When does IV identification not restrict outcomes?

Leonard Goff*

This version: January 6, 2026

Abstract

Many identification results in instrumental variables (IV) models hold without restrictions on the distribution of potential outcomes, or how those outcomes are related to selection behavior. This enables IV models to allow for arbitrary heterogeneity in treatment effects and the possibility of selection on gains in the outcome. I show that when the available instruments take a finite number of values, a condition that is easily seen to be sufficient for identification without restricting outcomes is also necessary. The condition generalizes the LATE monotonicity assumption, and I provide a new characterization of it that reveals a common structure behind a wide variety of known IV identification results for binary and discrete-valued treatments. The characterization yields an approach to enumerate all models of selection that allow for the identification of local average treatment effect type parameters without restricting outcomes. This search uncovers new selection models that yield identification, and provides impossibility results for others. An application considers the identification of complementarities between two cross-randomized binary treatments, obtaining a necessary and sufficient condition on selection for local average interaction effects to be identified without imposing restrictions on outcomes.

*Department of Economics, University of Calgary. I thank Simon Lee as well as Pat Kline, Eric Mbakop and Adam Rosen for helpful conversations about these ideas. Email: leonard.goff@ucalgary.ca.

1 Introduction

To leverage instrumental variables with heterogeneous treatment effects, researchers often make assumptions about selection into treatment, such as the "monotonicity" assumption of the local average treatment effects (LATE) model (Imbens and Angrist, 1994). Given this monotonicity assumption, the average treatment effect among compliers is point identified, with no restrictions imposed on the distribution of potential outcomes beyond them being independent of the instrument.

In this paper I analyze how a similar result holds broadly across instrumental variables (IV) models, when the treatment is not necessarily binary. I first show that if restrictions on selection behavior are sufficient to establish a particular generalization of the monotonicity assumption, then corresponding local average treatment effect parameters are identified. When the instruments have finite support, this sufficient condition for IV identification can be seen as a consequence of existing results by Heckman and Pinto (2018) and Navjeevan et al. (2023).

The main theoretical contribution of this paper is to show that the identification result, strikingly, has a converse. Consider any parameter that takes the form of an average treatment effect among a subgroup of the population defined by their selection behavior. For this treatment effect parameter to be point identified without the researcher imposing further restrictions on the joint distribution of potential outcomes, the selection model *must* permit the above generalization of monotonicity to hold. I es-

tablish this result using a novel geometric representation of monotonicity-type assumptions, which relates the parameter of interest to a vector space that depends on the model of selection maintained by the researcher.

Together, the above results yield a necessary and sufficient condition for identification of a given treatment effect to avoid reliance on ad-hoc assumptions about the outcome such as treatment effect homogeneity: what I call *outcome-nonrestrictive* identification. Outcome-nonrestrictive identification allows for what Heckman et al. (2006) call essential heterogeneity: not only can gains from treatment be heterogeneous, but selection behavior may in part depend on an individual's idiosyncratic gains.

Using this condition I show that it is straightforward to check whether there exist any local average treatment effect parameters that are identified without restricting outcomes, given a model of selection. By then enumerating over alternative selection models, the researcher can build a transparent map between assumptions about selection and identified parameters. I propose algorithms to implement this insight and build an exhaustive catalog of outcome-nonrestrictive identification results for treatment effects with binary or ternary instruments and treatments. In an application to settings with two-cross randomized binary instruments, I show that the necessary and sufficient condition for the interaction between two treatments to be identified without outcome restrictions is substantive but economically meaningful. Using this I revisit two empirical studies.

My results show quite generally that the price of outcome-nonrestrictive identification is the need to make assumptions about selection. This trade-

off appears to be quite appealing to researchers in practice. In "design-based" studies the researcher often has contextual knowledge about factors that affect treatment uptake in a given setting (Card, 2022), but may be reluctant to make assumptions about the very causal effect being studied (e.g. that it is homogeneous across individuals). Outcome-nonrestrictive identification results also have the practical benefit of paving the way for analysis to be repeated across multiple outcome variables, with the same assumptions about selection in a given (natural) experiment aiding identification for each outcome.[1] This is particularly useful in experimental work, when many outcome variables can be collected at minimal additional cost.

The perspective of outcome-nonrestrictive identification turns out to unify a wide variety of existing IV identification results in the literature. My positive identification result provides a simple and unified proof of point identification for settings including: i) the original LATE model (Imbens and Angrist, 1994); ii) the marginal treatment effect (Heckman and Vytlacil 2001; Heckman and Vytlacil 2005) and its generalization to multivalued treatments (Lee and Salanié, 2018); iii) unordered monotonicity (Heckman and Pinto, 2018); iv) vector monotonicity with multiple instruments (Goff, 2024); v) restrictions on choice and/or knowledge of second-best options (Kirkeboen et al., 2016); vi) interaction effects between two treatments (Blackwell, 2017); and vii) notions of monotonicity that are

[1] An outcome-nonrestrictive identification result is not fully indifferent to *which* variable one uses as an outcome: one must make the standard independence and exclusion restrictions for each. But in settings where treatment is as-good-as-randomly assigned and there are limited opportunities for the assignment to affect anything except via treatment (e.g. many experimental settings), exclusion may be quite natural without much further justification specific to each outcome.

only required to hold between particular *pairs* of instrument values (Sun and Wüthrich, 2024; van't Hoff et al., 2023; Sigstad, 2024). I contrast the above results with other identification results from the IV literature that weaken assumptions about selection while leveraging additional assumptions about outcomes and are therefore not outcome-nonrestrictive (Kolesár, 2013; de Chaisemartin, 2017; Comey et al., 2023).

In the other direction, this paper is to my knowledge the first to demonstrate a *necessary* condition for local average treatment effect parameters to be identified at a high level of generality. Without also establishing necessity of the condition, the interplay between assumptions about selection and identification cannot be fully discerned. In recent work, Navjeevan, Pinto and Santos (2023) give a necessary and sufficient condition for the identification of unconditional moments in IV models, but do not demonstrate that this condition is necessary to identify moments of potential outcomes that condition on selection. I show that their condition is equivalent to mine, and that their condition is thus indeed also necessary to identify local average treatment effect parameters.

These results are relevant both theoretically and for empirical practice. For researchers designing experiments or seeking a quasi-experimental research design, the map between selection restrictions and identified parameters outlines what instrument variation would yield identification that is robust to the failure of often poorly-motivated assumptions like treatment effect homogeneity. Given the available instrument variation and a selection model, the results yield a simple way to check whether any treatment

effects are identified, and if so for which subgroups of the population. On a theoretical level, this paper reveals the structure of the trade-off between assumptions about selection and assumptions about outcomes. When researchers prefer the former, my results show the upside to being flexible about one's parameter of interest (e.g. settling for a LATE). If instead one is interested only in e.g. an overall average treatment effect, I show that point identification is typically not possible without restricting outcomes.

2 Setup and a general identification result

2.1 Notation and basic IV assumptions

Let treatment t take values in a finite set $\mathcal{T}$. Denote potential outcomes as $Y_i(t, z)$ and potential treatments as $T_i(z)$, where Z_i are instruments with support $\mathcal{Z}$. I'll refer to Z_i as "the instruments", since in general it can be a vector of instrumental variables. The index i corresponds to observational units, i.e. "individuals".

Let $D_i^{[t]}(z) = \mathbb{1}(T_i(z) = t)$ be an indicator that i takes treatment t when the instruments are equal to z. I impose throughout the exclusion restriction that Z_i only affects Y_i through $T_i = T_i(Z_i)$, so that $Y_i(t, z) = Y_i(t)$ for all i and $z \in \mathcal{T}$. The observed outcome Y_i is then:

$$Y_i = Y_i(T_i(Z_i)) = \sum_{t \in \mathcal{T}} D_i^{[t]}(Z_i) \cdot Y_i(t) \tag{1}$$

Let $G_i : \mathcal{Z} \to \mathcal{T}$ the function that yields individual i's counterfactual treatment value for each possible instrument value z. Following Heckman

and Pinto (2018), call G_i individual i's "response type". Let us denote the full set of conceivable functions from $\mathcal{Z}$ to $\mathcal{T}$ as $\mathcal{T}^{\mathcal{Z}}$. I refer to any subset $\mathcal{G} \subseteq \mathcal{T}^{\mathcal{Z}}$ as a *selection model* or *choice model*, where $\mathcal{G}$ denotes a set of response types that are admissible according to the model (i.e. $supp\{G_i\} \subseteq \mathcal{G}$). When $\mathcal{G}$ is a strict subset of $\mathcal{T}^{\mathcal{Z}}$, it reflects a substantive restriction on the response types that are allowed.

I will also assume throughout that the instruments are exogenous in the sense that

$$Z_i \perp\!\!\!\perp (\tilde{Y}_i, G_i) \tag{2}$$

where $\tilde{Y} = \{Y_i(t)\}_{t \in \mathcal{T}}$ is a vector of potential outcomes across all treatments t. Eq. (2) says that potential outcomes and response types are jointly independent of the instruments. In applications, researchers often defend a *conditional* version of (2), i.e. $\{Z_i \perp\!\!\!\perp (\tilde{Y}_i, G_i)\}|X_i$, where X_i are observed covariates unaffected by treatment. Since my focus in this paper is on identification and not estimation of treatment effects, I suppress throughout conditioning on any such covariates for ease of exposition, and consider them in the empirical application and in Appendix J.2.

2.2 Outcome-nonrestrictive IV identification

I say that a causal parameter is *outcome-nonrestrictive* identified if identification of that parameter holds without restrictions on the distribution of $(\tilde{Y}_i, G_i)$ apart from independence from the instruments (2). That identification not restrict the distribution of $\tilde{Y}_i$ is often implicit in work that

leverages for identification assumptions about selection, e.g. Imbens and Angrist (1994) and Heckman and Pinto (2018).

Specifically, we say that a parameter θ is outcome-nonrestrictive identified if the identified set for θ is a singleton, for all distributions of $(\tilde{Y}, G, Z)$ that are compatible with the model M. Here the model M imposes that Eq. (2) and $supp\{G_i\} \subseteq \mathcal{G}$ hold, but places no restrictions on the distribution of $(\tilde{Y}, G)$.[2] For a parameter to be outcome-nonrestrictive identified, it must be point-identified for all distributions of observables that are consistent with the model, and not just particular distributions that happen to yield a singleton for the identified set. This is in line with typical definitions of a parameter being identified (Matzkin, 2007; Lewbel, 2019).

Defining the notion of outcome-nonrestrictive identification formally involves accounting for possible restrictions on observables implied by a given selection model $\mathcal{G}$. This requires some notation that will not be necessary for the main exposition of this paper, so I refer the reader to Appendices A.2 and A.3 for the formal definition. The definition of outcome-nonrestrictive identification does not restrict the support of Y_i in any way (e.g. that it be discrete or bounded), and the results of this paper do not require such a restriction.

2.3 Parameters of interest

This paper focuses on parameters θ that take the form of a conditional counterfactual mean $\mu_c^t := \mathbb{E}[Y_i(t)|c(G_i) = 1]$, a conditional treatment

[2]The definition of outcome-nonrestrictive identification in A.3 considers only distributions such that $\mathbb{E}[Y_i(t)|G_i = g]$ exists and is finite for each t, g and such that the parameter of interest is well-defined.

effect $\Delta_c^{t,t'} := \mathbb{E}[Y_i(t') - Y_i(t)|c(G_i) = 1] = \mu_c^{t'} - \mu_c^t$, or a probability $P(c(G_i) = 1)$, defined given a function $c : \mathcal{G} \to \{0, 1\}$. The function c represents inclusion in a collection of response types. For example, in the LATE model, the LATE is a conditional treatment effect $\Delta_c^{t,t'}$ with $t' = 1, t = 0$ and $c(g) = \mathbb{1}(g = \text{complier})$.

Note that identification of μ_c^t and $\mu_c^{t'}$ immediately implies identification of $\Delta_c^{t,t'}$, since $\Delta_c^{t,t'} = \mu_c^{t'} - \mu_c^t$. With outcome-nonrestrictive identification, this implication goes the other way as well: outcome nonrestrictive identification of $\Delta_c^{t,t'}$ requires outcome-nonrestrictive identification of each of the terms μ_c^t and $\mu_c^{t'}$. The intuition is that absent assumptions about the joint distribution of potential outcomes, data from individuals with $T_i = t$ provide no information about $Y_i(t')$ for a different treatment $t' \neq t$, and vice-versa. This idea is formalized in the following result:

Proposition 1. *for $t' \neq t$ is outcome-nonrestrictive identified if and only if μ_c^t and $\mu_c^{t'}$ are.*

Proofs of this and subsequent results are in Appendix B.

Although researchers are typically more interested in treatment effect parameters like $\Delta_c^{t,t'}$ than they are in counterfactual means, we can, given Proposition 1, begin our analysis of outcome-nonrestrictive identification with the simpler counterfactual mean parameters μ_c^t.[3]

[3]The proof extends the conclusion of Proposition 1 to treatment effect parameters that involve any number of separate treatment states. This is useful for the study of interaction effects in Section 5.

2.4 Identifying counterfactual means

I begin with a simple argument establishing outcome-nonrestrictive identification of $\mu_c^t = \mathbb{E}[Y_i(t)|c(G_i) = 1]$ under a condition on c (given t). As shown in Appendix D, this result can be seen as a corollary to Theorem T-2 of Heckman and Pinto (2018), and to Corrollary 4.4 of Navjeevan et al. (2023). This section constitutes a minor generalization of these results in which the instruments themselves need not be discrete. This additional generality allows us to view identification results for the marginal treatment effect (Heckman and Vytlacil, 2001; Lee and Salanié, 2018) as limiting cases. This connection is developed explicitly in Appendix G.

Consider any finite collection of distinct instrument values $z_k \in \mathcal{Z}$ for $k = 1 \dots K$ and corresponding coefficients α_k. Observe that the following quantity is then identified:

$$\begin{aligned}\sum_{k=1}^{K} \alpha_k \cdot \mathbb{E}\left[Y_i \cdot D_i^{[t]} | Z_i = z_k\right] &= \sum_{k=1}^{K} \alpha_k \cdot \mathbb{E}\left[Y_i(t) \cdot D_i^{[t]}(z_k) | Z_i = z_k\right] \\ &= \mathbb{E}\left[Y_i(t) \cdot \left(\sum_{k=1}^{K} \alpha_k \cdot D_i^{[t]}(z_k)\right)\right]\end{aligned}$$

where $D_i^{[t]} = D_i^{[t]}(Z_i) = \mathbb{1}(T_i = t)$ and the second equality follows from independence (2).

Suppose that the z_k and α_k could be chosen in such a way to guarantee that the linear combination $\sum_k \alpha_k \cdot D_i^{[t]}(z_k)$ (in parentheses above) is equal

to 0 or 1 for any given i. The above then simplifies to

$$\sum_{k=1}^{K} \alpha_k \cdot \mathbb{E}\left[Y_i \cdot D_i^{[t]} | Z_i = z_k\right] \\ = P\left(\sum_k \alpha_k \cdot D_i^{[t]}(z_k) = 1\right) \cdot \mathbb{E}\left[Y_i(t) \middle| \sum_k \alpha_k \cdot D_i^{[t]}(z_k) = 1\right]$$

Meanwhile, by similar steps

$$\sum_{k=1}^{K} \alpha_k \cdot \mathbb{E}\left[D_i^{[t]} | Z_i = z_k\right] = P\left(\sum_k \alpha_k \cdot D_i^{[t]}(z_k) = 1\right) \tag{3}$$

Therefore, provided that $P\left(\sum_k \alpha_k \cdot D_i^{[t]}(z_k) = 1\right) > 0$, we have:

$$\mathbb{E}\left[Y_i(t) \middle| \sum_k \alpha_k \cdot D_i^{[t]}(z_k) = 1\right] = \frac{\sum_{k=1}^{K} \alpha_k \cdot \mathbb{E}\left[Y_i \cdot D_i^{[t]} | Z_i = z_k\right]}{\sum_{k=1}^{K} \alpha_k \cdot \mathbb{E}\left[D_i^{[t]} | Z_i = z_k\right]} \tag{4}$$

Eq. (4) represents a generalization of the "Wald ratio" form common among IV estimands, and turns out to nest a surprising variety of point identification results from the IV literature.

Example 1 (LATE model). *Consider the model of Imbens and Angrist (1994) with a binary instrument, where* $\mathcal{Z} = \mathcal{T} = \{0, 1\}$ *and we rule out "defiers", i.e. those who would have* $T_i(0) = 1$, $T_i(1) = 0$. *That is,* $\mathcal{G}$ *consists of never-takers, always-takers, and compliers. Then* $(D_i^{[1]}(1) - D_i^{[1]}(0)) \in \{0, 1\}$ *for all* i *(*1 *for compliers and* 0 *for the other groups), corresponding to* $K = 2$ *with* $(z_1, \alpha_1) = (1, 1)$ *and* $(z_2, \alpha_2) = (0, -1)$. *Thus, by* (4), $\mathbb{E}[Y_i(1) | i \text{ is complier}] = \frac{\mathbb{E}[Y_i \cdot D_i^{[1]} | Z_i = 1] - \mathbb{E}[Y_i \cdot D_i^{[1]} | Z_i = 0]}{\mathbb{E}[D_i^{[1]} | Z_i = 1] - \mathbb{E}[D_i^{[1]} | Z_i = 0]}$.

Example 2 (a model with both compliers and defiers). *Continu-*

ing with the case of a binary treatment and binary instrument, consider a model that allows defiers but instead rules out always-takers or never-takers, i.e. imposes that all individuals' treatment is altered by the instrument. Then $D_i^{[1]}(1) \in \{0,1\}$ *for all* i *(*1 *for compliers and* 0 *for defiers), and thus* $\mathbb{E}[Y_i(1) - Y_i(0) | i \text{ is complier}] = \frac{\mathbb{E}[Y_i \cdot D_i | Z_i = 1]}{\mathbb{E}[D_i | Z_i = 1]}$ *by Eq.* (4).

Appendix G discusses several further examples from the literature, including cases where the treatment is not binary.

Note that we can think of the coefficients α_k in Eq. (4) as a function α on $\mathcal{Z}$, where the set of points z where $\alpha(z)$ differs from zero is finite. If $\mathcal{Z}$ is itself finite, then the coefficients α_k can in turn be represented simply as a vector in $\mathbb{R}^{|\mathcal{Z}|}$, a fact that I make use of in Section 3. Given a pair (t, α) the value of $\sum_k \alpha_k \cdot D_i^{[t]}(z_k)$ depends only on the response type G_i of individual i. We can therefore write the event that $\sum_k \alpha_k \cdot D_i^{[t]}(z_k) = 1$ as $c(G_i) = 1$, where $c : \mathcal{G} \to \{0,1\}$ is a function that depends on the treatment value t and coefficient function α. The the parameter on the LHS of Eq. (4) is therefore of the form μ_c^t introduced in the last section.

We can now summarize the result of Equation (4) as follows:

Theorem 1. *Given independence Eq. (2) and a pair* (t, α) *with the property that* $P\left(\sum_k \alpha_k \cdot D_i^{[t]}(z_k) \in \{0,1\}\}\right) = 1$*, the quantity* $P(c(G_i) = 1)$ *is identified by the LHS of* (3)*. If additionally* $P(c(G_i) = 1) > 0$*, the conditional counterfactual mean* μ_c^t *is identified by the RHS of (4). Since the only restrictions used in deriving* (4) *are that* (2) *holds and that* $P(c(G_i) = 1) > 0$*, identification of* μ_c^t *is outcome nonrestrictive.*

The key to observing that the above identification of $\mu_c^t := \mathbb{E}[Y_i(t)|c(G_i) = 1]$ is outcome-nonrestrictive is that whether the requirement $P(c(G_i) = 1) > 0$ holds depends only on the marginal distribution of G_i, and whether each $z_k \in \mathcal{Z}$ depends only on the distribution of Z_i. This implies nothing about the distribution of potential outcomes or their relation to G_i.

2.5 From counterfactual means to treatment effects

While Theorem 1 yields identification of conditional counterfactual means μ_c^t, we can furthermore identify *treatment effects* when two treatment values t and t' admit of identification of counterfactual means that condition on the same group of response types, i.e. $c^{[t,\alpha]}(\cdot) = c^{[t',\alpha']}(\cdot)$ for some α, α', where $c^{[t,\alpha]}$ denotes the function corresponding to (t, α). Let $c(\cdot)$ denote this common function. Then Theorem 1 implies that $\Delta_c^{t,t'} = \mu_c^{t'} - \mu_c^t$ is outcome-nonrestrictive identified as well. Appendix A.4 provides some general results regarding when this can be done, and Appendix A.5 discusses how this typically leads to testable implications of the model. The next section considers the identification of treatment effects in detail, when the instruments have finite support.

In both Examples 1 and 2 above, treatment effects of the form $\Delta_c^{t,t'}$ (and not just counterfactual means μ_c^t) are outcome-nonrestrictive identified. I return to these examples in the next section.

Note: By replacing $Y_i(t)$ by $\mathbb{1}(Y_i(t) \leq y)$ we can also identify the conditional distributions $F_{Y(t)|c(G)=1}$ and compute e.g. quantile treatment ef-

fects or establish bounds on the distribution of treatment effects among the $c(G_i) = 1$ group (Fan and Park, 2010), whenever a $\Delta_c^{t,t'}$ is outcome-nonrestrictive identified.

3 Necessity with discrete instruments

The remainder of the paper now specializes to settings in which the instruments Z_i are discrete and take only a finite number of values $\mathcal{Z}$. This allows us to establish that the condition from Theorem 1 is also *necessary* for outcome-nonrestrictive identification to occur. Combining with Theorem 1 and Proposition 1, this provides a full characterization of outcome-nonrestrictive identification, with a simple geometric interpretation.

When $\mathcal{Z}$ is discrete and finite, a function α can be associated with a vector in $\mathbb{R}^{|\mathcal{Z}|}$. Across a finite set of treatments, note that $\mathcal{G}$ can then also only take finitely many values, i.e. $|\mathcal{G}| \leq |\mathcal{T}|^{|\mathcal{Z}|}$. A function $c : \mathcal{G} \to \{0, 1\}$ defining a causal parameter like $\mu_c^t = \mathbb{E}[Y_i(t)|c(G_i) = 1]$ can now be associated with a $|\mathcal{G}|$-component vector c with components $c_g = c(g)$ for each $g \in \mathcal{G}$.

In this setting, we can also express the content of the selection model $\mathcal{G}$ through a $|\mathcal{Z}| \times |\mathcal{G}|$ matrix A, where component A_{zg} gives the common treatment $T_i(z)$ that all units i with $G_i = g$ take, when the instruments are equal to z. The restrictions imposed by selection model $\mathcal{G}$ correspond to deleting columns from the $|\mathcal{Z}| \times |\mathcal{T}|^{|\mathcal{Z}|}$ matrix that would include all $|\mathcal{T}|^{|\mathcal{Z}|}$ imaginable response types given $\mathcal{T}$ and $\mathcal{Z}$.

Define for any treatment t a matrix $A^{[t]}$ having binary entries $[A^{[t]}]_{zg} = \mathbb{1}(A_{zg} = t)$, which records the common value of $D_i^{[t]}(z)$ for any individual i having $G_i = g$. The matrix $A^{[t]}$ simply tells us whether units take treatment t versus any other treatment. A matrix analogous to $A^{[t]}$ is used heavily in Heckman and Pinto (2018).

Example 1 (continued from Section 2.4). *In this example:*

$$A^{[1]} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix} \qquad \textit{and} \qquad A^{[0]} = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}$$

where the first row of each matrix represents $z = 0$ and the second $z = 1$, while the columns correspond to never-takers, always-takers, and compliers, respectively.

Example 2 (continued from Section 2.4). *In this example:*

$$A^{[1]} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \qquad \textit{and} \qquad A^{[0]} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

where the first row of each matrix represents $z = 0$ and the second $z = 1$, while the columns correspond to compliers and defiers, respectively.

In both of the examples above, the entries of $A^{[0]}$ are simply one minus the corresponding entry of $A^{[1]}$. This is a general property of the matrices $A^{[t]}$ with a binary treatment, but does not extend to non-binary treatments. In general, we instead have $\sum_{t \in \mathcal{T}} [A^{[t]}]_{zg} = 1$ for each $z \in \mathcal{Z}$ and $t \in \mathcal{T}$.

3.1 The Theorem 1 condition is necessary

Consider a parameter of the form μ_c^t. Recall from above the representation of $c(\cdot)$ as a binary-valued vector $c \in \mathbb{R}^{|\mathcal{G}|}$ with $c_g \in \{0,1\}$ for each $g \in \mathcal{G}$. For any matrix B let $rowspace(B)$ or $rs(B)$ denote its rowspace, and B' its transpose. We have the following converse of Theorem 1:

Theorem 2. *Suppose that $\mathcal{Z}$ and $\mathcal{T}$ are finite. Then if $\mu_c^t := E[Y_i(t)|c(G_i) = 1]$ is outcome-nonrestrictive identified, then $c \in rowspace(A^{[t]})$.*

The rowspace of $A^{[t]}$ is the set of vectors $c \in \mathbb{R}^{|\mathcal{G}|}$ such that $c' = \alpha' A^{[t]}$ for some $\alpha \in \mathbb{R}^{|\mathcal{Z}|}$. Note that in a setting with finite $|\mathcal{Z}|$ (and hence a finite space of response types as well), the condition $P\left(\sum_k \alpha_k \cdot D_i^{[t]}(z_k) \in \{0,1\}\}\right) = 1$ from Theorem 1 occurs precisely when $\alpha' A^{[t]}$ is a binary-valued vector in $\mathbb{R}^{|\mathcal{G}|}$, i.e. when $c^{[\alpha,t]} \in rowspace(A^{[t]})$.

Theorem 2 thus establishes that if the instruments are finite, then Theorem 1 covers *all* instances in which a counterfactual mean that conditions on response types can be identified in an outcome-nonrestrictive way.

Combining Theorems 1 and 2, we have in the case of finite discrete instruments that given a selection model $\mathcal{G}$, a conditional counterfactual mean of the form $\mathbb{E}[Y_i(t)|c(G_i) = 1]$ is outcome-nonrestrictive identified *if and only if* the vector representation $c \in \{0,1\}^{|\mathcal{G}|}$ of $c(\cdot)$ lies in the rowspace of $A^{[t]}$. Using Proposition 1, it then follows that a treatment effect parameter $\mathbb{E}[Y_i(t') - Y_i(t)|c(G_i) = 1]$ is outcome-nonrestrictive identified *if and only if* c lies in the rowspaces of both $A^{[t]}$ *and* $A^{[t']}$, i.e. $c \in (rs(A^{[t']}) \cap rs(A^{[t]}))$. Note that as an intersection of two vector spaces, $(rs(A^{[t']}) \cap$

$rs(A^{[t]})$) is also a vector space in $\mathbb{R}^{|\mathcal{G}|}$.

Example 1 (continued). *The rowspaces of $A^{[1]}$ and $A^{[0]}$ can be found by row-reducing each matrix, revealing that $rs(A^{[1]})$ is the plane in $\mathbb{R}^3$ spanned by the compliers and always-takers, while $rs(A^{[0]})$ is the plane spanned by the compliers and never-takers.*

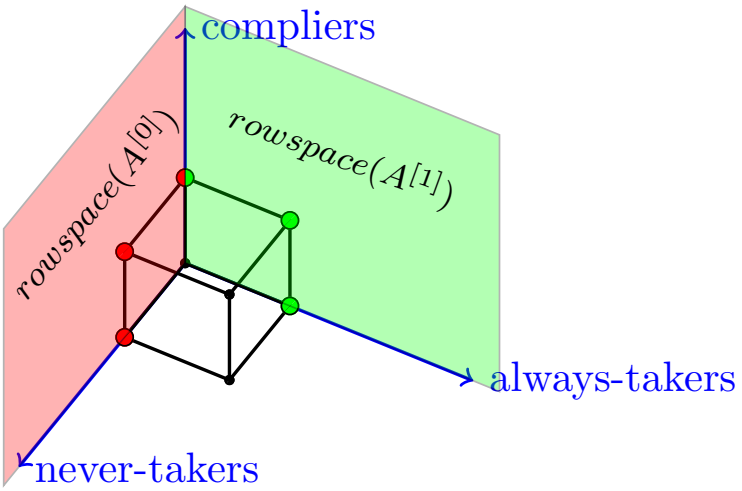

Figure 1: Identification in the LATE model with a binary instrument. The vector $c = (0, 0, 1)'$ belongs to both $rs(A^{[1]})$ and $rs(A^{[0]})$ and hence the LATE parameter $\mathbb{E}[Y_i(1) - Y_i(0)|i \text{ is a complier}]$ is identified, and this identification is outcome-nonrestrictive. $(0, 0, 1)'$ is the only vertex of the unit cube that belongs to $rs(A^{[1]}) \cap rs(A^{[0]})$, and is thus the only response type for which average treatment effects can be identified without restricting outcomes.

By Theorem 1, we can thus identify the mean of $Y_i(1)$ among always-takers or among compliers (or among both), and we can identify the mean of $Y_i(0)$ among never-takers or among compliers (or both). As depicted in Figure 1, these correspond to the non-zero vertices of the unit cube in $\mathbb{R}^3$ that take a value of zero in the never-takers "direction", or a value of zero in the always-takers "direction", respectively.

Note that $(0, 0, 1)'$ the unique non-zero vertex of the unit cube in $\mathbb{R}^3$ that belongs to both $rs(A^{[1]})$ and to $rs(A^{[0]})$. The local average treatment effect

*Δ_c is outcome-nonrestrictive identified for the compliers $c = (0, 0, 1)'$, because this c belongs to both $rs(A^{[1]})$ and $rs(A^{[0]})$. Theorem 2 demonstrates that the LATE among compliers is the **only** treatment effect parameter Δ_c that is outcome-nonrestrictive identified in the LATE model.*

Example 2 (continued). *The rowspaces of $A^{[1]}$ and $A^{[0]}$ are the same and both span $\mathbb{R}^2$. Thus we identify the LATE for compliers as well as the LATE for defiers. For example, using $\alpha = (0, 1)$ for $t = 1$ and $\alpha = (1, 0)$ for $t = 0$, we have that:*

$$\mathbb{E}[Y_i(1) - Y_i(0)|i \text{ is complier}] = \frac{\mathbb{E}[Y_i \cdot D_i^{[1]}|Z_i = 1]}{\mathbb{E}[D_i^{[1]}|Z_i = 1]} - \frac{\mathbb{E}[Y_i \cdot D_i^{[0]}|Z_i = 0]}{\mathbb{E}[D_i^{[0]}|Z_i = 0]}$$

An analagous construction identifies the average treatment effect among defiers: $\mathbb{E}[Y_i(1)-Y_i(0)|i \text{ is defier}] = \frac{\mathbb{E}[Y_i \cdot D_i^{[1]}|Z_i=0]}{\mathbb{E}[D_i^{[1]}|Z_i=0]} - \frac{\mathbb{E}[Y_i \cdot D_i^{[0]}|Z_i=1]}{\mathbb{E}[D_i^{[0]}|Z_i=1]}$. See Figure 4 in Appendix F.0.1 for a visualization similar to Figure 1.

Appendix D discusses how Theorems 1 and 2 relate to recent necessary and sufficient conditions for identification in IV models by Navjeevan et al. (2023) and by Heckman and Pinto (2018). While Theorem 1 can be seen a corollary to results of both Navjeevan et al. (2023) and Heckman and Pinto (2018), Theorem 2 is novel. Further, neither work obtains the condition $c \in (rs(A^{[t']}) \cap rs(A^{[t]}))$ for the identification of treatment effect parameters. This observation enables an exhaustive search for outcome-nonrestrictive identification results for treatment effects when $|\mathcal{Z}|$ is finite.

Remark: Theorems 1 and 2 both extend to the more general family of

target parameters that can be defined by functions c that depend on Z_i in addition to response types G_i. This is useful to nest parameters like the average treatment effect on the treated, or certain parameters that can arise in settings with multiple instruments. See Appendix H for details.

3.2 Discussion of Theorem 2

Although Theorem 2 synthesizes a wide variety of existing IV identification results (detailed in Appendix G), the requirement of *outcome-nonrestrictive* identification—rather than point identification in general—is important.

Point identification, given a distribution $\mathcal{P}_{obs}$ of the observables (Y, T, Z), requires that there not exist two distributions $\mathcal{P}$ and $\mathcal{P}'$ of the underlying model variables $(\tilde{Y}, G, Z)$ such that both $\mathcal{P}$ and $\mathcal{P}'$ recover $\mathcal{P}_{obs}$, but $\theta(\mathcal{P}) \neq \theta(\mathcal{P}')$. The proof of Theorem 2 shows that if $c \notin rs(A^{[t]})$, there always exist $\mathcal{P} \in M$ that enable us to construct such a $\mathcal{P}'$, where recall that M is the set of permissible DGPs according to the model. If one restricts the model space M by imposing further assumptions about the distribution of potential outcomes (or how they are related to response type G_i), the constructed distribution $\mathcal{P}'$ may violate those assumptions and thus not belong to M, enabling identification even if $c \notin rs(A^{[t]})$.

For example, it is known that $\mathbb{E}[Y_i(t') - Y_i(t)]$—the unconditional average treatment effect (ATE) between t and t'—is identified under an assumption of "no selection on gains" (NSOG). NSOG imposes that $Y_i(t) - Y_i(t')$ be mean independent of T_i and Z_i for all $t, t' \in \mathcal{T}$ (Kolesár, 2013;

Arora et al., 2021). Homogeneous treatment effects—that $Y_i(t) - Y_i(t')$ is the same for all i—is a special case of NSOG. In Appendix C, I show that unconditional counterfactual means $\mu^t := \mathbb{E}[Y_i(t)]$ are identified under NSOG, extending existing results for ATEs. Identification of μ^t under NSOG requires no assumptions on selection beyond an order condition that can be checked empirically.

Note that the parameter μ^t is a parameter of the form μ^t_c, where $c = (1, 1 \dots 1)'$ with a one for every response type in $supp\{G_i\}$. Absent a substantive selection model $\mathcal{G}$ that rules out some of the conceivable response types in $\mathcal{T}^{\mathcal{Z}}$, the vector $c = (1, 1 \dots 1)'$ corresponding to μ^t will never be in the row space of $A^{[t]}$. In particular, as long as there is a "never-takers" group $g_0(t)$ for treatment t such that $T_i(z) \neq t$ for all $z \in \mathcal{Z}$ when $G_i = g_0(t)$, there will exist a zero in the corresponding entry of any c in the rowspace of $A^{[t]}$. I show explicitly in Appendix C.2 that the construction $\mathcal{P}'$ in the proof of Theorem 2 can only satisfy NSOG if $\mathcal{G}$ is such that $(1, 1 \dots 1)' \in rs(A^{[t]})$ for all $t \in \mathcal{T}$. That is, if $(1, 1 \dots 1)' \notin rs(A^{[t]})$, then although the distribution $\mathcal{P}'$ can be chosen to satisfy all of the *other* assumptions of the IV model aside from NSOG and yield $\theta(\mathcal{P}) \neq \theta(\mathcal{P}')$, the distribution $\mathcal{P}'$ necessarily violates NSOG. Thus $\mu^t_c = \theta(\mathcal{P})$ can remain identified under NSOG, even if $c \notin rs(A^{[t]})$. This illustrates how Theorem 2 can coexist with identification results that operate by restricting treatment effect heterogeneity rather than response type heterogeneity.[4]

[4]Intuitively, NSOG is strong enough to let the researcher *impute* the value of $\mathbb{E}[Y_i(t)|G_i = g_0(t)]$ if such a never-taker group exists. This eliminates the dependence of any identifying estimand for μ^t on the distribution of Y_i among individuals such that $G_i = g_0(t)$ and $T_i = t$ (which is unobservable).

Further examples of IV identification results that are not covered by Theorem 2, because they restrict outcomes or consider different types of causal parameters, are discussed in Appendix C.3. These examples include results from de Chaisemartin (2017), Comey et al. (2023), and Kline and Walters (2016).

4 Making use of the equivalence result

This section shows how the necessary and sufficient condition $c \in rs(A^{[t]})$ can be useful in understanding existing identification results for treatment effects, generating new ones, and ruling out further opportunities for identification in a given selection model. We continue here our focus from Section 3 on settings in which the instruments have finite support.

4.1 A geometric characterization of identification

Let $\mathcal{C}(t)$ be the set of c in the rowspace of $A^{[t]}$ that have entries of only zero or one: i.e.[5]

$$\mathcal{C}(t) = rs(A^{[t]}) \cap \{0,1\}^{|\mathcal{G}|} \tag{5}$$

It is always the case that $C(t) \neq \emptyset$ provided that $P(T_i = t) > 0$.[6] An identified treatment effect in turn arises when $\mathcal{C}(t) \cap \mathcal{C}(t') \neq \emptyset$.[7] When treatment is binary, this observation can be used to establish that $\Delta_c^{0,1}$ is

[5] Melo and Winter (2019) study the cardinality of the intersection between the unit cube in $\mathbb{R}^n$ and any linear subspace of $\mathbb{R}^n$. Their result implies that $\mathcal{C}(t)$ has a cardinality of at most 2^k where $k = rank(A^{[t]})$.

[6] To see this, note that $\mathbb{E}[Y_i(t)|T_i(z) = t] = \frac{\mathbb{E}[Y_i \cdot D_i^{[t]}|Z_i=z]}{\mathbb{E}[D_i^{[t]}|Z_i=z]}$, which considers all units that take treatment t when $Z_i = z$. This corresponds to Eq. (4) with $\alpha_{z'} = \mathbb{1}(z' = z)$ and $c_g = \mathbb{1}(T_g(z) = t)$.

[7] For an example of a selection model in which $\mathcal{C}(t) \cap \mathcal{C}(t') = \emptyset$, and thus no identified $\Delta_c^{t,t'}$ for $t' \neq t$ exist despite identified μ_c^t existing, see the model described in Proposition 8 of Lee and Salanié (2023).

outcome-nonrestrictive identified if μ_c^1 is, with coefficients α_z that add to zero.

Proposition. *Let $\mathbb{1}_n$ denote a vector of ones in $\mathbb{R}^n$. If $\mathcal{T} = \{0, 1\}$ and $\alpha' \mathbb{1}_{|\mathcal{Z}|} = 0$ and $A^{[1]\prime}\alpha \in \mathcal{C}(1)$, then $A^{[1]\prime}\alpha = -A^{[0]\prime}\alpha$ so that $A^{[0]\prime}(-\alpha) \in \mathcal{C}(0)$ and hence $\mathbb{E}[Y_i(1) - Y_i(0)|c_{G_i}]$ is outcome-nonrestrictive identified.*

As an example, we saw in Example 1 (the LATE model), that using Eq. (4) with with $\alpha = (-1, +1)$, $t = 1$ that $\mathbb{E}[Y_i(1)|i \text{ is complier}] = \frac{\mathbb{E}[Y_i \cdot D_i^{[1]}|Z_i=1] - \mathbb{E}[Y_i \cdot D_i^{[1]}|Z_i=0]}{\mathbb{E}[D_i^{[1]}|Z_i=1] - \mathbb{E}[D_i^{[1]}|Z_i=0]}$. This implies by the above Proposition that we can simply flip the signs of the coefficients to $\alpha = (+1, -1)$ for $t = 0$ to obtain that $\mathbb{E}[Y_i(0)|i \text{ is complier}] = \frac{-\mathbb{E}[Y_i \cdot D_i^{[0]}|Z_i=1] + \mathbb{E}[Y_i \cdot D_i^{[0]}|Z_i=0]}{-\mathbb{E}[D_i^{[0]}|Z_i=1] + \mathbb{E}[D_i^{[0]}|Z_i=0]}$. Combining and using that $D_i^{[0]} + D_i^{[1]} = 1$, we recover the familiar formula that $\mathbb{E}[Y_i(1) - Y_i(0)|i \text{ is complier}] = \frac{\mathbb{E}[Y_i|Z_i=1] - \mathbb{E}[Y_i|Z_i=0]}{\mathbb{E}[D_i|Z_i=1] - \mathbb{E}[D_i|Z_i=0]}$ where $D_i := D_i^{[1]}$. The above Proposition is a restatement of Proposition A.1 in Appendix A.[8]

The unconditional average treatment effect

With a binary treatment, a well-studied parameter of interest is the overall unconditional ATE: $\Delta^{0,1} = \mathbb{E}[Y_i(1) - Y_i(0)]$, i.e. $\Delta_c^{0,1}$ with $c = (1, 1, \ldots 1)'$. For non-binary treatments, an ATE $\Delta^{t,t'} := \mathbb{E}[Y_i(t') - Y_i(t)]$ can be defined between any two treatment values $t, t' \in \mathcal{T}$. As a Corollary of Theorems 1 and 2 we have the following result for unconditional ATEs:

Corollary 1. *Suppose that $\mathcal{Z}$ is finite. Then the unconditional counterfactual mean $\mathbb{E}[Y_i(t)]$ is outcome-nonrestrictive identified given selection*

[8] A simple proof is that since $A^{[0]} = \mathbb{1}_{|\mathcal{Z}|}\mathbb{1}'_{|\mathcal{G}|} - A^{[1]}$, so $\alpha' A^{[0]} = \cancel{\alpha' \mathbb{1}_{|\mathcal{Z}|}} \mathbb{1}'_{|\mathcal{G}|} - \alpha' A^{[1]} = (-\alpha')A^{[1]}$.

model $\mathcal{G}$ *if and only if* $(1, 1, \ldots 1)' \in rs(A^{[t]})$*, and the average treatment effect* $\mathbb{E}[Y_i(t') - Y_i(t)]$ *is outcome-nonrestrictive identified given selection model* $\mathcal{G}$ *if and only if* $(1, 1, \ldots 1)' \in (rs(A^{[t']}) \cap rs(A^{[t]}))$*.*

Since the presence of never-takers with respect to treatment t implies that $(1, 1, \ldots 1)' \notin rs(A^{[t]})$,[9] Corollary 1 implies that ATEs and unconditional counterfactual means are never point-identified in an outcome-nonrestrictive manner absent restrictions on selection.

Corollary 1 also relates my results to recent work by Bai et al. (2024) on the partial identification power of monotonicity for these parameters. Bai et al. (2024) show that selection models can have limited additional identifying power for ATEs provided that they include a restriction that the authors call *generalized monotonicity*, and the outcome is discrete and bounded. I show in Appendix I.1 that under generalized monotonicty, $\mu^t_{(1,1,\ldots 1)}$ can either be point identified for a given t without restrictions on selection, or otherwise $(1, 1, \ldots 1)' \notin rs(A^{[t]})$ (and thus it cannot be point identified without restrictions on outcomes). These results underscore the upside to focusing on target parameters beyond the ATE (i.e. $c \neq (1, 1, \ldots 1)'$) when one is willing to impose restrictions on selection.

4.2 Applying the characterization to search for identified treatment effect parameters

Given Theorem 2, what can we say about the set of possible treatment effect parameters $\mathbb{E}[Y_i(t') - Y_i(t)|c_{G_i} = 1]$ for a given $t' \neq t$ that are

[9] If such never-takers are allowed in $\mathcal{G}$, this introduces a column of all zeroes in the matrix $A^{[t]}$.

outcome-nonrestrictive identified, i.e where $c \in rs(A^{[t]}) \cap rs(A^{[t']}) \cap \{0,1\}^{|\mathcal{G}|}$?

For ease of notation, let us without loss of generality label the treatment values of interest $t' = 1$ and $t = 0$. Accordingly, denote $\alpha^{[t']}$ by α_1 and $\alpha^{[t]}$ by α_0 (each of these is a $|\mathcal{Z}|$-component vector). Then for some $c \in \{0,1\}^{|\mathcal{G}|}$ and $\alpha_0, \alpha_1 \in \mathbb{R}^{|\mathcal{G}|}$, we have an identified conditional treatment effect parameter when $c' = \alpha_1' A^{[1]} = \alpha_0' A^{[0]}$. This occurs if and only if

$$(\alpha_1', -\alpha_0') \begin{bmatrix} A^{[1]} \\ A^{[0]} \end{bmatrix} := \alpha' A^{[1,0]} = \mathbf{0}^{|\mathcal{G}|} \tag{6}$$

with $c' = \alpha_1' A^{[1]}$, where we let $A^{[1,0]}$ denote a $2 \cdot |\mathcal{Z}| \times |\mathcal{G}|$ matrix composed of the rows of $A^{[1]}$ followed by the rows of $A^{[0]}$, and $\alpha = (\alpha_1', -\alpha_0')'$ is a $2 \cdot |\mathcal{Z}| \times 1$ vector. For any α in the left null-space $ns(A^{[1,0]})$ of $A^{[1,0]}$, let $c(\alpha)$ denote the value $c = A^{[1]\prime}\alpha_1 = A^{[0]\prime}\alpha_0$ where α_1 is a vector of the the first $|\mathcal{Z}|$ components of α and α_0 is a vector of minus one times each of the last $|\mathcal{Z}|$ components of α. In general then $\mathcal{C}(t) \cap \mathcal{C}(t') = \{c(\alpha) : \alpha \in ns(A^{[t',t]})\} \cap \{0,1\}^{|\mathcal{G}|}$, where $A^{[t',t]}$ is composed from $A^{[t']}$ and $A^{[t]}$ as above. This characterization proves useful in the search for new IV identification results to follow.

The following result greatly reduces the complexity of building an index of outcome-nonrestrictive identified parameters, searching over vectors α:

Proposition 2. *If $c \in rs(A^{[t]})$ for some t and $c \in \{0,1\}^{|\mathcal{G}|}$, then the equation $c' = \alpha' A^{[t]}$ can be satisfied by a vector α having elements that are rational and belong to the set $\mathcal{C}_n := \left\{\frac{a}{b} : a, b, \in \mathcal{D}_{|\mathcal{Z}|}\right\}$, where $\mathcal{D}_n :=$*

$\{det(B) : B \in \{0,1\}^{n\times n}\}$ *is the set of possible determinant values for an* $n \times n$ *matrix* B *having entries in* $\{0,1\}$.

Proposition 2 implies that when searching for $c = c(\alpha)$, we can always restrict the components of α to belong to the finite set $\mathcal{D}_{|\mathcal{Z}|}$. For $n \leq 7$, the set $\mathcal{D}_n$ is known to consist of consecutive integers symmetric about zero (Craigen, 1990). For example, $\mathcal{D}_1 = \mathcal{D}_2 = \{-1, 0, 1\}$ and $\mathcal{D}_3 = \{-2, -1, 0, 1, 2\}$.[10] It follows that for $|\mathcal{Z}| \leq 2$, we can restrict a search over α_z to the set $\mathcal{C}_1 = \mathcal{C}_2 = \{-1, 0, 1\}$. For $|\mathcal{Z}| = 3$, we can restrict to the set $\mathcal{C}_3 = \{-2, -1, -1/2, 0, 1/2, 1, 2\}$, and so on.

4.3 Algorithms for enumerating identified treatment effects

In this section I implement a brute-force algorithm that uses Proposition 2 above to perform an exhaustive search for local average treatment effect parameters that are outcome-agnostic identified in settings with $|\mathcal{Z}|, |\mathcal{T}| \leq 3$. This search uncovers several novel identification results for treatment effects in IV models.[11] This exercise has clear value within a particular selection model $\mathcal{G}$, as it may reveal new treatment effect parameters that are identified given assumptions about selection that the researcher has already accepted. By iterating over selection models $\mathcal{G}$, it can also suggest relaxations of assumptions regarding selection under which a given parameter remains identified. The search can also be useful to researchers de-

[10] Meanwhile, $\mathcal{D}_4 = \{-3, \ldots, -1, 0, 1, \ldots, 3\}$, $\mathcal{D}_5, = \{-5, \ldots, -1, 0, 1, \ldots, 5\}$, $\mathcal{D}_6 = \{-8, \ldots, -1, 0, 1, \ldots, 8\}$. For $n \geq 8$, $\mathcal{D}_n$ remains a bounded set of integers for any given n, but $\mathcal{D}_n$ generally skips some consecutive integers. For example, it is not possible for a 7×7 binary matrix to have a determinant of 28 but one can achieve a determinant of 32 (Craigen, 1990).

[11] I am grateful to Simon Lee for suggesting this idea to me.

signing experimental interventions, as the amount of instrument variation in the experiment determines $\mathcal{Z}$ (and hence the number of rows available to build $rs(A^{[t]})$). To be clear, the existence of an outcome-nonrestrictive identification result for a given $\mathcal{G}$ should generally not be used as a motivation to search for an ex-post justification of the selection model $\mathcal{G}$. Rather, the plausibility of the restriction to a given $\mathcal{G}$ should be evaluated on its own merits.

I compare two versions of the algorithm, which are laid out explicitly in Appendix E. The first is a "naive" approach that iterates over all possible selection models $\mathcal{G}$ given $\mathcal{Z}$ and $\mathcal{T}$ and then finds identified $\Delta_c^{t,t'}$ within that selection model. For a given selection model $\mathcal{G}$, there are $2^{|\mathcal{G}|}$ possible values of the vector c, and a certificate of whether c corresponds belongs to $\mathcal{C}(t) \cap \mathcal{C}(t')$ for a given t', t can be verified by testing whether $c = c(\alpha)$ for some α in the left nullspace of matrix $A^{[t',t]}$ defined in Eq. (6). A second algorithm makes use of Proposition 2 to instead iterate over the possible $2 \cdot |\mathcal{Z}|$-component vectors α, rather than over selection models $\mathcal{G}$. This comes at great computational benefit, as computations for a single α are useful for studying many selection models at once. Given Proposition 2, we can without loss of generality restrict the search over α to those having components in the discrete and finite set $\mathcal{C}_{|\mathcal{Z}|}$. Compared with Algorithm 1 above, which quickly becomes infeasible for $|\mathcal{Z}| \geq 3$, this second approach runs on $|\mathcal{Z}| = 3$ within minutes. The reason is that the number of possible selection models $2^{|\mathcal{T}|^{|\mathcal{Z}|}}$ scales much more quickly with $|\mathcal{Z}|$ than the number $(\mathcal{C}_{|\mathcal{Z}|})^{2|\mathcal{Z}|}$ of possible α vectors, as shown in Appendix Table E.1.

4.4 Overview of computational results

Table 1 below presents an overview of results of the two algorithms for settings with $|\mathcal{Z}|, |\mathcal{T}| \leq 3$.[12]Appendix F makes illustrative observations from each combination $(|\mathcal{Z}|, |\mathcal{T}|)$ in detail, and a full catalog of the identification results is provided in Appendix K. While the settings reported in Table 1 are "small" $(|\mathcal{Z}|, |\mathcal{T}| \leq 3)$, they turn out to contain a rich structure of identification results, which varies considerably by $\mathcal{T}$ and $\mathcal{Z}$.

$\lvert\mathcal{T}\rvert$	$\lvert\mathcal{Z}\rvert$	# SMs	# TEs	Algorithm 2 run-time		Algorithm 1 run-time
				Initial search	Organizing/Paring	
2	2	2	4	0.08 seconds	.06 seconds	0.11 seconds
3	2	5	5	0.08 seconds	.12 seconds	13.9 seconds
2	3	11	30	55 seconds	.20 seconds	4.3 seconds
3	3	251	251	18 minutes	65 minutes	N/A (estimate: 22 days)

"# TEs" = number of distinct treatment effect parameters identified, "# SMs" = number of distinct maximal selection models. See text below for precise definitions.

Table 1

The third column in Table 1 counts the number of distinct selection models for a given support of the instruments and treatments, for which at least one treatment effect parameter of the form $\Delta_c^{t,t'}$ is outcome-nonrestrictive identified. For instance, the first row finds that Example 1 and Example 2 discussed in Sections 2-3 are the *only* selection models admitting of outcome-nonrestrictive identified treatment effect parameters, with a binary treatment and binary instrument.

The detailed description of Algorithm 2 in Appendix E describes how starting from an identification result for a parameter $\Delta_c^{t,t'}$ using some $\alpha \in \mathbb{R}^{2\cdot|\mathcal{Z}|}$, we can define a "maximal" selection model $\mathcal{G}(\alpha)$ with the property

that α now identifies $\Delta_{\tilde{c}}^{t,t'}$ for some $\tilde{c} \in \{0,1\}^{|\mathcal{G}|}$ within any selection model $\mathcal{G} \subseteq \mathcal{G}(\alpha)$ that is more restrictive that $\mathcal{G}(\alpha)$.[13]

The fourth column in Table 1 counts the number of distinct vectors α for a given support of the instruments and treatments. Each such α yields a distinct identified treatment effect parameter $\Delta_c^{t,t'}$. Although a given α may correspond to identified treatment effect parameters under more restrictive selection models $\mathcal{G} \subseteq \mathcal{G}(\alpha)$ as well (as described above), Table 1 treats these as the same identification result. The preamble to Appendix K provides a detailed example illustrating how the counting is done.

5 Application: interaction effects in cross-randomized designs

This section applies Theorems 1 and 2 to study the identification of complementarities between two binary treatment variables, representing a setting in which $|\mathcal{T}| = |\mathcal{Z}| = 4$ but where some additional structure is natural. Appendix F.1 considers a second application with $|\mathcal{T}| = |\mathcal{Z}| = 4$, where a different structure is natural: spillovers between pairs of observational units and binary treatments.

[12]Run times are with R version 4.3.2 with a 3600MHz processor (AMD Ryzen Threadripper PRO 5975WX), 128GB RAM. While Algorithm 1 is parallelized across 31 cores, Algorithm 2 computation uses a single core. Algorithm 2 is not trivial to parallelize across processors given the need to check for redundancies, but does enable Algorithm 2 to be feasibly extended to $|\mathcal{Z}| = 4$ on this computer setup.

[13]For example, let $\mathcal{G}$ be the choice model described in Section F.0.4 from case ii of Proposition 2 of Kirkeboen et al. (2016). After removing two response types from $\mathcal{G}$, a second treatment effect parameter becomes identified, which is listed under a different selection model $\mathcal{G}' \subset \mathcal{G}$ counted in Table 1.

5.1 Background and empirical practice

In many experimental settings, researchers cross randomize two treatments A and B, and investigate whether there are interaction effects between the treatments, i.e. whether the effect of receiving both A and B differs from the sum of the effects of each of A and B alone. In some such settings instrumental variables methods are not needed, because compliance is perfect or the intent-to-treat effect is the policy-relevant effect of direct interest (see e.g. Duflo et al. 2015; Mbiti et al. 2019). Intent-to-treat (ITT) effects can be straightforwardly estimated by the regression:

$$Y_i = \gamma_0 + \gamma_1 \cdot \mathbb{1}(Z_i = A) + \gamma_2 \cdot \mathbb{1}(Z_i = B) + \gamma_3 \cdot \mathbb{1}(Z_i = C) + \nu_i \qquad (7)$$

where $Z_i = C$ indicates the treatment arm for *both* treatments A and B. Such cross-randomized experiments are often referred to as "factorial designs".[14]

In many factorial designs, the treatment arms $Z_i \in \{A, B, C\}$ represent *offers* for A or B or both, respectively, and researchers obtain data on whether A and B were actually received. For example, Angelucci and Bennett (2024) study the effects of pharmacotherapy (medication) and livelihood assistance (personalized training and support around income generation), among adults with depression in Karnataka, India. Across the three treatment arms of the experiment, roughly 65% of participants actually undertake pharmacotherapy (defined as attending at least one

[14]See Muralidharan et al. (2023) for a review of empirical practice in factorial designs.

psychiatric consultation), receive livelihoods assistance (attending at least one livelihoods workshop), or both. Further many adults assigned to receive both pharmacotherapy *and* assistance undertake only one of the two treatments, although they are offered both. In this setting, individuals do not typically have access to pharmacotherapy or livelihoods assistance except through the field experiment, so the non-compliance is one-sided.

When compliance is imperfect and researchers care about the effects of A and B themselves as treatments (rather than as *offers* of treatment), it is necessary to move beyond intent-to-treat regression (7). Denote the possible treatments as $\mathcal{T} = \{0, A, B, C\}$, with associated potential outcomes $Y_i(t)$ for $t \in \mathcal{T}$. For example, $Y_i(0)$ is the outcome i would experience with neither of the two treatments A and B.

Meanwhile, the instrument values are:

$$\mathcal{Z} = \{\text{offered neither},\ \text{offered just A},\ \text{offered just B},\ \text{offered both}\}$$

If subjects are offered both A and B, they may choose to take treatment A only, treatment B only, or both (treatment C). Treatments A and B are otherwise not available to participants, so non-compliance is one-sided.

For a single individual, we can say that A and B exhibit *complementarity* if the effect of both treatments is greater than the sum of the effects of each treatment separately, i.e. $\{Y_i(C) - Y_i(0)\} > \{Y_i(A) - Y_i(0)\} + \{Y_i(B) - Y_i(0)\}$. Of course, testing for complementarity at the individual

is infeasible due to the fundamental problem of causal inference. Let

$$H_0 : \mathbb{E}[Y_i(C) - Y_i(A) - Y_i(B) + Y_i(0)] = 0 \tag{8}$$

instead be the two-sided hypothesis of no interaction *on average*.[15] Under perfect compliance, H_0 is equivalent to the hypothesis $\gamma_3 - \gamma_1 - \gamma_2 > 0$ from the ITT regression (7). This test is employed for example by Angelucci and Bennett (2024), using only data on assignment and ignoring information about compliance. However, the interpretation of this test may be misleading if compliance is not perfect:

Proposition 3. *If there is imperfect compliance, the parameter* $\gamma_3 - \gamma_1 - \gamma_2$ *in Eq.* (7) *may be zero even when* H_0 *does not hold, and may be non-zero even when* H_0 *holds.*

The intuition behind Proposition 3 is that regression (7) tells us nothing about complementarity effects among individuals who do not align their actual treatments T_i with their treatment assignment Z_i. The result suggests that the common empirical practice of using ITT regressions (rather than focusing on treatment effects per-se) is problematic given that compliance is often known to be far from perfect. However, there are limited identification results for researchers to make use of to estimate interaction effects with imperfect compliance and effect heterogeneity.

One solution is to restrict outcomes, assuming sufficient treatment effect homogeneity to get around Proposition 3. For example, if we assume

[15] The LHS of (8) is the average of the individual-level "interaction effect" $\{Y_i(C) - Y_i(0)\} - \{Y_i(A) - Y_i(0)\} + \{Y_i(B) - Y_i(0)\} = Y_i(C) - Y_i(A) - Y_i(B) + Y_i(0)$.

that no selection on gains (NSOG) holds, the four unconditional counterfactual means $\mathbb{E}[Y_i(C)]$, $\mathbb{E}[Y_i(B)]$, $\mathbb{E}[Y_i(A)]$, and $\mathbb{E}[Y_i(0)]$ are identified under general conditions given in Appendix C. Identification is constructive and corresponds to the estimand of a two-stage least squares (2SLS) regression of Y_i on indicators for each of the four treatments (and no constant), instrumented by indicators for each of the four treatment assignment arms.[16] We can then test H_0 by testing $\beta_3 = 0$ in the equation $Y_i = \beta_0 + \beta_1 \cdot \mathbb{1}(T_i \in \{A, C\}) + \beta_2 \cdot \mathbb{1}(T_i \in \{B, C\}) + \beta_3 \cdot \mathbb{1}(T_i = C) + \epsilon_i$, estimated using the instruments $\mathbb{1}(Z_i = A)$, $\mathbb{1}(Z_i = B)$, $\mathbb{1}(Z_i = \text{both})$ and a constant.

Nevertheless, NSOG is a very restrictive assumption. It suggests for example that individuals do not have any knowledge of their specific gains from the various treatments which informs their selection behavior. When NSOG does not hold, Kormos et al. (2024) detail how the 2SLS estimand β_3 generally mixes interaction effects with terms that simply reflect treatment effect heterogeneity. It is thus desirable to pursue an alternative approach that leads to an interpretable causal estimand without restricting outcomes.

[16] In particular, since there are four instrument values and four treatment values, we can use a result derived in Appendix C under NSOG, that $\mathbb{E}[Y_i(t)] = \sum_z \Sigma_{tz}^{-1} \cdot \mathbb{E}[Y_i \cdot \mathbb{1}(Z_i = z)]$, provided that the matrix with entries $\Sigma_{zt} = P(Z_i = z, T_i = t)$ is invertible. Some algebra shows that this coincides with the two-stage least squares estimand mentioned above.

5.2 Identifying the local average interaction effect among compliers

We now use Theorems 1 and 2 to examine to what extent NSOG can be meaningfully relaxed. Ex-ante, there are $2 \times 2 \times 4 = 16$ response types that respect one-sided non-compliance.[17] However, assuming that the weak-axiom of revealed preference (WARP) holds, we are left with the nine response groups enumerated in Table 2.[18] I refer to these nine response types as $\mathcal{G}^{WARP}$. $\mathcal{G}^{WARP}$ represents the weakest selection model consistent with rational choice and one-sided non-compliance in a factorial design.

offered ↓	n.t.	complier	A only	B only	only both	A+	B+	favor A	favor B
neither	0	0	0	0	0	0	0	0	0
just A	0	A	A	0	0	A	0	A	A
just B	0	B	0	B	0	0	B	B	B
both	0	C	A	B	C	C	C	A	B

Table 2: Response types that satisfy WARP in the cross-randomized offer design. The columns in black correpond to the response types allowed by Proposition 4, while the gray columns correspond to the remaining response types that are compatible with WARP.

Given a selection model $\mathcal{G}$ and a function $c : \mathcal{G} \to \{0, 1\}$, let us refer to $LAIE(c) := \mathbb{E}[Y_i(C) - Y_i(A) - Y_i(B) + Y_i(0) | c(G_i) = 1]$ as the *local average interaction effect* among the subgroup of $g \in \mathcal{G}$ such that $c(g) = 1$. LAIEs are causal quantities like the local treatment effect parameters introduced in Section A, except that they involve the potential outcomes for all four

[17] These response types correspond to the choices individuals would make across three decisions: whether to take treatment A if A only is offered, B if B only is offered, and which of the four treatment combinations to take if both are offered.

[18] The reduction from 16 to 9 response types comes from the additional restrictions that $\{T_i(\text{offered both}) = A \implies T_i(\text{offered A}) = A\}$, $\{T_i(\text{offered both}) = B \implies T_i(\text{offered B}) = B\}$, $\{T_i(\text{offered both}) = 0 \implies T_i(\text{offered A}) = T_i(\text{offered B}) = 0\}$, $\{T_i(\text{offered A}) = A \implies T_i(\text{offered both}) \neq 0\}$, and $\{T_i(\text{offered B}) = B \implies T_i(\text{offered both}) \neq 0\}$.

treatments rather than just two. The terminology "LAIE" follows that used in Blackwell (2017) and Kormos et al. (2024).

The following Proposition uses Theorems 1 and 2 to establish when $LAIE(c)$ is identified in a manner that does not restrict outcomes:

Proposition 4. *Given one-sided noncompliance and WARP, $LAIE(c)$ is outcome-nonrestrictive identified if and only if $c(g) = \mathbb{1}(g = complier)$ and $\mathcal{G} \subseteq \{n.t., complier, A\ only, B\ only\}$.*

Proposition 4 follows from a brute-force enumeration over all of the 511 selection models $\mathcal{G} \subseteq \mathcal{G}^{WARP}$, and the $c \in \{0,1\}^{|\mathcal{G}|}$ within each of them. Given such a selection model $\mathcal{G}$, Theorems 1 and 2 along with an extension of Proposition 1 to parameters that involve more than two treatment states (proved in Appendix B) shows that $LAIE(c)$ is outcome non-restrictive identified iff $c \in rs(A^{[0]}) \cap rs(A^{[A]}) \cap rs(A^{[B]}) \cap rs(A^{[C]})$.

Proposition 4 establishes that identifying complementarities in a cross-randomized design without outcome restrictions requires substantive restrictions on selection: many of the response types in $\mathcal{G}^{WARP}$ not included in $\{\text{n.t.}, \text{complier}, \text{A only}, \text{B only}\}$ are ex-ante plausible. For example, let $U_i(t)$ denote the interpret $U_i(t)$ as the net utility of treatment $t \in \mathcal{T}$ relative to no treatment for individual i (thus normalizing $U_i(0) = 0$). Without loss of generality, consider a random coefficients form for the utility function: $U_i(t) = \pi_{Ai} \cdot \mathbb{1}(t = A) + \pi_{Bi} \cdot \mathbb{1}(t = B) + \pi_{Ci} \cdot \mathbb{1}(t = C)$. If the vector $\pi_i = (\pi_{Ai}, \pi_{Bi}, \pi_{Ci})'$ has support in an open neighborhood of the origin in $\mathbb{R}^3$, all nine groups from Table 2 will be present in the population.

Appendix J.1 shows that we can rationalize the restriction made in Proposition 4 by supposing that individuals choose *separately* whether to receive treatment A or B, rather than as a single joint decision. That is, individuals choose as if they evaluate the costs and benefits of each treatment A or B separately, and choose all treatments offered to them for which benefits outweigh costs. For this reason, let us denote the largest selection model in which the local average interaction effect among compliers is identified as $\mathcal{G}^{sep} := \{\text{n.t.}, \text{complier}, \text{A only}, \text{B only}\}$. Given one-sided noncompliance, the selection model $\mathcal{G}^{sep}$ is also equivalent to what Blackwell (2017) calls a "treatment exclusion" restriction that the instrument for treatment A does not affect uptake of treatment B (and vice versa). Blackwell (2017) shows that in this case the interaction coefficient β_3 from a 2SLS regression identifies the local average interaction effect among compliers. Proposition 4 shows that treatment exclusion is furthermore *necessary* to identify this parameter without restricting outcomes. That is, given WARP and one-sided noncompliance, the assumption of treatment exclusion cannot be relaxed without restricting outcomes.

Estimating treatment effects and interaction effects among compliers: The function c corersponding to the LAIE among compliers is $c(g) = \mathbb{1}(g = \text{complier})$, or in vector notation $c' = (0, 1, 0, 0)'$. Let us

denote the functions $\alpha^{[t]}(z)$ in vector form as α_t, in which case

$$\alpha_0 = (1,-1,-1,1)', \quad \alpha_A = (0,1,0,-1)',$$
$$\alpha_B = (0,0,1,-1)', \quad \alpha_C = (0,0,0,1)' \tag{9}$$

One can verify directly that for each $t \in \mathcal{R}$, $\alpha_t' A^{[t]} = (0,1,0,0)$ where the matrix A is defined from the first four columns of Table 2. For brevity, let us denote $LAIE((0,1,0,0)') = \mathbb{E}[Y_i(C) - Y_i(A) - Y_i(B) + Y_i(0)|g = \text{complier}]$ as simply LAIE (with this c implicit).

Applying Eq. (4) with the coefficients in (9) implies a cumbersome expression for LAIE, but some simplification shows that $LAIE = \theta^{ITT}/p$, where $p = P(G_i = \text{complier})$ and $\theta^{ITT} := \gamma_3 - \gamma_1 - \gamma_2$ is the measure of average complimentary from the intent-to-treat regression Eq. (7). This delivers the following useful consequence of Proposition 4:

Corollary 2. *Given $\mathcal{G} \subseteq \mathcal{G}^{sep}$, the sign of the local average interaction effect among compliers $LAIE$ is the same as $\gamma_3 - \gamma_1 - \gamma_2$ from the intent-to-treat regression* (7)*.*

The algebra that leads to $LAIE = \theta^{ITT}/p$ is given in Appendix J.5, where it is also extended to the case in which covariates are included in Eq. (7).

Thus while the ITT condition $\gamma_3 - \gamma_1 - \gamma_2$ cannot be used to test the hypothesis of overall unconditional complementarity (without outcome restrictions), it can by Propisition 2 be used to test for the sign of local average interaction effect among compliers. This latter interpretation requires no outcome restrictions, but instead the non-trivial selection model

$\mathcal{G} \subseteq \mathcal{G}^{sep}$. Corollary 2 thereby formally justifies the test for complementarity used by Angelucci and Bennett (2024) within this selection model.

I note finally that the selection model $\mathcal{G} \subseteq \mathcal{G}^{sep}$ yields three overidentification restrictions for the share of compliers, obtained by applying Eq. (3) with (9).[19] This testable implication is new to the literature and can be used to assess the substantive assumption $\mathcal{G} \subseteq \mathcal{G}^{sep}$.[20]

5.3 Empirical application

I use the replication data from Angelucci and Bennett (2024) as an empirical implementation of the above findings. First, I assess the testable implication above of $\mathcal{G} \subseteq \mathcal{G}^{sep}$. Unable to reject the over-identifying restrictions, I estimate LAIE following Proposition 4.

In Appendix J.7 I consider a second empirical setting, from Angrist et al. (2009), in which students were cross-randomized into academic support and financial incentives for good grades. In that setting, I find by contrast that the testable implication of $\mathcal{G} \subseteq \mathcal{G}^{sep}$ is rejected, and therefore local average interaction effect parameters cannot be identified absent restrictions on outcomes. This illustrates that the general over-identifying restrictions highlighted in Section A.5 have power in an empirically relevant way.

[19] The restriction is

$$\begin{aligned} p &:= P(T_i = C | Z_i = \text{both}) \\ &= P(T_i = A | Z_i = \text{just A}) - P(T_i = A | Z_i = \text{both}) \\ &= P(T_i = B | Z_i = \text{just B}) - P(T_i = B | Z_i = \text{both}) \\ &= 1 + P(T_i = 0 | Z_i = \text{both}) - P(T_i = 0 | Z_i = \text{just A}) - P(T_i = 0 | Z_i = \text{just B}) \end{aligned} \tag{10}$$

for some value $p \in [0, 1]$ which identifies $P(G_i = \text{complier})$.

[20] These are stronger than testable implications mentioned by Blackwell (2017), which give $P(A \text{ or } C | \text{both}) = P(A | \text{just A})$ and $P(B \text{ or } C | \text{both}) = P(B | \text{just B})$ in the case of one-sided noncompliance. Those do not imply the last line of Eq. (10).

Since the experiment reported in Angelucci and Bennett (2024) stratifies randomization into nine strata (defined by district and terciles of a village poverty index), implementation requires some extensions to the basic results of this paper that allow randomization to hold based on observed covariates X_i. As described in Appendix J.3, conditional expectations that need to be estimated are assumed to be additively separable between instruments Z_i and indicators for the strata X_i for simplicity.

Testable implications of $\mathcal{G} \subseteq \mathcal{G}^{sep}$: Each of the four expressions the proportion p of compliers in Eq. (10) can be estimated using regressions of the various $D_i^{[t]}$ on instrument indicators as well as indicators for strata X_i. The extension of (10) to the case with strata fixed effects is given in Appendix J.3. Following Angelucci and Bennett (2024), I use cluster robust inference by village (the level of treatment assignment).

The point estimates for $p := P(G_i = \text{complier})$ are 36.7%, 40.2%, 39.7%, and 43.7%, respectively. A chi-squared test for equality of all four estimates of p yields a p-value of 65%. This indicates that we cannot reject these overidentification restrictions at all conventional levels. This provides evidence in favor of the choice model $\mathcal{G} \subseteq \mathcal{G}^{sep}$.

However, the equality restrictions (10) are not the only observable implications of $\mathcal{G} \subseteq \mathcal{G}^{sep}$. In Appendix J.6, I describe how all of the observable first-stage information can be aggregated into a system of linear equations $\mathcal{A}x = \beta$, where $\mathcal{A}$ is a known matrix defined from the $A^{[t]}$, β is a vector of observed treatment choice probabilities, and x is a vector of the (non-

negative) unobserved occupancies $x_g = P(G_i = g)$ of each response type. Maintaining the weaker assumption of $\mathcal{G} \subseteq \mathcal{G}^{WARP}$, we can test whether $\mathcal{G}$ is furthermore a subset of $\mathcal{G}^{set}$ by computing a lower bound on the sum of the components of x_g for $g \in \mathcal{G}^{WARP} - \mathcal{G}^{sep}$, subject to the constraints that $\mathcal{A}x = \beta$, each $x_g \geq 0$ and the x_g sum to unity.[21]

Solving this linear program with point estimates of the observed treatment choice probabilities suggests that $P(G_i \in \mathcal{G}^{WARP} - \mathcal{G}^{sep})$ is at least 6.3% (and is no more than 80.8%). In the absence of sampling uncertainty, this would provide some evidence against the restriction $\mathcal{G} \subseteq \mathcal{G}^{sep}$. However, this lower bound for $P(G_i \in \mathcal{G}^{WARP} - \mathcal{G}^{sep})$ is not statistically significant. Fang et al. (2023) provide a method for testing whether there exists componentwise non-negative solutions x to systems of the form $\mathcal{A}x = \beta$ like the above, when β is estimated from the data. This method yields a 95% confidence interval of $[0, 0.83]$ for the share of offending response types. We therefore cannot reject that $\mathcal{G} \subseteq \mathcal{G}^{sep}$ within the weaker assumption $\mathcal{G} \subseteq \mathcal{G}^{WARP}$, even using the full observable information on treatment uptake. Appendix J.6 provides further details.[22]

[21] In principle, this exercise could be implemented by strata to test $\mathcal{G} \subseteq \mathcal{G}^{sep}$ among the individuals within each. To increase statistical power given the small sample, I pool the data across all strata for this exercise. This is valid under the assumption that the response-type distribution is common across strata.Note that this resrtiction does not require *potential outcomes* to be uncorrelated with stratum.

[22] I implement the FSST method using the `R` package `lpinfer`. This method does not involve any clustering and is designed for *i.i.d.* data, so the confidence interval reported above may undercover the parameter $P(G_i \in \mathcal{G}^{WARP} - \mathcal{G}^{sep})$ if one considers uncertainty as arising from treatment assignment as well. Since the proportion of each cluster (in this case village) that is sampled is small (on average about two individuals), results for OLS suggest that the influence of clustering in treatment assignment may be minimal, even considering both uncertainty arising from clustered treatment assignment as well as sampling (Abadie et al., 2022). Similar results are provided using alternative methods for inference on linear systems introduced by Romano and Shaikh (2008) and Cho and Russell (2024).

Estimates of local average interaction among compliers: The data from Angelucci and Bennett (2024) follow 1,000 respondents over five survey waves. I use their main outcome variable, which is a standardized version of the PHQ-9 score for depression, with higher values indicating more severe depression. I focus on longer-run outcomes in the fourth and fifth survey waves, which occured between one and two years after treatment. In these longer waves, the authors estimate $\gamma_3 - \gamma_1 - \gamma_2$ from the ITT regression to be marginally significant at the 10% level with a p-value of .10. Meanwhile, they find that the combination (treatment "C") of pharmacotherapy (treatment "A") and livelihoods assistance (treatment "B") reduces depression symptoms even after the intervention that is significant at the 95% level, while the effects of treatments A or B alone are insignificant (cf. their Table 2, panel B). These estimates however come from an ITT regression that ignores actual treatment uptake, and may be attenuated or otherwise distorted when interpreted as effects of the treatments themselves rather than as effects of assignment.

Column (1) of Table 3 implements this ITT regression of the outcome on instrument indicators (and strata fixed effects). Departing slightly from Angelucci and Bennett (2024), I focus on a minimal specification and do not control for baseline values of the outcome. However, the findings are qualitatively the same and quantitatively similar. In line with Angelucci and Bennett (2024) only the effect of treatment C (pharmacotherapy and livelihoods assistance) is statistically significant. Column (2) uses data on treatment uptake and implements a 2SLS regression as described in Section

	(1) ITT	(2) 2SLS	(3) Eq. (4)	(4) GMM
$\mathbb{E}[Y(C) - Y(0)\|c(G) = 1]$	-0.28**	-0.68**	-0.49*	-0.29*
	(0.08)	(0.22)	(0.21)	(0.11)
$\mathbb{E}[Y(A) - Y(0)\|c(G) = 1]$	-0.05	-0.09	-0.07	-0.09
	(0.08)	(0.15)	(0.21)	(0.21)
$\mathbb{E}[Y(B) - Y(0)\|c(G) = 1]$	-0.03	-0.04	0.13	0.25
	(0.08)	(0.10)	(0.24)	(0.17)
Interaction effect parameter	AIE	AIE/LAIE	LAIE	LAIE
Interaction point estimate	-0.20	-0.55	-0.56	-0.50
P-val: no complementarity	0.10	0.08	0.07	0.15
c(G)	compliers/all	all individuals	compliers	compliers
p(c(G)=1)	1	1	.4	.41
Identifying assumption	perfect compliance	NSOG	$\mathcal{G} \subseteq \mathcal{G}^{sep}$	$\mathcal{G} \subseteq \mathcal{G}^{sep}$
Sample size	1650	1650	1650	1650

Table 3: Treatment effects and interaction effect estimates, where A is pharmacotherapy ("PC"), treatment B is livelihoods assistance ("LA"), and treatment C is receiving both. Outcome variable is the PHQ-9 depression score, expressed in units of its sample standard deviation. ITT effect estimates can be interpreted as effects of receiving treatment under perfect compliance, though this is rejected by the data. 2SLS estimates assume no-selection-on-gains (NSOG). The interaction effect estimated in columns (1) or (2) is the overall average interaction effect (AIE) appearing in Eq. (8), and in (3) and (4) are the local average interaction effect (LAIE) among compliers. All columns include strata controls and cluster standard errors by village.

5.1. These treatment effects estimates are larger in magnitude and have the same pattern of significance, which is intuitive given imperfect compliance. However the main treatment effect estimates (besides the LAIE) in Column (2) invoke the strong assumption of no-selection-on-gains (NSOG) to be interpreted causally.[23]

By contrast, none of the estimates reported in Columns (3) and (4) require no restrictions on outcomes to be causally interpreted and compared. Column (3) uses simple sample estimators of the expectations from Eq. (4) (extended for strata fixed effects) along with the α vectors in Eq. (9) that isolate compliers. (see Appendix J.3 for details). Column (4) re-estimates

[23]Thm. 2 of Blackwell (2017) shows how the various 2SLS coefficients average effects over different groups of response types even given $\mathcal{G} \subseteq \mathcal{G}^{sep}$, making them not comparable to one another without outcome restrictions like NSOG.

Column (3) while further imposing the overidentification restrictions (10) for the share of compliers, using a generalized method of moments (GMM) estimator (see Appendix J.4 for details). While the GMM estimator does not reduce the standard error of LAIE in this setting, it yields a statistically significant estimate of $\mathbb{E}[Y_i(C) - Y_i(0) | i \textit{ is complier}]$ at the 95% level.

The three estimates of $LAIE$ in columns (2)-(4) are valid under the same assumption that $\mathcal{G} \subseteq \mathcal{G}^{sep}$, and suggest that pharmacotherapy and livelihoods assistance are complementary: they have an interaction effect of about half of a standard deviation of PHQ-9 among compliers. 2SLS provides the most precise estimate of this parameter, which is signficant at the 10% level.

The large positive magnitude of the effect of livelihoods assistance (treatment B) in columns (3) and (4) raises the question of whether this intervention may in fact *exacerbate* depression symptoms among compliers, when it is not accompanied by pharmacotherapy (treatment A). This finding is not evident in the ITT estimates from Angelucci and Bennett (2024) that do not adjust for non-compliance, or from 2SLS results. The estimate is not quite significant at the 10% level even with the GMM estimator however (t-statistic $2.47/1.67 = 1.48$), so this finding should be caveated accordingly. The estimates reported in Table 3 otherwise confirm the qualitative findings of Angelucci and Bennett (2024), while offering quantitative treatment effects that account for the partial compliance.

6 Conclusion

This paper has shown that given discrete instruments, outcome non-restrictive identification using instrumental variables is equivalent to the existence of linear combinations of counterfactual treatment indicators that add up to zero or one for all response types in the assumed selection model. A selection model only allows for treatment effects to be identified in an outcome-nonrestrictive way when a matrix that summarizes the selection behavior allowed by the model satisfies a particular geometric property.[24]

This insight yields a systematic approach to enumerating all selection models that afford identification of treatment effects in a manner that does not restrict outcomes. The search delivers a multiplicity of new identification results, despite its computational complexity scaling rapidly with the size of support of the instruments and treatment. Perhaps more importantly, it traces out the *limits* of point identification leveraging restrictions on selection alone. This is illustrated in the application to cross-randomized assignment to two binary treatments, where I establish that substantive restrictions on choice cannot be relaxed without sacrificing point identification of the local average interaction effect.

References

ABADIE, A., ATHEY, S., IMBENS, G. W. and WOOLDRIDGE, J. M. (Oct. 2022). "When Should You Adjust Standard Errors for Clustering?" *The Quarterly Journal of Economics* 138 (1), pp. 1–35.

[24] In particular, that the matrix $A^{[1,0]}$ from Section 4.2 with respect to two particular treatment values has a non-trivial null-space that intersects the unit cube in the space of types allowed by the model.

ANGELUCCI, M. and BENNETT, D. (2024). "The Economic Impact of Depression Treatment in India: Evidence from Community-Based Provision of Pharmacotherapy". *American Economic Review* 114 (1), 169–98.

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

ARORA, A., GOFF, L. and HJORT, J. (2021). "Pure-Chance Jobs versus a Labor Market: The Impact on Careers of a Random Serial Dictatorship for First Job Seekers". *AEA Papers and Proceedings* 111, pp. 470–75.

BAI, Y., HUANG, S., MOON, S., SANTOS, A., SHAIKH, A. M. and VYTLACIL, E. J. (2025). *Inference for Treatment Effects Conditional on Generalized Principal Strata using Instrumental Variables*. arXiv: 2411.05220 [econ.EM].

BAI, Y., HUANG, S., MOON, S., SHAIKH, A. M. and VYTLACIL, E. J. (2024). *On the Identifying Power of Monotonicity for Average Treatment Effects*. arXiv: 2405.14104 [econ.EM].

BLACKWELL, M. (2017). "Instrumental Variable Methods for Conditional Effects and Causal Interaction in Voter Mobilization Experiments". *Journal of the American Statistical Association* 112 (518), pp. 590–599.

CARD, D. (2022). "Design-Based Research in Empirical Microeconomics". *American Economic Review* 112 (6), 1773–81.

CHO, J. and RUSSELL, T. M. (2024). "Simple Inference on Functionals of Set-Identified Parameters Defined by Linear Moments". *Journal of Business & Economic Statistics* 42 (2), pp. 563–578.

COMEY, M. L., ENG, A. R. and PEI, Z. (2023). *Supercompliers*. arXiv: 2212.14105 [econ.EM].

CRAIGEN, R. (1990). "The Range of the Determinant Function on the Set of n X n (0,1)-Matrices". *Journal of Combinatorial Mathematics and Combinatorial Computing* 8.

DE CHAISEMARTIN, C. (2017). "Tolerating defiance? Local average treatment effects without monotonicity". *Quantitative Economics* 8 (2), pp. 367–396.

DUFLO, E., DUPAS, P. and KREMER, M. (2015). "Education, HIV, and Early Fertility: Experimental Evidence from Kenya". *American Economic Review* 105 (9), 2757–97.

FAN, Y. and PARK, S. S. (2010). "Sharp Bounds on the Distribution of the Treatment Effect and Their Statistical Inference". *Econometric Theory* 26 (3), pp. 931–951.

FANG, Z., SANTOS, A., SHAIKH, A. M. and TORGOVITSKY, A. (2023). "Inference for Large-Scale Linear Systems With Known Coefficients". *Econometrica* 91 (1), pp. 299–327.

GOFF, L. (2024). "A Vector Monotonicity Assumption for Multiple Instruments". *Journal of Econometrics* 241 (1).

HECKMAN, J. J. and PINTO, R. (2018). "Unordered Monotonicity". *Econometrica* 86 (1), pp. 1–35.

HECKMAN, J. J., URZUA, S. and VYTLACIL, E. J. (2006). "Understanding Instrumental Variables in Models with Essential Heterogeneity". *The Review of Economics and Statistics* 88 (3), pp. 389–432.

HECKMAN, J. J. and VYTLACIL, E. (2005). "Structural Equations, Treatment Effects, and Econometric Policy Evaluation". *Econometrica* 73 (3), pp. 669–738.

HECKMAN, J. J. and VYTLACIL, E. J. (2001). "Local Instrumental Variables". *Nonlinear Statistical Modeling: Proceedings of the Thirteenth International Symposium in Economic Theory and Econometrics: Essays in Honor of Takeshi Amemiya.*

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

KIRKEBOEN, L. J., LEUVEN, E. and MOGSTAD, M. (May 2016). "Field of Study, Earnings, and Self-Selection*". *The Quarterly Journal of Economics* 131 (3), pp. 1057–1111.

KLINE, P. and WALTERS, C. R. (July 2016). "Evaluating Public Programs with Close Substitutes: The Case of Head Start*". *The Quarterly Journal of Economics* 131 (4), pp. 1795–1848.

KOLESÁR, M. (2013). "Estimation in an Instrumental Variables Model With Treatment Effect Heterogeneity". Working paper.

KORMOS, M., LIELI, R. P. and HUBER, M. (2024). *Interacting Treatments with Endogenous Takeup*. arXiv: 2301.04876 [econ.EM].

LEE, S. and SALANIÉ, B. (2018). "Identifying Effects of Multivalued Treatments". *Econometrica* 86 (6), pp. 1939–1963.

LEE, S. and SALANIÉ, B. (2023). *Treatment Effects with Targeting Instruments*. arXiv: 2007.10432 [econ.EM].

LEWBEL, A. (2019). "The Identification Zoo: Meanings of Identification in Econometrics". *Journal of Economic Literature* 57 (4), pp. 835–903.

MATZKIN, R. (2007). "Nonparametric identification". *Handbook of Econometrics*. Ed. by J. HECKMAN and E. LEAMER. 1st ed. Vol. 6B. Elsevier. Chap. 73.

MBITI, I., MURALIDHARAN, K., ROMERO, M., SCHIPPER, Y., MANDA, C. and RAJANI, R. (Apr. 2019). "Inputs, Incentives, and Complementarities in Education: Experimental Evidence from Tanzania*". *The Quarterly Journal of Economics* 134 (3), pp. 1627–1673.

MELO, N. and WINTER, A. (2019). "Intersection patterns of linear subspaces with the hypercube". *Journal of Combinatorial Theory, Series A* 164, pp. 60–71.

MURALIDHARAN, K., ROMERO, M. and WÜTHRICH, K. (Mar. 2023). "Factorial Designs, Model Selection, and (Incorrect) Inference in Randomized Experiments". *The Review of Economics and Statistics*, pp. 1–44.

NAVJEEVAN, M., PINTO, R. and SANTOS, A. (2023). *Identification and Estimation in a Class of Potential Outcomes Models*. arXiv: 2310.05311 [econ.EM].

VAN'T HOFF, N., LEWBEL, A. and MELLACE, G. (May 2023). *Limited Monotonicity and the Combined Compliers LATE*. Boston College Working Papers in Economics 1059.

ROMANO, J. P. and SHAIKH, A. M. (2008). "Inference for identifiable parameters in partially identified econometric models". *Journal of Statistical Planning and Inference* 138 (9). Special Issue in Honor of Theodore Wilbur Anderson, Jr. on the Occasion of his 90th Birthday, pp. 2786–2807.

SIGSTAD, H. (2024). *Marginal Treatment Effects and Monotonicity*. arXiv: 2404.03235 [econ.EM].

SUN, Z. and WÜTHRICH, K. (2024). *Pairwise Valid Instruments*. arXiv: 2203.08050 [econ.EM].