(1) - \bar{E}_2(1)}{P(1|0) - P(1|2)}, \\ \mathbb{E}\left[Y_i(2)|i \in C_{002} \bigcup C_{012} \bigcup C_{112}\right] &= \frac{\bar{E}_2(2) - \bar{E}_0(2)}{P(2|2) - P(2|0)}, \\ \mathbb{E}\left[Y_i(2)|i \in C_{212}\right] &= \frac{\bar{E}_0(2) - \bar{E}_1(2)}{P(2|0) - P(2|1)}.\end{aligned}$$

By itself, Proposition 9 does not allow us to identify an average treatment effect for *any* (even composite) response-group. Suppose for instance that we want to identify $E(Y_i(1) - Y_i(0)|i \in C)$ for some group C. Then C needs to exclude A_1, C_{112}, and C_{212}, since $E(Y_i(0)|i \in C')$ is not identified for any group C' that contains A_1, C_{112}, or C_{212}. Since we only know the mean outcome of treatment 1 for groups that contain one of these three subgroups, the conclusion follows.

Note, however, that if $E(Y_i(1)|i \in C_{112}) = E(Y_i(1)|i \in C_{212})$ then we can subtract the sixth equation of Proposition 9 from the fifth (after reweighting) to obtain $E(Y_i(1)|i \in C_{010} \bigcup C_{012})$ and identify the average effect of treatment 1 on this composite complier group. The following corollary exploits such homogeneity assumptions.

Assumption 9 (Homogeneity in the 3×3 model)**.** *Either or both of the following assumptions hold:*

$$\mathbb{E}\left[Y_i(1)|i \in C_{112}\right] = \mathbb{E}\left[Y_i(1)|i \in C_{212}\right], \tag{2.7}$$

$$\mathbb{E}\left[Y_i(2)|i \in C_{112}\right] = \mathbb{E}\left[Y_i(2)|i \in C_{212}\right]. \tag{2.8}$$

Corollary 3 (Point-identifying treatment effects in the 3 × 3 model)**.**

- *Under* (2.7)*, the average treatment effect*

$$\mathbb{E}\left[Y_i(1) - Y_i(0)|i \in C_{010} \bigcup C_{012}\right]$$

is identified as

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

- *Under* (2.8)*, the average treatment effect*

$$\mathbb{E}\left[Y_i(2) - Y_i(0)|i \in C_{010} \bigcup C_{012}\right]$$

is identified as

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

- *If both* (2.7) *and* (2.8) *hold, the average treatment effect*

$$\mathbb{E}\left[Y_i(1) - Y_i(2)|i \in C_{112} \bigcup C_{212}\right]$$

is identified as

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

To interpret the homogeneity conditions in Assumption 9, consider a hypothetical program to encourage college attendance. Let $z = 1$ be a tuition subsidy that can only be used for a STEM curriculum, and $z = 2$ a tuition subsidy for a non-STEM curriculum. The treatments are: not going to college ($t = 0$), studying STEM in college ($t = 1$), and opting for a non-STEM college curriculum ($t = 2$); the outcome Y is later earnings. The response groups C_{112} and C_{212} are "college always-takers" who choose the major for which they receive a subsidy. They only differ in the major they would choose in the absence of a subsidy; that difference may be negligible relative to what separates them from the other groups of compliers (C_{0**}), who would not go to college without a subsidy. The homogeneity assumption (9) formalizes this intution. It allows us to identify three different local average treatment effects (LATEs):

- $\mathbb{E}\left[Y_i(1) - Y_i(0) | i \in C_{010} \bigcup C_{012}\right]$ is the return to a STEM major for "STEM-major compliers" ($C_{010} \bigcup C_{012}$);
- $\mathbb{E}\left[Y_i(2) - Y_i(0) | i \in C_{002} \bigcup C_{012}\right]$ is the return to non-STEM major for "non-STEM-major compliers" ($C_{002} \bigcup C_{012}$);
- and $\mathbb{E}\left[Y_i(1) - Y_i(2) | i \in C_{112} \bigcup C_{212}\right]$, that is, the difference in earnings between STEM and non-STEM majors for "college always-takers".

Kirkeboen, Leuven, and Mogstad (2016) used a 3×3 model to study a similar question: 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. They call the least stringent version (condition (iii) in their Proposition 2) "irrelevance and information on next-best alternatives". As they make clear, this consists of two quite distinct parts. "Irrelevance" is a set of exclusion restrictions, while "additional information" assumes the availability of additional data. We show in Appendix C that our combination of Assumption 3 and Assumption 6 yields exactly the same identifying restrictions as the *irrelevance* condition in Kirkeboen, Leuven, and Mogstad (2016), by a quite different path.

The irrelevance condition (or strict targeting) in itself is too weak to give the TSLS 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 dummy variables $\mathbb{1}(T_i = 1)$ and $\mathbb{1}(T_i = 2)$, with instruments Z_i.

Proposition 10 (TSLS in the 3×3 model under strict, one-to-one targeting)**.** *The parameters β_1 and β_2 satisfy*

$$\begin{pmatrix} \Pr(i \in C_{010} \bigcup C_{012} \bigcup C_{212}) & -\Pr(i \in C_{212}) \\ -\Pr(i \in C_{112}) & \Pr(i \in C_{002} \bigcup C_{012} \bigcup C_{112}) \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}$$
$$= \begin{pmatrix} \mathbb{E}[\{Y_i(1) - Y_i(0)\}\mathbb{1}(i \in C_{010} \bigcup C_{012} \bigcup C_{212})] - \mathbb{E}[\{Y_i(2) - Y_i(0)\}\mathbb{1}(i \in C_{212})] \\ \mathbb{E}[\{Y_i(2) - Y_i(0)\}\mathbb{1}(i \in C_{002} \bigcup C_{012} \bigcup C_{112})] - \mathbb{E}[\{Y_i(1) - Y_i(0)\}\mathbb{1}(i \in C_{112})] \end{pmatrix}.$$

Proposition 10 implies that β_1 and β_2 are weighted averages of the following four local average treatment effects: $\mathbb{E}[Y_i(1) - Y_i(0)|i \in C_{010} \bigcup C_{012} \bigcup C_{212}]$, $\mathbb{E}[Y_i(1) - Y_i(0)|i \in C_{112}]$, $\mathbb{E}[Y_i(2) - Y_i(0)|i \in C_{212}]$, and $\mathbb{E}[Y_i(2) - Y_i(0)|i \in C_{002} \bigcup C_{012} \bigcup C_{112}]$. While the weights are identifiable functions of $\Pr(i \in C_{010} \bigcup C_{012} \bigcup C_{212})$, $\Pr(i \in C_{212})$, $\Pr(i \in C_{112})$, and $\Pr(i \in C_{002} \bigcup C_{012} \bigcup C_{112})$, they can be either positive or negative, which complicates interpretation further[12].

Using the additional information on next-best alternatives in Kirkeboen, Leuven, and Mogstad (2016) amounts to (in our notation) dropping response-groups C_{212} and C_{112} from the data. It is easy to see that if there is no observation i in $C_{212} \bigcup C_{112}$, Proposition 10 reduces to

$$\beta_1 = \mathbb{E}[Y_i(1) - Y_i(0)|i \in C_{010} \bigcup C_{012}],$$
$$\beta_2 = \mathbb{E}[Y_i(2) - Y_i(0)|i \in C_{002} \bigcup C_{012}],$$

which reproduces Proposition 2 (iii) of Kirkeboen, Leuven, and Mogstad (2016).

Additional information of the type used by Kirkeboen, Leuven, and Mogstad (2016) often is not available. We now show that an alternative set of assumptions can be combined with strict targeting to interpret the TSLS estimators as local average treatment effects. Remember that under strict targeting, five response-groups have $T(1) = 1$: $C_{010}, C_{012}, C_{112}$, C_{212}, and of course A_1. We call the first four groups the *1-compliers* as they take treatment $T = 1$ if and only if $z = 1$. The effect of moving a member of one of these groups from $T = 0$ to $T = 1$ a priori varies both within the group, and between the four groups. The next corollary shows that if the average effect for group C_{112} coincides with the average effect across the four groups of 1-compliers, and a similar condition applies to group C_{212} within 2-compliers, then the TSLS estimators identify well-defined LATEs.

[12]Mogstad, Torgovitsky, and Walters (2020a) give a set of assumptions under which the weights are positive in a model with multiple binary instruments.

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

$$\mathcal{C}_1 = C_{010} \bigcup C_{012} \bigcup C_{112} \bigcup C_{212},$$
$$\mathcal{C}_2 = C_{002} \bigcup C_{012} \bigcup C_{112} \bigcup C_{212}$$

denote the groups of 1-compliers and 2-compliers. Assume that

$$\begin{aligned} &\mathbb{E}[Y_i(1) - Y_i(0) | i \in C_{112}] = \mathbb{E}[Y_i(1) - Y_i(0) | i \in \mathcal{C}_1], \\ &\mathbb{E}[Y_i(2) - Y_i(0) | i \in C_{212}] = \mathbb{E}[Y_i(2) - Y_i(0) | i \in \mathcal{C}_2]. \end{aligned} \tag{2.9}$$

Then, if

$$\Pr(i \in C_{010} \bigcup C_{012} \bigcup C_{212}) \Pr(i \in C_{002} \bigcup C_{012} \bigcup C_{112}) \neq \Pr(i \in C_{212}) \Pr(i \in C_{112}), \tag{2.10}$$

the two-stage least squares estimators β_1 and β_2 satisfy

$$\beta_1 = \mathbb{E}[Y_i(1) - Y_i(0) | i \in C_{010} \bigcup C_{012} \bigcup C_{212} \bigcup C_{112}],$$
$$\beta_2 = \mathbb{E}[Y_i(2) - Y_i(0) | i \in C_{002} \bigcup C_{012} \bigcup C_{112} \bigcup C_{212}].$$

The regularity condition (2.10) ensures that the 2×2 matrix that premultiplies $(\beta_1, \beta_2)'$ in Proposition 10 is invertible. Condition (2.9) is more demanding—perhaps more than Assumption 9, although neither implies the other. Note that the groups C_{112} and C_{212} share one property: they attend college independently of the value of the instrument. The first part of Equation (2.9), for instance, imposes that the effect of taking STEM in college is on average the same for the "always-college" subgroup of the 1-compliers and for the whole group.

To summarize, it appears that the TSLS estimators in the 3×3 model are difficult to interpret unless additional information is available and/or some homogeneity assumptions such as (2.9) are imposed. The slightly more complex estimands we developed in Corollary 3 require different homogeneity assumptions. We recommend using both TSLS estimators and our own in order to explore heterogeneous treatment effects in the unfiltered 3×3 model.

3 The Filtered Treatment Model

We now turn to filtered versions of the treatment model we analyzed in the previous section. That is, we impose Assumption 4 throughout this section and consider a model with a treatment variable $D_i \in \mathcal{D}$, where the set of filtered treatment values $\mathcal{D}$ is a non-trivial

partition of the set of unfiltered treatment values $\mathcal{T} = 0, \ldots, |\mathcal{T}| - 1$. By definition, $2 \leqslant |\mathcal{D}| < |\mathcal{T}|$. We impose ARUM (Assumption 3) on the unfiltered treatment model.

Recall that $M : \mathcal{T} \to \mathcal{D}$ denotes the "filtering map": for any $d \in \mathcal{D}$, the set of unfiltered t's that generate the observation $D = d$ is $M^{-1}(d)$. The statistics that can be identified from the data are obtained by summing their unfiltered equivalent over $t \in M^{-1}(d)$.

To make this more precise, we add superscripts T or D to response groups, conditional probabilities and expectations to indicate whether they pertain to the unfiltered treatment model or to the filtered treatment model. For instance, C^T refers to a response group in the unfiltered treatment model (a "T-response group"). The filtering map transforms C^T into a "D-response group" C^D straightforwardly: if $C^T_{(z)} = t$, then $C^D_{(z)} = M(t)$. Define $\bar{M}$ to be the component-by-component extension of M, so that $\bar{M}(C^T) \equiv (M(t_1), \ldots, M(t_{|\mathcal{Z}|}))$ for $(t_1, \ldots, t_{|\mathcal{Z}|}) \in C^T$. Then the D-response groups are

$$C^D = \bigcup_{C^T \,|\, \bar{M}(C^T) = C^D} C^T,$$

with probabilities

$$\Pr(i \in C^D) = \sum_{C^T \,|\, \bar{M}(C^T) = C^D} \Pr(i \in C^T).$$

We let $P^T(t|z)$ denote the generalized propensity scores, and $\bar{E}^T_z(t)$ the conditional average outcomes of Definition 5. Their equivalents in the filtered treatment model are

$$P^D(d|z) \equiv \Pr(D_i = d | Z_i = z) = \sum_{t \in M^{-1}(d)} P^T(t|z) \tag{3.1}$$

and

$$\bar{E}^D_z(d) \equiv \mathbb{E}(Y_i \mathbb{1}(D_i = d) | Z_i = z) = \sum_{t \in M^{-1}(d)} \bar{E}^T_z(t). \tag{3.2}$$

Since we do not observe T_i, only the left-hand sides in equations (3.1) and (3.2) are identified from the data. Finally, we let $T_i(z)$ and $D_i(z)$ denote the counterfactual treatments, and $Y^T_i(t)$ and $Y^D_i(d)$ the counterfactual outcomes.

3.1 Applications

It would be easy, but perhaps not that useful, to translate the general results of Section 2.3 and Section 2.4 to the filtered treatment model. We choose to focus here on two useful classes of examples in which the unfiltered treatment model satisfies strict, one-to-one targeting.

3.1.1 Filtering in the $2 \times T$ model

Let us first return to the binary instrument/multiple unfiltered treatment model (Example 2). Since $z = 1$ targets unfiltered treatment $t = 1$, it seems natural to start with a binary filtered treatment: $D_i = \mathbb{1}(T_i = 1)$. This corresponds to a filtering map M_1 defined by

- $M_1(1) = 1$
- $M_1(t) = 0$ for $t \neq 1$.

In this case, the analyst can observe whether an observation i took the targeted treatment; if not, then i could be in any other treatment cell.

The mapping of T-response groups to D-response groups is straightforward. The groups of always takers of treatment $t = d = 1$ coincide: $A_1^D = A_1^T$. The other always-takers map into the single group $A_0^D = \bigcup_{t \neq 1} A_t^T$; and the compliers C_{t1}^T combine into $C_{01}^D = \bigcup_{t \neq 1} C_{t1}^T$. Under M_1, we have $P^D(1|z) = P^T(1|z)$ for $z = 0, 1$. That is the sum of our information on group probabilities. Moving to treatment effects, we observe $\bar{E}_z^D(1) = \bar{E}_z^T(1)$ and

$$\bar{E}_z^D(0) = \sum_{t \neq 1} \bar{E}_z^T(t) \tag{3.3}$$

for $z = 0, 1$.

This allows us to identify the probabilities of D-response group and a weighted LATE, with unknown weights this time.

Proposition 11 (Identification in the M_1-filtered $2 \times T$ model (1))**.** *(i) The probabilities of the three D-response groups are point-identified:*

$$\begin{aligned}
\Pr(A_1^D) &= \Pr(A_1^T) = P^D(1|0), \\
\Pr(A_0^D) &= \sum_{t \neq 1} \Pr(A_t^T) = 1 - P^D(1|1), \\
\Pr(C_{01}^D) &= \sum_{t \neq 1} \Pr(C_{t1}^T) = P^D(1|1) - P^D(1|0)
\end{aligned}$$

with the testable implication $P^D(1|1) \geqslant P^D(1|0)$.

(ii) The following counterfactual expectations are identified:

$$\begin{aligned}
\mathbb{E}(Y_i^D(0)|i \in A_0^D) &= \frac{\bar{E}_1^D(0)}{1 - P^D(1|1)}, \\
\mathbb{E}(Y_i^D(1)|i \in A_1^D) &= \frac{\bar{E}_0^D(1)}{P^D(1|0)}.
\end{aligned}$$

(iii) The standard Wald estimator identifies the following combination of LATEs:

$$\begin{aligned}\frac{\mathbb{E}(Y_i|Z_i=1)-\mathbb{E}(Y_i|Z_i=0)}{\Pr(D_i=1|Z_i=1)-\Pr(D_i=0|Z_i=0)} &= \frac{(\bar{E}_1^D(1)-\bar{E}_0^D(1))-(\bar{E}_0^D(0)-\bar{E}_1^D(0))}{P^D(1|1)-P^D(1|0)} \\ &= \mathbb{E}(Y_i^D(1)|i\in C_{01}^D)-\sum_{t\neq 1}\alpha_t^T\mathbb{E}(Y_i^T(t)|i\in C_{t1}^T), \end{aligned} \tag{3.4}$$

where the numbers $\alpha_t^T = \Pr(i \in C_{t1}^T | i \in C_{01}^D)$ are unidentified positive weights that sum to one.

The right-hand side of Equation (3.4) is a particular form of weighted LATE: the substitution of $\mathbb{E}(Y_i^D(0)|i \in C_{01}^D)$ by the weighted average in its second term reflects the lack of information of the analyst on the respective sizes of the groups C_{t1}^T within C_{01}^D, and on the dispersion of the average counterfactual outcomes when $z = 0$ across these groups. If these outcomes are homogeneous, then we get a stronger (if somewhat obvious) result.

Corollary 5 (Identification in the M_1-filtered $2\times T$ model (2))**.** *Assume that $\mathbb{E}(Y_i^T(t)|i \in C_{t1}^T)$ is the same for all $t \neq 1$. Then*

$$\mathbb{E}(Y_i^D(0)|i\in C_{01}^D)=\sum_{t\neq 1}\alpha_t^T\mathbb{E}(Y_i^T(t)|i\in C_{t1}^T)$$

and the standard Wald estimator identifies the LATE on D-compliers:

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

Now suppose that $|\mathcal{T}| \geqslant 3$. If we interpret $t = 0$ as a control group and all other values (including $t = 1$) as alternative treatments, then the analyst may only know whether observation i received some kind of treatment. The corresponding filtering map would be

- $M_2(0) = 0$,
- $M_2(t) = 1$ for $t > 0$.

Given the structure of the problem, this is very limited information. It becomes more useful if we combine it with the information from M_1. Let M_3 be the join of M_1 and M_2:

- $M_3(0) = 0$,
- $M_3(1) = 1$,
- $M_3(t) = 2$ for $t > 1$.

It allows the analyst to know whether an observation was treated, and if treated, whether it received the targeted treatment. The D-response groups consist of the always-takers $A_0^D = A_0^T$, $A_1^D = A_1^T$, $A_2^D = \bigcup_{t>1} A_t^T$; and the complier groups $C_{01}^D = C_{01}^T$ and $C_{21}^D = \bigcup_{t>1} C_{t1}^T$.

Proposition 12 (Identification in the M_3-filtered $2 \times T$ model)**.** *(i) The probabilities of the five D-response groups are point-identified:*

$$\begin{aligned}
&\Pr(i \in A_0^D) = P^D(0|1),\\
&\Pr(i \in A_1^D) = P^D(1|0),\\
&\Pr(i \in A_2^D) = P^D(2|1),\\
&\Pr(i \in C_{01}^D) = P^D(0|0) - P^D(0|1),\\
&\Pr(i \in C_{21}^D) = P^D(2|0) - P^D(2|1)
\end{aligned}$$

with the testable implications $P^D(0|0) \geqslant P^D(0|1)$ and $P^D(2|0) \geqslant P^D(2|1)$.

(ii) The standard Wald estimator identifies the following combination of LATEs:

$$\begin{aligned}
(3.5)\qquad &\mathbb{E}(Y_i^D(1)|i \in C_{01}^D) - \alpha_0^D \mathbb{E}(Y_i^T(0)|i \in C_{01}^T) - (1-\alpha_0^D)\sum_{t>1} \beta_t^T \mathbb{E}(Y_i^T(t)|i \in C_{t1}^T)\\
&= \frac{\mathbb{E}(Y_i|Z_i=1) - \mathbb{E}(Y_i|Z_i=0)}{\Pr(D_i=1|Z_i=1) - \Pr(D_i=1|Z_i=0)},
\end{aligned}$$

where

- $\alpha_0^D = \Pr(i \in C_{01}^D | i \in C_{01}^D \bigcup C_{21}^D)$ *is a positive weight, smaller than 1, identified by* $(P^D(0|0) - P^D(0|1))/(P^D(1|1) - P^D(1|0))$*;*
- *the numbers* $\beta_t^T = \Pr(i \in C_{t1}^T | i \in C_{21}^D)$ *are unidentified positive weights that sum to one.*

The extension to more general filters is trivial: any finer partition will identify more α_d^D parameters and allow the analyst to gain more information on the sizes of D-complier groups and to refine the interpretation of the Wald estimator.

3.1.2 Filtering in the 3×3 model

Let us now turn to the 3×3 unfiltered treatment model of Example 3. Remember that $z = 1$ subsidizes $t = 1$ and $z = 2$ subsidizes $t = 2$. Suppose now that the analyst only observes whether an individual took one of the subsidized treatments ($d = 1$ iff $t > 0$) or not ($d = t = 0$). Then $M^{-1}(0) = 0$ and $M^{-1}(1) = \{1, 2\}$. The 3×3 unfiltered treatment model

becomes a 3×2 filtered treatment model. The eight T-response groups of Proposition 7 combine into five D-response groups:

$$\begin{aligned} A_0^D &= A_0^T, \\ A_1^D &= A_1^T \bigcup A_2^T \bigcup C_{112}^T \bigcup C_{212}^T, \\ C_{001}^D &= C_{002}^T, \\ C_{010}^D &= C_{010}^T, \\ C_{011}^D &= C_{012}^T. \end{aligned}$$

We observe the conditional probabilities $P^D(1|z)$ and the average outcomes $\bar{E}_z^D(0)$ and $\bar{E}_z^D(1)$ for $z = 0, 1, 2$.

This simple example illustrates how even after imposing our strongest assumptions on the unfiltered treatment model, filtering allows for two-way flows between treatments D. Take two observations i and j such that i (resp. j) is a "pure 1-complier" (resp. a "pure 2-complier") in the unfiltered treatment model: $i \in C_{010}^T$ and $j \in C_{002}^T$. Then $T_i(1) = 1$ and $D_i(1) = M(1) = 1$, while $T_i(2) = 0$ and $D_i(2) = M(0) = 0$: when the instrument switches from $z = 1$ to $z = 2$, observation i moves from $D_i = 1$ to $D_i = 0$. On the other hand, $T_j(1) = 0$ and $T_j(2) = 2$ so that $D_j(1) = M(0) = 0$ and $D_j(2) = M(2) = 1$: observation j moves in the opposite direction to i as the instrument switches from 1 to 2. Another way to see this is that $i \in C_{001}^D$ and $j \in C_{010}^D$, while two such groups cannot both be non-empty in an unfiltered model under ARUM.

Proposition 13 (Identification in the 3×2-filtered 3×3 model)**.** *(i) The probability of the always-taker group A_1^D is point-identified as $P^D(1|0)$. The other four D-response groups probabilities are connected by three equations:*

$$\begin{aligned} \Pr(C_{01*}^D) &= \Pr(C_{010}^D) + \Pr(C_{011}^D) = P^D(0|0) - P^D(0|1), \\ \Pr(C_{0*1}^D) &= \Pr(C_{001}^D) + \Pr(C_{011}^D) = P^D(0|0) - P^D(0|2), \\ \Pr(C_{00*}^D) &= \Pr(C_{001}^D) + \Pr(A_0^D) = P^D(0|1). \end{aligned}$$

with the testable implications $P^D(0|0) \geqslant P^D(0|1)$ and $P^D(0|0) \geqslant P^D(0|2)$.

The four partially-identified probabilities can be parameterized as

$$\begin{aligned}
\Pr(C^D_{011}) &= p, \\
\Pr(C^D_{010}) &= P^D(0|0) - P^D(0|1) - p, \\
\Pr(C^D_{001}) &= P^D(0|0) - P^D(0|2) - p, \\
\Pr(A^D_0) &= P^D(0|2) + P^D(0|1) - P^D(0|0) + p,
\end{aligned}$$

where

$$\max(0, P^D(0|0) - P^D(0|1) - P^D(0|2)) \leqslant p \leqslant P^D(0|0) - \max(P^D(0|1), P^D(0|2)).$$

(ii) The following average conditional counterfactual outcomes are point-identified:

$$\begin{aligned}
\mathbb{E}(Y_i^D(0)|i \in C^D_{00*}) &= \frac{\bar{E}^D_1(0)}{P^D(0|1)}, \\
\mathbb{E}(Y_i^D(0)|i \in C^D_{01*}) &= \frac{\bar{E}^D_0(0) - \bar{E}^D_1(0)}{P^D(0|0) - P^D(0|1)}, \\
\mathbb{E}(Y_i^D(1)|i \in C^D_{01*}) &= \frac{\bar{E}^D_1(1) - \bar{E}^D_0(1)}{P^D(0|0) - P^D(0|1)}, \\
\mathbb{E}(Y_i^D(1)|i \in A^D_1) &= \frac{\bar{E}^D_1(0)}{P^D(1|0)}.
\end{aligned}$$

(iii) The standard Wald estimators identify the LATE on C^D_{01*} and on C^D_{0*1}:

$$\mathbb{E}(Y_i^D(1) - Y_i(0)|i \in C^D_{01*}) = \frac{\mathbb{E}(Y_i|Z_i = 1) - \mathbb{E}(Y_i|Z_i = 0)}{\Pr(D_i = 1|Z_i = 1) - \Pr(D_i = 1|Z_i = 0)}, \tag{3.6}$$

$$\mathbb{E}(Y_i^D(1) - Y_i(0)|i \in C^D_{0*1}) = \frac{\mathbb{E}(Y_i|Z_i = 2) - \mathbb{E}(Y_i|Z_i = 0)}{\Pr(D_i = 1|Z_i = 2) - \Pr(D_i = 1|Z_i = 0)}. \tag{3.7}$$

Note that the width of the interval on the unknown p cannot be larger than $\min(P^D(0|1), P^D(0|2))$: if either instrument $z = 1, 2$ is very effective at getting people into treatment, the sizes of all D-response groups will be almost point-identified. Since the average counterfactual outcomes on elemental D-response groups are connected by equations like

$$\mathbb{E}(Y_i(d)|i \in C^D_{01*}) = q\mathbb{E}(Y_i(d)|i \in C^D_{011}) + (1 - q)\mathbb{E}(Y_i(d)|i \in C^D_{010})$$

with $q = p/(P^D(0|0) - P^D(0|1))$, one could go further and impose homogeneity assumptions to improve the identification of elemental LATEs.

3.2 Filtered Factorial Design

We now return to Example 5, which featured a 4×3 design. Recall that we had $z = 0\times 0, 0\times 1, 1\times 0, 1\times 1$, and $\mathcal{T} = \{0, 1, 2\}$. Each instrument combines two binary instruments: the first one is meant to promote treatment $t = 1$ and the second one promotes $t = 2$. We focus here on the case when there is no complementarity (positive or negative) between the two binary instruments[13]: $\Delta^T_{1\times 1}(1) = \Delta^T_{1\times 0}(1)$ and $\Delta^T_{1\times 1}(2) = \Delta^T_{0\times 1}(2)$. This would hold for instance if each binary instrument is a price subsidy and prices enter mean values additively—a common assumption in discrete choice models. As we saw in Section 2.1, we have $\bar{Z}(1) = \{1\times 0, 1\times 1\}$ and $\bar{Z}(2) = \{0\times 1, 1\times 1\}$, so that this treatment model does not statisfy one-to-one targeting. On the other hand, we also saw that strict targeting holds if each binary instrument only has an effect on the treatment value that it targets. We will impose the corresponding assumptions $\Delta^T_{0\times 1}(1) = \Delta^T_{0\times 0}(1)$ and $\Delta^T_{1\times 0}(2) = \Delta^T_{0\times 0}(2)$.

Let us now introduce a filter, so that the analyst only observes $D_i = \mathbb{1}(T_i > 0)$. This yields a 4×2 filtered treatment model. Compared to the previous subsection, there are two important differences—the instrument takes four values rather than three, and we imposed several constraints on the mean values:

$$\begin{aligned} \Delta^T_{1\times 1}(1) &= \Delta^T_{1\times 0}(1) \\ \Delta^T_{1\times 1}(2) &= \Delta^T_{0\times 1}(2) \\ \Delta^T_{0\times 1}(1) &= \Delta^T_{0\times 0}(1) \\ \Delta^T_{1\times 0}(2) &= \Delta^T_{0\times 0}(2). \end{aligned} \tag{3.8}$$

In spite of the filtering, they will allow us to point-identify the relevant LATEs. To see this, first note that for any given observation i,

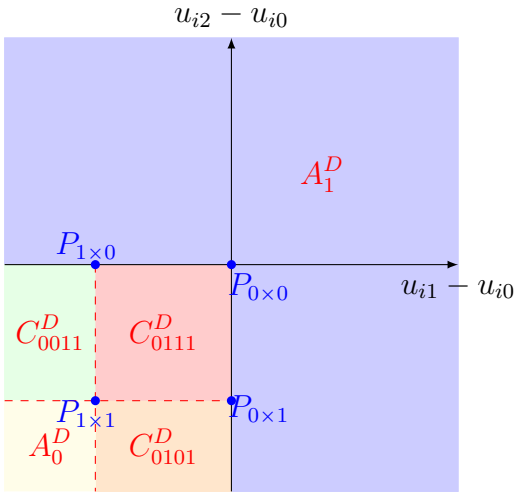
$$D_i(z) = 0 \text{ iff } u_{i0} > \max(\Delta^T_z(1) + u_{i1}, \Delta^T_z(2) + u_{i2}), \tag{3.9}$$

so that the filtered treatment model has the structure of a double hurdle model (Example 1).

First note that under our assumptions, the right hand side is largest when $z = 1 \times 1$. Therefore if $D_i(1\times 1) = 0$, observation i always takes $d = 0$. If on the other hand $D_i(0\times 0) = 1$, then i is in A^D_1 since the right-hand side can only be larger for the other instrument values. Denote indices in response-groups in the order $0\times 0, 1\times 0, 0\times 1, 1\times 1$. The preceding arguments leave only the D-response groups C^D_{0**1}. The group C^D_{0001} cannot exist since in the absence

[13] We use a superscript T to remind the reader that the argument in parentheses is an unfiltered treatment value in $\mathcal{T}$.

Figure 8: Filtered Factorial Design



of complementarity between the binary instruments,

$$\max(\Delta_{1\times 1}^T(1) + u_{i1}, \Delta_{1\times 1}^T(2) + u_{i2}) = \max(\Delta_{1\times 0}^T(1) + u_{i1}, \Delta_{0\times 1}^T(2) + u_{i2}).$$

The three other groups are[14]:

- the eager compliers C_{0111}^D: any instrument except 0×0 causes them to adopt $d = 1$,
- the reluctant compliers C_{0011}^D and C_{0101}^D: they only adopt $d = 1$ if the right binary instrument is switched on.

The resulting five D-response groups are shown in Figure 8. Table 3 shows which groups take $D_i = d$ when $Z_i = z$.

Table 3: D-response Groups

	$D_i(z) = 0$	$D_i(z) = 1$
$z = 0 \times 0$	$C_{0***}^D = A_0^D \bigcup C_{0011}^D \bigcup C_{0101}^D \bigcup C_{0111}^D$	A_1^D
$z = 1 \times 0$	$C_{00**}^D = A_0^D \bigcup C_{0011}^D$	$C_{*1*1}^D = A_1^D \bigcup C_{0101}^D \bigcup C_{0111}^D$
$z = 0 \times 1$	$C_{0*0*}^D = A_0^D \bigcup C_{0101}^D$	$C_{**11}^D = A_1^D \bigcup C_{0011}^D \bigcup C_{0111}^D$
$z = 1 \times 1$	A_0^D	$C_{***1}^D = A_1^D \bigcup C_{0011}^D \bigcup C_{0101}^D \bigcup C_{0111}^D$

[14]We borrow here the terminology of Mogstad, Torgovitsky, and Walters (2020a), which they apply to a rather different model.

Proposition 14 (Identifying the Filtered Factorial Design Model)**.** *(i) The probabilities of the D-response groups are point-identified by*

$$\begin{aligned}
\Pr(A_0^D) &= P^D(0|1\times 1)\\
\Pr(A_1^D) &= P^D(1|0\times 0)\\
\Pr(C_{0011}^D) &= P^D(0|1\times 0) - P^D(0|1\times 1)\\
\Pr(C_{0101}^D) &= P^D(0|0\times 1) - P^D(0|1\times 1)\\
\Pr(C_{0111}^D) &= P^D(0|0\times 0) + P^D(0|1\times 1) - P^D(0|1\times 0) - P^D(0|0\times 1),
\end{aligned}$$

and the model has three testable implications:

$$\begin{aligned}
P^D(0|1\times 0) &\geqslant P^D(0|1\times 1),\\
P^D(0|0\times 1) &\geqslant P^D(0|1\times 1),\\
P^D(0|0\times 0) + P^D(0|1\times 1) &\geqslant P^D(0|1\times 0) + P^D(0|0\times 1).
\end{aligned}$$

(ii) The LATEs on the three groups of compliers are point-identified by

$$\begin{aligned}
\mathbb{E}(Y_i^D(1) - Y_i^D(0)|i\in C_{0101}^D) &= \frac{\mathbb{E}(Y|Z=1\times 1) - \mathbb{E}(Y|Z=0\times 1)}{P^D(1|1\times 1) - P^D(1|0\times 1)}\\
\mathbb{E}(Y_i^D(1) - Y_i^D(0)|i\in C_{0011}^D) &= \frac{\mathbb{E}(Y|Z=1\times 1) - \mathbb{E}(Y|Z=1\times 0)}{P^D(1|1\times 1) - P^D(1|1\times 0)}\\
\mathbb{E}(Y_i^D(1) - Y_i^D(0)|i\in C_{0111}^D) &= \\
&\frac{\mathbb{E}(Y|Z=1\times 0) + \mathbb{E}(Y|Z=0\times 1) - \mathbb{E}(Y|Z=1\times 1) - \mathbb{E}(Y|Z=0\times 0)}{P^D(1|1\times 0) + P^D(1|1\times 0) - P^D(1|1\times 1) - P^D(1|0\times 0)}.
\end{aligned}$$

Proposition 14 states that (i) the average treatment effects for reluctant compliers are identified by suitable Wald statistics and that (ii) the average treatment effect for eager compliers is identified by a ratio between difference-in-differences (DiD) population quantities. The latter estimand can be viewed as a two-dimensional version of Wald statistics.

4 Empirical Examples

4.1 The Student Achievement and Retention Project

In this section, we revisit Angrist, Lang, and Oreopoulos (2009), who analyzed the Student Achievement and Retention Project. STAR was a randomized evaluation of academic services and incentives for college freshmen at a Canadian university. It was a factorial design, with

two binary instruments. The Student Fellowship Program (SFP) offered students the chance to win merit scholarships for good grades in the first year; the Student Support Program (SSP) offered students access to both a peer-advising service and a supplemental instruction service. Entering first-year undergraduates were randomly assigned to one of four groups: a control group ($z = 0 \times 0$), SFP only ($z = 0 \times 1$), SSP only ($z = 1 \times 0$), and an intervention offering both (SFSP, $z = 1 \times 1$).

The data indicates whether a student who was offered a program signed up, and whether those who were offered SSP or SFSP and signed up made use of the services of SSP. Angrist, Lang, and Oreopoulos (2009) used the sign-up as the treatment variable. They comment that "in the SSP and SFSP, a further distinction can be made between compliance via sign up and compliance via service use" (p. 147). Many students who sign up did not in fact use the services; this suggests defining an unfiltered treatment variable as a pair $T_i = (A_i, S_i)$, where $A_i = 1$ (for "accept") denotes that student i signed up and $S_i = 1$ (for "services") that (s)he used the services provided by SSP.

Obviously, $S_i = 0$ if $A_i = 0$. Hence T_i can only take three values: $(0,0), (1,0)$, and $(1,1)$. With a slight change in notation, we model the choice as

$$T_i(z) = \arg\max \left(u_i(0,0), \Delta_z^T(1,0) + u_i(1,0), \Delta_z^T(1,1) + u_i(1,1) \right).$$

While there are four instrument values and three treatment values, this is in fact a 3×3 model, with some specific features. First note that $T_i = 0$ for all observations in the control group; this allows us to set $\Delta_{0\times0}^T(1,0)$ and $\Delta_{0\times0}^T(1,1)$ at minus infinity. In addition, S_i can only be zero if $z = 0 \times 1$, so that we can set $\Delta_{0\times1}^T(1,0)$ and $\Delta_{0\times1}^T(1,1)$ at minus infinity too. As a consequence, we do not lose any information by redefining the control group to be $0 \equiv \{0 \times 0, 0 \times 1\}$.

In addition, students should be more likely to sign up under $z = 1 \times 1$ than under $z = 1 \times 0$, as the former adds the lure of a fellowship. We will also assume that it makes them more likely to use the services—an assumption that we will test below. Then both treatment values $(1,0)$ and $(1,1)$ are targeted by 1×1, but they cannot be strictly targeted. Take for instance $\bar{Z}(1,1) = \{1 \times 1\}$; strict targeting would require $\Delta_{1\times0}^T(1,1) = \Delta_0^T(1,1)$, which is minus infinity.

Rather than to pursue with the unfiltered treatment model, let us move on to filtered models. In our terminology, Angrist, Lang, and Oreopoulos (2009) chose to use a particular filter $M(A,S) = A$, which is close to intent-to-treat as they point out. Here we take

$M(A,S) = S$ instead: we define

$$D_i(z) = S_i(z) = \mathbb{1}\left(\Delta_z^T(1,1) + u_i(1,1) > \max(u_i(0,0), \Delta_z^T(1,0) + u_i(1,0))\right). \tag{4.1}$$

Since the SFP incentives applied to the first year grades only, we take the grades in the second year as our outcome variable Y_i.

Equation (4.1) has a similar structure to the double hurdle model of Equation (3.9). The models are quite different, however. This new filtered model has $D_i(0) = 0$ with probability one; and we are assuming that an offer of a fellowship makes students more likely to use the services. Of the a priori four possible response-groups $C^D_{0,d,d'}$ for $d, d' = 0, 1$, this assumption eliminates one: if $D_i(1 \times 1) = 0$ then a fortiori $D_i(1 \times 0) = 0$. This leaves three groups: the never-takers A^D_0, and two groups of compliers C^D_{001} and C^D_{011}. The group C^D_{001} consists of reluctant compliers, who only use SSP if it is offered along with SFP. Those in C^D_{011} are eager compliers: they use SSP whenever it is offered to them with or without a fellowship.

Remember that $P^D(1|z) := \Pr(D_i = 1|Z_i = z)$ for $z = 0, 1 \times 0, 1 \times 1$. Then $P^D(1|0) = 0$ and the proportions of the three response-groups are given by

$$\begin{aligned}
\Pr(A^D_0) &= 1 - P^D(1|1 \times 1) \\
\Pr(C^D_{001}) &= P^D(1|1 \times 1) - P^D(1|1 \times 0) \\
\Pr(C^D_{011}) &= P^D(1|1 \times 0).
\end{aligned}$$

Note that given Equation (4.1),

$$\begin{aligned}
P^D(1|1 \times 0) = \Pr\big(&u_i(1,1) - u_i(0,0) > -\Delta^T_{1\times 0}(1,1) \\
&\text{and } \; u_i(1,1) - u_i(1,0) > \Delta^T_{1\times 0}(1,0) - \Delta^T_{1\times 0}(1,1)\big) \\
P^D(1|1 \times 1) = \Pr\big(&u_i(1,1) - u_i(0,0) > -\Delta^T_{1\times 1}(1,1) \\
&\text{and } \; u_i(1,1) - u_i(1,0) > \Delta^T_{1\times 1}(1,0) - \Delta^T_{1\times 1}(1,1)\big).
\end{aligned}$$

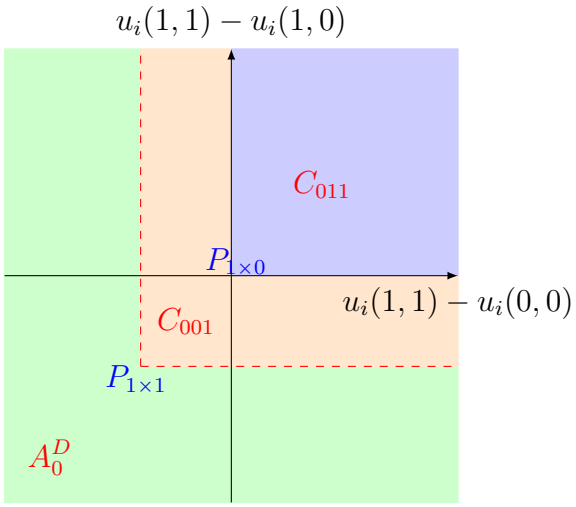
Our assumption that students are more likely to use the services under SFSP translates into

$$\Delta^T_{1\times 1}(1,1) > \Delta^T_{1\times 0}(1,1) \;\text{ and }\; \Delta^T_{1\times 1}(1,0) - \Delta^T_{1\times 1}(1,1) < \Delta^T_{1\times 0}(1,0) - \Delta^T_{1\times 0}(1,1).$$

Figure 9 illustrates a configuration in which these inequalities hold, where

$$P_{1\times 0} = (-\Delta^T_{1\times 0}(1,1), \Delta^T_{1\times 0}(1,0) - \Delta^T_{1\times 0}(1,1)) \text{ and } P_{1\times 1} = (-\Delta^T_{1\times 1}(1,1), \Delta^T_{1\times 1}(1,0) - \Delta^T_{1\times 1}(1,1)).$$

Figure 9: STAR example



Under our assumptions, it is straightforward to show that

$$\mathbb{E}[Y_i^D|Z_i = 1\times 0] - \mathbb{E}[Y_i^D|Z_i = 0] = \mathbb{E}[Y_i^D(1) - Y_i^D(0)|i \in C_{011}^D]\Pr(i \in C_{011}^D),$$
$$\mathbb{E}[Y_i^D|Z_i = 1\times 1] - \mathbb{E}[Y_i^D|Z_i = 1\times 0] = \mathbb{E}[Y_i^D(1) - Y_i^D(0)|i \in C_{001}^D]\Pr(i \in C_{001}^D).$$

Therefore, we have

$$\mathbb{E}[Y_i^D(1) - Y_i^D(0)|i \in C_{011}^D] = \frac{\mathbb{E}[Y_i^D|Z_i = 1\times 0] - \mathbb{E}[Y_i^D|Z_i = 0]}{P^D(1|1\times 0)},$$
$$\mathbb{E}[Y_i^D(1) - Y_i^D(0)|i \in C_{001}^D] = \frac{\mathbb{E}[Y_i|Z_i = 1\times 1] - \mathbb{E}[Y_i^D|Z_i = 1\times 0]}{P^D(1|1\times 1) - P^D(1|1\times 0)}.$$

Since $\Pr(D_i = 1|Z_i = 0) = 0$, the first estimand is the two-stage least squares formula of Bloom (1984); the second estimand is the LATE formula of Imbens and Angrist (1994).

Table 4 reports estimation results. We only focus on the subsample of women since the STAR program had no effect on men. Panel A of Table 4 shows the estimated proportions of the two complier groups: 0.288 for C_{011}^D and 0.245 for C_{001}^D. The majority group is the never-takers whose share is 0.467. This is because the usage of SSP was low. Panel B reveals remarkable heterogeneity between the two complier groups. We do not find any significant treatment effect for C_{011}^D, whereas we do find sizeable and significant impact on probation/withdrawal and good standing for C_{001}^D.[15] As can be seen in Figure 9, C_{001}^D is

[15]The point estimates for probation/withdrawal and good standing are very large in absolute value; however, the standard errors are large as well, resulting in wide confidence intervals. This is partially because the sample size is relatively small and partially because the estimand is the ratio of two population quantities with the small denominator.

Table 4: Empirical Results from STAR

Panel A.	Proportion of Compliers
$\Pr(i \in C^D_{011})$	0.288
	(0.040)
$\Pr(i \in C^D_{001})$	0.245
	(0.069)

Panel B.	GPA	On probation or withdrawal	Good standing	Credits earned
$\mathbb{E}[Y_i\|Z_i = 1 \times 0] - \mathbb{E}[Y_i\|Z_i = 0]$	0.084	0.045	-0.039	-0.065
	(0.088)	(0.043)	(0.046)	(0.147)
$\mathbb{E}[Y_i(1) - Y_i(0)\|i \in C^D_{011}]$	0.291	0.156	-0.137	-0.225
	(0.303)	(0.152)	(0.161)	(0.516)
$\mathbb{E}[Y_i\|Z_i = 1 \times 1] - \mathbb{E}[Y_i\|Z_i = 1 \times 0]$	0.186	-0.141	0.163	0.350
	(0.127)	(0.058)	(0.065)	(0.208)
$\mathbb{E}[Y_i(1) - Y_i(0)\|i \in C^D_{001}]$	0.758	-0.576	0.664	1.427
	(0.532)	(0.265)	(0.305)	(0.887)

Notes. Standard errors are in the parentheses. The estimation sample consists of women in the control, SSP and SFSP groups. The sample size is $n = 837$. The outcome variables are second year GPA, an indicator of probation or withdrawal in the second year, a "good standing" variable that indicates whether students attempted at least four credits and were not on probation, and the credits earned. Estimates of treatment effects are computed based on linear regression models using the full set of controls used in Angrist, Lang, and Oreopoulos (2009).

closer to the group of never-takers: they have higher unobserved disutilities of using academic support services than those in C^D_{011}. However, those in C^D_{001} reaped greater benefits of using the SSP by avoiding probation or withdrawal in the second year.

The main parameter of interest in Angrist, Lang, and Oreopoulos (2009) was the intent-to-treat (ITT) effect of the SFSP program: $\mathbb{E}[Y_i|Z_i = 1 \times 1] - \mathbb{E}[Y_i|Z_i = 0 \times 0]$ in our notation. Our analysis suggests that the ITT effect of the SFSP program is a mix of two very different treatment effects. This highlights the importance of unbundling heterogeneous complier groups.

4.2 Head Start

Let us now reexamine Kline and Walters's (2016) analysis of the Head Start Impact Study (HSIS) using our framework. The structure of HSIS is identical to that of Example 2. The

treatments consist of no preschool (n), Head Start (h), and other preschool centers (c): $\mathcal{T} = \{n, h, c\}$. We will take $t_0 = n$ as our reference treatment. The instrument is binary, with a control group ($z = 0$) and a group that is offered admission to Head Start ($z = 1$). This gives five response groups: $A_n = C_{nn}$, $A_c = C_{cc}$, $A_h = C_{hh}$, C_{nh}, and C_{ch}. The first three groups are always-takers and the last two groups are compliers.

4.2.1 Group proportions and counterfactual means

Their proportions in the sample are given by (2.2) in Proposition 6; 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 as per Proposition 8. Among those that are point-identified, the average test scores are the highest for the groups who always choose other preschool centers (about 0.3 standard deviation). There is a noticeable difference between the two complier groups: $\mathbb{E}[Y_i(n)|i \in C_{nh}]$ is negative, but $\mathbb{E}[Y_i(c)|i \in C_{ch}]$ is above 0.1 standard deviation. This indicates that among compliers, the children who used other centers had higher scores than those who stayed at home. Head Start may therefore attract some children who already were at a good preschool. Kline and Walters (2016) call this pattern the "substitution effect" of Head Start. However, the way we achieve the identification of $\mathbb{E}[Y_i(n)|i \in C_{nh}]$ and $\mathbb{E}[Y_i(c)|i \in C_{ch}]$ is new.

4.2.2 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, under Proposition 8, they are only partially identified by

$$\begin{aligned}
&\mathbb{E}\left[Y_i(h)|i \in C_{nh}\right]\left\{\Pr(T_i = n|Z_i = 0) - \Pr(T_i = n|Z_i = 1)\right\} \\
&+ \mathbb{E}\left[Y_i(h)|i \in C_{ch}\right]\left\{\Pr(T_i = c|Z_i = 0) - \Pr(T_i = c|Z_i = 1)\right\} \\
&= \mathbb{E}\left[Y_i \mathbb{1}(T_i = h)|Z_i = 1\right] - \mathbb{E}\left[Y_i \mathbb{1}(T_h = 1)|Z_i = 0\right].
\end{aligned}$$

This is exactly the formula on Kline and Walters (2016, pp.1811), where they point out that the LATE for Head Start is a weighted average of "subLATEs" with weights S_c and $(1 - S_c)$ with

$$S_c := \frac{\Pr(C_{ch})}{\Pr(C_{nh}) + \Pr(C_{ch})} = \frac{\Pr(T_i = c|Z_i = 0) - \Pr(T_i = c|Z_i = 1)}{\Pr(T_i \neq h|Z_i = 0) - \Pr(T_i \neq h|Z_i = 1)}.$$

Kline and Walters (2016) first tried to estimate $\mathbb{E}[Y_i(h) - Y_i(c)|i \in C_{ch}]$ and $\mathbb{E}[Y_i(h) - Y_i(n)|i \in C_{nh}]$ separately using two-stage least squares (TSLS), using interaction of the instrument with covariates or experimental sites in an attempt to generate enough variation. They

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 6			
Always — no preschool (A_n)	0.092	0.099	0.095
Always — Head Start (A_h)	0.147	0.122	0.136
Always — other centers (A_c)	0.058	0.114	0.083
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 8			
$\mathbb{E}[Y_i(n)\vert \in A_n]$	-0.050	-0.017	-0.035
$\mathbb{E}[Y_i(h)\vert \in A_h]$	0.007	-0.080	-0.028
$\mathbb{E}[Y_i(c)\vert \in A_c]$	0.293	0.330	0.316
$\mathbb{E}[Y_i(n)\vert i \in C_{nh}]$	-0.027	-0.116	-0.062
$\mathbb{E}[Y_i(c)\vert i \in C_{ch}]$	0.112	0.144	0.129
Panel C. Counterfactual Means of Test Scores via Corollary 2			
$\mathbb{E}[Y_i(h)\vert i \in C_{nh}] = \mathbb{E}[Y_i(h)\vert i \in C_{ch}]$	0.252	0.169	0.216
Panel D. Treatment Effects via Corollary 2			
$\mathbb{E}[Y_i(h) - Y_i(n)\vert i \in C_{nh}]$ for compliers from 'n' to 'h'	0.279	0.285	0.278
	(0.063)	(0.076)	(0.050)
$\mathbb{E}[Y_i(h) - Y_i(c)\vert i \in C_{ch}]$ for compliers from 'c' to 'h'	0.140	0.025	0.087
	(0.089)	(0.097)	(0.063)
$\mathbb{E}[Y_i(h) - Y_i(n)\vert i \in C_{nh}] - \mathbb{E}[Y_i(h) - Y_i(c)\vert i \in C_{ch}]$	0.139	0.260	0.191
	(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.

acknowledged the limitations of this interacted TSLS approach and developed a parametric selection model à la Heckman (1979). Using a parametric selection model and pooled cohorts, Kline and Walters (2016, Table VIII, column (4) full model) obtain estimates of the treatment effect of $0.370\,(0.088)$ for C_{nh} and $-0.093\,(0.154)$ for C_{ch} respectively (standard errors in parentheses).

Our Corollary 2 provides an alternative approach to separating the two treatment effects. If we assume that $\mathbb{E}\left[Y_i(h)|i \in C_{nh}\right] = \mathbb{E}\left[Y_i(h)|i \in C_{ch}\right]$, we can point-identify the average treatment effects for both groups of compliers. The resulting estimates are shown in Panels C and D of Table 5. The average impact on test scores of participating in Head Start is around 0.28 for C_{nh}, whereas it is smaller and insignificant for C_{ch}. Their difference is significantly different when the two cohorts are pooled together.

We obtained these estimates of the treatment effects by a completely different route than Kline and Walters (2016). While the two sets of estimates are similar, our estimate of the difference between the treatment effects on the two groups of compliers is twice smaller. Our homogeneity assumption $\mathbb{E}[Y_i(h)|i \in C_{nh}] = \mathbb{E}[Y_i(h)|i \in C_{ch}]$ may be too strong. It might be more plausible to assume that

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

as children who would not attend preschool in the absence of offer to Head Start are likely to be less well-prepared than children who would attend other preschools. Then our estimated difference between the two complier groups will be a lower bound of the true difference.

Concluding Remarks

We have shown that our targeting and filtering concepts are a useful way to analyze models with multivalued treatments and multivalued instruments. While our characterization is sharpest under strict, one-to-one targeting (Corollary 1), our framework remains useful even without strict targeting.

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 filtering 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) can be seen as analyzing continuous-instruments/discrete-treatments filtered models; so does Mountjoy's (2019)'s study of 2-year colleges. Extending our analysis to models with continuous treatments is an interesting topic for further research.

Appendices

A Proofs for Section 2

Proof of Proposition 1. Let $T_i(\bar{z}(t)) = 0$ for some $t \in \mathcal{T}^*$. Then $u_{i0} > \bar{\Delta}_t + u_{it}$. However, if $z \neq \bar{z}(t)$ then $\bar{\Delta}_t > \Delta_z(t)$ under Assumption 5. Therefore $u_{i0} > \Delta_z(t) + u_{it}$, and $T_i(z)$ cannot be t. □

Proof of Lemma 1. The lemma is proved in the main text. □

Proof of Proposition 2. Take any observation i and an instrument value $z \in \mathcal{Z}$. The treatment $T_i(z)$ must maximize $(U_z(t) + u_{it})$ over $t \in \mathcal{T}$. Under Assumption 6, for any t we have

- $U_z(t) = U_z(0) + \bar{\Delta}_t$ if $t \in \bar{T}(z)$
- $U_z(t) = U_z(0) + \underline{\Delta}_t$ otherwise.

Therefore, eliminating $U_z(0)$,

$$T_i(z) \in \arg\max \left(\max_{t \notin \bar{T}(z)} (\underline{\Delta}_t + u_{it}), \max_{t \in \bar{T}(z)} (\bar{\Delta}_t + u_{it}) \right). \tag{A.1}$$

Since $\bar{\Delta}_t \geqslant \underline{\Delta}_t$ for all $t \in \mathcal{T}$, a fortiori $\bar{\Delta}_t + u_{it} \geqslant \underline{\Delta}_t + u_{it}$ when $t \in \bar{T}(z)$. As a consequence, we can rewrite Equation (A.1) as

$$T_i(z) \in \arg\max \left(\Delta_i^*,\ V_i^*(z) \right).$$

(i) If $z \in \mathcal{Z}^*$, then $\bar{T}(z)$ is not empty and the maximizer can be either in τ_i^* or in $T_i^*(z)$.

(ii) If $z \in \mathcal{Z} \setminus \mathcal{Z}^*$, then z can only be 0. $\bar{T}(0) = \varnothing$ and $T_i(0)$ can only be in τ_i^*.

□

Proof of Proposition 3. Take an observation i and define $A_i = \{z \in \mathcal{Z}^* \mid T_i(z) = T_i^*(z)\}$.

(i) By definition, $A_i \subset \mathcal{Z}^*$; therefore $A_i \neq \mathcal{Z}$.

(ii) If $z \in \mathcal{Z}^* \backslash A_i$, then by construction $T_i(z) \neq T_i^*(z)$. By Proposition 2(i), $T_i(z)$ can only be τ_i^*. If $z \notin \mathcal{Z}^*$, then $z = 0$ and we know from Proposition 2(ii) that $T_i(0) = \tau_i^*$.

(iii) Assume that $\tau_i^* = \tau \in \mathcal{T}^*$. Then $\bar{Z}(\tau) \neq \varnothing$. For any z in $\bar{Z}(\tau)$,

$$V_i^*(z) \geqslant \bar{\Delta}_\tau + u_{i\tau} > \underline{\Delta}_\tau + u_{i\tau} = \Delta_i^*;$$

therefore $z \in A_i$. This proves that $\bar{Z}(\tau) \subset A_i$.

□

Proof of Corollary 1. It follows directly from Proposition 3. □

Proof of Proposition 4. The set A of Corollary 1 must be a subset of $\mathcal{Z}^*$. For each such subset, τ can take any value in $\mathcal{T} \setminus \mathcal{T}^*$; and if $\tau \in \mathcal{T}^*$ then τ must be in A. Each subset A of $\mathcal{Z}^*$ with a elements therefore allows for $(a + |\mathcal{T}| - |\mathcal{T}^*|)$ values of τ. This gives a total of

$$\sum_{a=0}^{|\mathcal{Z}^*|} \binom{|\mathcal{Z}^*|}{a} (a + |\mathcal{T}| - |\mathcal{T}^*|)$$

response-types. Moreover, we know that $|\mathcal{T}^*| = |\mathcal{Z}^*|$ under one-to-one targeting. Using the identities

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

we obtain a total of $(2|\mathcal{T}| - |\mathcal{Z}^*|) \times 2^{|\mathcal{Z}^*|-1}$ types. □

Proof of Proposition 5. Take $z \in \mathcal{Z}$ and $t \in \mathcal{T}$, and let observation i belong to a class $c(A, \tau)$ for some $A \subset \mathcal{Z}^*$ and $\tau \in \bar{t}(A) \bigcup (\mathcal{T} \setminus \mathcal{T}^*)$. There are only two ways to obtain $T_i(z) = t$:

- if $z \notin A$, then $T_i(z) = \tau$; therefore $\tau = t$. Summing over all subsets A of $\mathcal{Z}^*$ that exclude z gives the first term of (2.1).
- if $z \in A$ (which implies $z \in \mathcal{Z}^*$), we know that $T_i(z) = \bar{t}(z)$ no matter what the value of τ is; hence $t = \bar{t}(z)$. Summing over all subsets A that include z and all values of $\tau \in \bar{t}(A) \bigcup (\mathcal{T} \setminus \mathcal{T}^*)$ gives the second line in (2.1).

□

Proof of Proposition 6. We apply equation (2.1) to a treatment value $t > 0$. With $\mathcal{Z}^* = \{1\}$, we can only have $A = \varnothing$ or $A = \{1\}$. The first line of equation (2.1) gives zero if $z = 1$, and for $z = 0$:

$$P(t|0) = \mathbb{1}(t \neq 1) \Pr(c(\varnothing, t)) + \Pr(c(\{1\}, 1)).$$

and

$$P(t|1) = \mathbb{1}(t \neq 1) \Pr(c(\varnothing, t)) + \mathbb{1}(t = 1) \sum_{\tau \in \mathcal{T}} \Pr(c(\{1\}, \tau)).$$

We already know that $c(\varnothing, t)$ is A_t and $c(\{1\}, \tau)$ is A_1 if $\tau = 1$ and the complier group $C_{\tau 1}$ otherwise. Therefore for $t = 1$

$$P(1|1) = \Pr(A_1) + \sum_{\tau \neq 1} \Pr(C_{\tau 1})$$

and $P(1|0) = \Pr(A_1)$; while for $t \neq 1$ we have $P(t|1) = \Pr(A_t)$ and $P(t|0) = \Pr(C_{t1}) + \Pr(A_t)$. □

Proof of Proposition 7. It is straightforward from Figure 6 and Table 2. □

Proof of Lemma 2. 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 $C_{(z)} = t$, 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 8. By Lemma 2, we get

$$\begin{aligned} \bar{E}_0(1) &= \mathbb{E}\left[Y_i(1)|i \in A_1\right] \Pr(i \in A_1) \\ \bar{E}_0(t) &= \mathbb{E}\left[Y_i(t)|i \in A_t\right] \Pr(i \in A_t) \\ &\quad + \mathbb{E}\left[Y_i(t)|i \in C_{t1}\right] \Pr(i \in C_{t1}) \text{ for } t \neq 1, \\ \bar{E}_1(1) &= \mathbb{E}\left[Y_i(1)|i \in A_1\right] \Pr(i \in A_1) \\ &\quad + \sum_{t \neq 1} \mathbb{E}\left[Y_i(1)|i \in C_{t1}\right] \Pr(i \in C_{t1}), \\ \bar{E}_1(t) &= \mathbb{E}\left[Y_i(t)|i \in A_t\right] \Pr(i \in A_t) \text{ for } t \neq 1. \end{aligned}$$

Since Proposition 6 identifies all type probabilities, the first and fourth equations give directly $\mathbb{E}(Y_i(t)|i \in A_t)$ for all t. Then the second equation identifies $\mathbb{E}(Y_i(t)|i \in C_{t1})$ for $t \neq 1$.

However, the values $\mathbb{E}(Y_i(1)|i \in C_{t1})$ for $t \neq 1$ are only linked by

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

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 6 and Lemma 2 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. The corollary follows directly from the proof of Proposition 8, as

$$\sum_{t\neq 1}\Pr(i \in C_{t1}) = \sum_{t\neq 1}(P(t|0) - P(t|1)) = P(1|1) - P(1|0)$$

gives $\mathbb{E}(Y_i(1)|i \in C_{t1}) = (\bar{E}_1(1) - \bar{E}_0(1))/(P(1|1) - P(1|0))$. □

Proof of Proposition 9. By Lemma 2, we obtain

$$\begin{aligned}\bar{E}_2(1) &= \mathbb{E}\left[Y_i(1)|i \in A_1\right]\Pr(i \in A_1),\\ \bar{E}_1(2) &= \mathbb{E}\left[Y_i(2)|i \in A_2\right]\Pr(i \in A_2),\\ \bar{E}_0(0) - \bar{E}_1(0) &= \mathbb{E}\left[Y_i(0)|i \in C_{010}\bigcup C_{012}\right]\Pr(i \in C_{010}\bigcup C_{012}),\\ \bar{E}_0(0) - \bar{E}_2(0) &= \mathbb{E}\left[Y_i(0)|i \in C_{002}\bigcup C_{012}\right]\Pr(i \in C_{002}\bigcup C_{012}),\\ \bar{E}_1(1) - \bar{E}_0(1) &= \mathbb{E}\left[Y_i(1)|i \in C_{010}\bigcup C_{012}\bigcup C_{212}\right]\Pr(i \in C_{010}\bigcup C_{012}\bigcup C_{212}),\\ \bar{E}_0(1) - \bar{E}_2(1) &= \mathbb{E}\left[Y_i(1)|i \in C_{112}\right]\Pr(i \in C_{112}),\\ \bar{E}_2(2) - \bar{E}_0(2) &= \mathbb{E}\left[Y_i(2)|i \in C_{002}\bigcup C_{012}\bigcup C_{112}\right]\Pr(i \in C_{002}\bigcup C_{012}\bigcup C_{112}),\\ \bar{E}_0(2) - \bar{E}_1(2) &= \mathbb{E}\left[Y_i(2)|i \in C_{212}\right]\Pr(i \in C_{212}).\end{aligned}$$

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

Proof of Corollary 3. The desired results follow from Proposition 9 under assumption 9. □

Proof of Proposition 10. Using counterfactual notation, write

$$Y_i = Y_i(0) + \sum_{t=1}^{2} [Y_i(t) - Y_i(0)] \mathbb{1}(T_i = t) \tag{A.2}$$

and

$$\mathbb{1}(T_i = t) = \mathbb{1}(T_i(0) = t) + \sum_{z=1}^{2} [\mathbb{1}(T_i(z) = t) - \mathbb{1}(T_i(0) = t)] \mathbb{1}(Z_i = z). \tag{A.3}$$

Combining equation (A.2) with equation (A.3) gives

$$\begin{aligned}
U_i &= Y_i - \beta_0 - \sum_{t=1}^{2} \beta_t \mathbb{1}(T_i = t) \\
&= [Y_i(0) - \beta_0] + \sum_{t=1}^{2} [Y_i(t) - Y_i(0) - \beta_t] \mathbb{1}(T_i = t) \\
&= [Y_i(0) - \beta_0] + \sum_{t=1}^{2} [Y_i(t) - Y_i(0) - \beta_t] \mathbb{1}(T_i(0) = t) \\
&+ \sum_{t=1}^{2} [Y_i(t) - Y_i(0) - \beta_t] \sum_{z=1}^{2} [\mathbb{1}(T_i(z) = t) - \mathbb{1}(T_i(0) = t)] \mathbb{1}(Z_i = z).
\end{aligned}$$

First, note that using the fact that Z_i is independent of $\{Y_i(t), T_i(z) : t, z = 0, 1, \ldots, T-1\}$,

$$\begin{aligned}
0 &= \mathbb{E}[\mathbb{1}(Z_i = 0) U_i] \\
&= \mathbb{E}\big[[Y_i(0) - \beta_0] \mathbb{1}(Z_i = 0)\big] + \sum_{t=1}^{2} \mathbb{E}\big[[Y_i(t) - Y_i(0) - \beta_t] \mathbb{1}(T_i(0) = t) \mathbb{1}(Z_i = 0)\big] \\
&= \mathbb{E}[Y_i(0) - \beta_0] \Pr(Z_i = 0) + \sum_{t=1}^{2} \mathbb{E}\big[[Y_i(t) - Y_i(0) - \beta_t] \mathbb{1}(T_i(0) = t)\big] \Pr(Z_i = 0),
\end{aligned}$$

which implies that

$$0 = \mathbb{E}[Y_i(0) - \beta_0] + \sum_{t=1}^{2} \mathbb{E}\big[[Y_i(t) - Y_i(0) - \beta_t] \mathbb{1}(T_i(0) = t)\big]. \tag{A.4}$$

Now, for $t = 1, 2$,

$$\begin{aligned}
0 &= \mathbb{E}[\mathbb{1}(Z_i = t)U_i] \\
&= \mathbb{E}\big[[Y_i(0) - \beta_0]\mathbb{1}(Z_i = t)\big] + \sum_{j=1}^{2} \mathbb{E}\big[[Y_i(j) - Y_i(0) - \beta_j]\mathbb{1}(T_i(0) = j)\mathbb{1}(Z_i = t)\big] \\
&+ \sum_{j=1}^{2} \mathbb{E}\big[[Y_i(j) - Y_i(0) - \beta_j][\mathbb{1}(T_i(t) = j) - \mathbb{1}(T_i(0) = j)]\mathbb{1}(Z_i = t)\big],
\end{aligned}$$

which implies that for $t = 1, 2$,

$$\begin{aligned}
0 &= \mathbb{E}[Y_i(0) - \beta_0] + \sum_{t=1}^{2} \mathbb{E}\big[[Y_i(t) - Y_i(0) - \beta_t]\mathbb{1}(T_i(0) = t)\big] \\
&+ \sum_{j=1}^{2} \mathbb{E}\big[[Y_i(j) - Y_i(0) - \beta_j][\mathbb{1}(T_i(t) = j) - \mathbb{1}(T_i(0) = j)]\big] \\
&= \sum_{j=1}^{2} \mathbb{E}\big[[Y_i(j) - Y_i(0) - \beta_j][\mathbb{1}(T_i(t) = j) - \mathbb{1}(T_i(0) = j)]\big],
\end{aligned}$$

where the last equality follows from equation (A.4).

First, consider $t = 1$. Then, we have that

$$\begin{aligned}
0 = \mathbb{E}\big[[Y_i(1) - Y_i(0) - \beta_1][\mathbb{1}(T_i(1) = 1) - \mathbb{1}(T_i(0) = 1)]\big] \\
+ \mathbb{E}\big[[Y_i(2) - Y_i(0) - \beta_2][\mathbb{1}(T_i(1) = 2) - \mathbb{1}(T_i(0) = 2)]\big].
\end{aligned}$$

Using the facts that (i) $[\mathbb{1}(T_i(1) = 1) - \mathbb{1}(T_i(0) = 1)]$ can be nonzero with a value of $+1$ if and only if $i \in C_{010} \bigcup C_{012} \bigcup C_{212}$ and (ii) $[\mathbb{1}(T_i(1) = 2) - \mathbb{1}(T_i(0) = 2)]$ can be nonzero with a value of -1 if and only if $i \in C_{212}$, we obtain

$$\begin{aligned}
&\{\beta_1 - \mathbb{E}[Y_i(1) - Y_i(0) | i \in C_{010} \bigcup C_{012} \bigcup C_{212}]\} \Pr(i \in C_{010} \bigcup C_{012} \bigcup C_{212}) \\
&= \{\beta_2 - \mathbb{E}[Y_i(2) - Y_i(0) | i \in C_{212}]\} \Pr(i \in C_{212}).
\end{aligned} \tag{A.5}$$

We now consider $t = 2$. Similarly, we have that

$$\begin{aligned}
0 = \mathbb{E}\big[[Y_i(1) - Y_i(0) - \beta_1][\mathbb{1}(T_i(2) = 1) - \mathbb{1}(T_i(0) = 1)]\big] \\
+ \mathbb{E}\big[[Y_i(2) - Y_i(0) - \beta_2][\mathbb{1}(T_i(2) = 2) - \mathbb{1}(T_i(0) = 2)]\big].
\end{aligned}$$

Using the facts that (i) $[\mathbb{1}(T_i(2) = 1) - \mathbb{1}(T_i(0) = 1)]$ can be nonzero with a value of -1 if and only if $i \in C_{112}$ and (ii) $[\mathbb{1}(T_i(2) = 2) - \mathbb{1}(T_i(0) = 2)]$ can be nonzero with a value of $+1$

if and only if $i \in\in C_{002} \bigcup C_{012} \bigcup C_{112}$, we obtain

$$\begin{aligned} &\{\beta_2 - \mathbb{E}[Y_i(2) - Y_i(0)|i \in C_{002} \bigcup C_{012} \bigcup C_{112}]\} \Pr(i \in C_{002} \bigcup C_{012} \bigcup C_{112}) \\ &= \{\beta_1 - \mathbb{E}[Y_i(1) - Y_i(0)|i \in C_{112}]\} \Pr(i \in C_{112}). \end{aligned} \tag{A.6}$$

The conclusion of the proposition follows from equations (A.5) and (A.6). □

Proof of Corollary 4. The corollary follows directly from Proposition 10 under equations (2.9) and (2.10). □

B Proofs for Section 3

Proof of Proposition 11. (i) It follows directly from Proposition 6 and from the mapping of types.

(ii) From Proposition 8, we have

$$\mathbb{E}(Y_i^D(1)|i \in A_1^D) = \mathbb{E}(Y_i^T(1)|i \in A_1^T) = \frac{\bar{E}_0^T(1)}{P^T(1|0)} = \frac{\bar{E}_0^D(1)}{P^D(1|0)}.$$

Moreover,

$$\begin{aligned} \mathbb{E}(Y_i^D(0)|i \in A_0^D) &= \mathbb{E}(Y_i^D(0)|i \in \bigcup_{t \neq 1} A_t^T) \\ &= \sum_{t \neq 1} \mathbb{E}(Y_i^T(t)|i \in A_t^T) \frac{P^T(t|1)}{1 - P^T(1|1)} \\ &= \sum_{t \neq 1} \frac{\bar{E}_1^T(t)}{1 - P^T(1|1)} \\ &= \frac{\bar{E}_1^D(0)}{1 - P^D(1|1)}. \end{aligned}$$

(iii) Now consider the weighted LATE $\sum_{t \neq 1} \alpha_t^T \mathbb{E}(Y_i^T(1) - Y_i^T(t)|i \in C_{t1}^T)$, which is identified in the unfiltered treatment model (equation 2.5). The weights $\alpha_t^T = (P^T(t|0) - P^T(t|1))/(P^T(1|1) - P^T(1|0))$ are not identified any more. Note however that for any variable W_i,

$$\sum_{t \neq 1} \alpha_t^T \mathbb{E}(W_i|i \in C_{t1}) = \mathbb{E}(W_i|i \in C_{01}^D);$$

therefore $\sum_{t \neq 1} \alpha_t^T \mathbb{E}(Y_i^D(1)|i \in C_{t1}^T) = \mathbb{E}(Y_i^D(1)|i \in C_{01}^D)$. The LHS of Equation (2.5)

becomes

$$\mathbb{E}(Y_i^D(1)|i \in C_{01}^D) - \sum_{t \neq 1} \alpha_t^T \mathbb{E}(Y_i^T(t)|i \in C_{t1}).$$

On the RHS we had

$$\frac{(\bar{E}_1^T(1) - \bar{E}_0^T(1)) - \sum_{t \neq 1}(\bar{E}_0^T(t) - \bar{E}_1^T(t))}{P^T(1|1) - P^T(1|0)}.$$

The denominator is still identified as $P^D(1|1) - P^D(1|0)$, as is the first term of the numerator, which equals $\bar{E}_1^D(1) - \bar{E}_0^D(1)$. From equation 3.3,

$$\sum_{t \neq 1}(\bar{E}_0^T(t) - \bar{E}_1^T(t)) = \bar{E}_1^D(0).$$

Therefore we identify

$$\mathbb{E}(Y_i^D(1)|i \in C_{01}^D) - \sum_{t \neq 1} \alpha_t^T \mathbb{E}(Y_i^T(t)|i \in C_{t1}^T) = \frac{(\bar{E}_1^D(1) - \bar{E}_0^D(1)) - (\bar{E}_0^D(0) - \bar{E}_1^D(0))}{P^D(1|1) - P^D(1|0)},$$

which is the standard Wald estimator.

□

Proof of Corollary 5. It is obvious by direct substitution into Equation (3.4). □

Proof of Proposition 12. (i) It follows directly from the mapping of groups.

(ii) Part (i) identifies the weight $\alpha_0^T = (P^D(0|0) - P^D(0|1))/(P^D(1|1) - P^D(1|0))$, which we denote α_0^D in the Proposition. The other terms obtain by simple factorization, with $1 - \alpha_D^0 = \Pr(i \in C_{t1}^T | t > 1)$.

□

Proof of Proposition 13. Recall that Table 3 shows which groups take $D_i = d$ when $Z_i = z$.

(i) We have $P^D(0|z) = P^T(0|z)$ for $z = 0, 1$. Given Proposition 7(i), this gives us $\Pr(C_{010}^T) + \Pr(C_{012}^T) = P^D(0|0) - P^D(0|1)$ and $\Pr(C_{002}^T) + \Pr(C_{012}^T) = P^D(0|0) - P^D(0|2)$, which map into

$$\Pr(C_{010}^D) + \Pr(C_{011}^D) = P^D(0|0) - P^D(0|1)$$
$$\Pr(C_{001}^D) + \Pr(C_{011}^D) = P^D(0|0) - P^D(0|2);$$

and the last equation in Proposition 7(i) maps into

$$\Pr(C_{001}^D) + \Pr(C_{010}^D) + \Pr(C_{011}^D) + \Pr(A_0^D) = P^D(0|0).$$

Finally, $\Pr(A_1^D) = P^D(1|0)$ from the table. Defining $p = \Pr(C_{011}^D)$ gives the equations in the proposition, along with the constraints on p. Note also that $\Pr(C_{0*0}^D) = \Pr(A_0^D) + \Pr(C_{010}^D) = P^D(0|2)$.

(ii) First note that $\bar{E}_0^D(1) = \mathbb{E}(Y_i \mathbb{1}(i \in A_1^D)) = \mathbb{E}(Y_i^D(1)|i \in A_1^D)\Pr(A_1^D)$. The other equations can be read from the table:

$$\begin{aligned}
\bar{E}_1^D(0) &= \mathbb{E}(Y^D(0)\mathbb{1}(C_{00*}^D)) \\
\bar{E}_2^D(0) &= \mathbb{E}(Y^D(0)\mathbb{1}(C_{0*0}^D)) \\
\bar{E}_0^D(0) &= \mathbb{E}(Y^D(0)\mathbb{1}(C_{0**}^D)) \\
\bar{E}_1^D(1) &= \mathbb{E}(Y^D(1)\mathbb{1}(C_{*1*}^D)) \\
\bar{E}_2^D(1) &= \mathbb{E}(Y^D(1)\mathbb{1}(C_{**1}^D)).
\end{aligned}$$

Part (i) showed that we point-identify $\Pr(A_1^D)$, $\Pr(C_{01*}^D)$, $\Pr(C_{0*1}^D)$, and $\Pr(C_{00*}^D)$. This allows us to rewrite the last three lines as

$$\begin{aligned}
\bar{E}_0^D(0) &= P^D(0|1)\,\mathbb{E}(Y^D(0)|C_{00*}^D) + (P^D(0|0) - P^D(0|1))\,\mathbb{E}(Y^D(0)|C_{01*}^D) \\
&= P^D(0|2)\,\mathbb{E}(Y^D(0)|C_{0*0}^D) + (P^D(0|0) - P^D(0|2))\,\mathbb{E}(Y^D(0)|C_{0*1}^D) \\
\bar{E}_1^D(1) &= (P^D(0|0) - P^D(0|1))\,\mathbb{E}(Y^D(1)|C_{01*}^D) + P^D(1|0)\,\mathbb{E}(Y^D(1)|A_1^D) \\
\bar{E}_2^D(1) &= (P^D(0|0) - P^D(0|2))\,\mathbb{E}(Y^D(1)|C_{0*1}^D) + P^D(1|0)\,\mathbb{E}(Y^D(1)|A_1^D),
\end{aligned}$$

where we used the fact that $C_{1*1}^D = C_{11*}^D = A_1^D$.

Simple calculations give

$$\begin{aligned}
\mathbb{E}(Y^D(0)|C_{00*}^D) &= \frac{\bar{E}_1^D(0)}{\Pr(C_{00*}^D)} = \frac{\bar{E}_1^D(0)}{P^D(0|1)} \\
\mathbb{E}(Y^D(0)|C_{0*0}^D) &= \frac{\bar{E}_2^D(0)}{\Pr(C_{0*0}^D)} = \frac{\bar{E}_2^D(0)}{P^D(0|2)} \\
\mathbb{E}(Y^D(1)|C_{01*}^D) &= \frac{\bar{E}_1^D(1) - \bar{E}_0^D(1)}{P^D(0|0) - P^D(0|1)} \\
\mathbb{E}(Y^D(1)|C_{0*1}^D) &= \frac{\bar{E}_2^D(1) - \bar{E}_0^D(1)}{P^D(0|0) - P^D(0|2)}.
\end{aligned}$$

Finally,

$$\mathbb{E}(Y^D(0)|C^D_{01*}) = \frac{\bar{E}^D_1(1) - \mathbb{E}(Y^D(0)\mathbb{1}(C^D_{00*}))}{P^D(0|0) - P^D(0|1)} = \frac{\bar{E}^D_1(1) - \bar{E}^D_1(0)}{P^D(0|0) - P^D(0|1)}$$
$$\mathbb{E}(Y^D(0)|C^D_{0*1}) = \frac{\bar{E}^D_2(1) - \mathbb{E}(Y^D(0)\mathbb{1}(C^D_{00*}))}{P^D(0|0) - P^D(0|1)} = \frac{\bar{E}^D_0(0) - \bar{E}^D_1(0)}{P^D(0|0) - P^D(0|1)}$$

(iii) From (ii) we obtain directly, using Lemma 2,

$$\mathbb{E}(Y^D(1) - Y^D(0)|C^D_{01*}) = \frac{\bar{E}^D_1(1) + \bar{E}^D_1(0) - \bar{E}^D_0(1) - \bar{E}^D_0(0)}{P^D(0|0) - P^D(0|1)} = \frac{\mathbb{E}(Y|Z=1) - \mathbb{E}(Y|Z=0)}{P^D(0|0) - P^D(0|1)}$$
$$\mathbb{E}(Y^D(1) - Y^D(0)|C^D_{0*1}) = \frac{\bar{E}^D_2(1) + \bar{E}^D_2(0) - \bar{E}^D_0(1) - \bar{E}^D_0(0)}{P^D(0|0) - P^D(0|2)} = \frac{\mathbb{E}(Y|Z=2) - \mathbb{E}(Y|Z=0)}{P^D(0|0) - P^D(0|2)}.$$

□

C Strict Targeting in the 3×3 Model

Just like ours, Kirkeboen, Leuven, and Mogstad (2016)'s approach to identification relies on a monotonicity assumption and a restriction on the mapping from instruments to treatments. We translate them here in our notation to show that in this model, their assumptions are equivalent to ours.

Kirkeboen, Leuven, and Mogstad (2016) 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 6.

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 (2019): "Multivalued Treatments and Decomposition Analysis: An Application to the WIA Program," *Journal of Business & Economic Statistics*, in press, https://doi.org/10.1080/07350015.2019.1660664.

Bloom, H. S. (1984): "Accounting for no-shows in experimental evaluation designs," *Evaluation Review*, 8(2), 225–246.

Caetano, C., and J. C. Escanciano (2020): "Identifying Multiple Marginal Effects with a Single Instrument," *Econometric Theory*, forthcoming, https://doi.org/10.1017/S0266466620000213.

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.

Feng, J. (2020): "Matching Points: Supplementing Instruments with Covariates in Triangular Models," arXiv:1904.01159, https://arxiv.org/abs/1904.01159.

Goff, L. (2020): "A Vector Monotonicity Assumption for Multiple Instruments," available at http://www.columbia.edu/~ltg2111/.

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.

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. (2020): "Identification of Program Access Effects with an Application to Head Start," arXiv:1711.02048, https://arxiv.org/abs/1711.02048.

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

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. Torgovitsky, and C. R. Walters (2020a): "Identification of Causal Effects with Multiple Instruments: Problems and Some Solutions," Working Paper 25691, National Bureau of Economic Research.

——— (2020b): "Policy Evaluation With Multiple Instrumental Variables," Working Paper 27546, National Bureau of Economic Research.

Mountjoy, J. (2019): "Community Colleges and Upward Mobility," Chicago Booth mimeo, Available at SSRN: `https://ssrn.com/abstract=3373801`.

Muralidharan, K., M. Romero, and K. Wüthrich (2019): "Factorial Designs, Model Selection, and (Incorrect) Inference in Randomized Experiments," Working Paper 26562, National Bureau of Economic Research.

Pinto, R. (2015): "Selection bias in a controlled experiment: the case of Moving to Opportunity," University of Chicago, mimeo.

——— (2019): "Noncompliance as a Rational Choice: A Framework that Exploits Compromises in Social Experiments to Identify Causal Effects," UCLA, mimeo, `https://www.rodrigopinto.net/`.

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.