On the Identifying Power of Generalized Monotonicity for Average Treatment Effects*

Yuehao Bai
Department of Economics
University of Southern California
yuehao.bai@usc.edu

Shunzhuang Huang
Booth School of Business
University of Chicago
shunzhuang.huang@chicagobooth.edu

Sarah Moon
Department of Economics
MIT
sarahmn@mit.edu

Azeem M. Shaikh
Department of Economics
University of Chicago
amshaikh@uchicago.edu

Edward J. Vytlacil
Department of Economics
Yale University
edward.vytlacil@yale.edu

November 18, 2025

Abstract

In the context of a binary outcome, treatment, and instrument, Balke and Pearl (1993, 1997) establish that the monotonicity condition of Imbens and Angrist (1994) has no identifying power beyond instrument exogeneity for average potential outcomes and average treatment effects in the sense that adding it to instrument exogeneity does not decrease the identified sets for those parameters whenever those restrictions are consistent with the distribution of the observable data. This paper shows that this phenomenon holds in a broader setting with a multi-valued outcome, treatment, and instrument, under an extension of the monotonicity condition that we refer to as generalized monotonicity. We further show that this phenomenon holds for any restriction on treatment response that is stronger than generalized monotonicity provided that these stronger restrictions do not restrict potential outcomes. Importantly, many models of potential treatments previously considered in the literature imply generalized monotonicity, including the types of monotonicity restrictions considered by Kline and Walters (2016), Kirkeboen et al. (2016), and Heckman and Pinto (2018), and the restriction that treatment selection is determined by particular classes of additive random utility models. We show through a series of examples that restrictions on potential treatments can provide identifying power beyond instrument exogeneity for average potential outcomes and average treatment effects when the restrictions imply that the generalized monotonicity condition is violated. In this way, our results shed light on the types of restrictions required for help in identifying average potential outcomes and average treatment effects.

KEYWORDS: Multi-valued Treatments, Average Treatment Effects, Endogeneity, Instrumental Variables.

JEL classification codes: C31, C35, C36

*We thank the associate editor and three anonymous referees, as well as Sukjin Han, Sokbae Lee, Derek Neal, Kirill Ponomarev, Vitor Possebom, Bernard Salanié and Panos Toulis for helpful comments. Shaikh acknowledges financial support from National Science Foundation Grant SES-2419008. Vytlacil acknowledges financial support from the Tobin Center for Economic Policy at Yale. Moon acknowledges financial support from the National Science Foundation Graduate Research Fellowship under Grant No. 2141064.

1 Introduction

In their analysis of a setting with a binary outcome, treatment, and instrument, Balke and Pearl (1993, 1997) establish that the monotonicity condition of Imbens and Angrist (1994) has no identifying power beyond instrument exogeneity for average potential outcomes and the average treatment effect (ATE). Here, by no identifying power beyond instrument exogeneity, we mean that adding the monotonicity condition of Imbens and Angrist (1994) to instrument exogeneity does not decrease the identified sets for those parameters whenever those restrictions are consistent with the distribution of the observable data.[1] In this way, their results contrast with the analysis of Imbens and Angrist (1994), who showed that their monotonicity condition and instrument exogeneity permitted identification of the local average treatment effect (LATE). This paper studies the extent to which this phenomenon holds in the broader context of a multi-valued outcome, treatment, and instrument.

We show that a generalization of the monotonicity condition of Imbens and Angrist (1994) to this richer setting also has no identifying power beyond instrument exogeneity for average potential outcomes and ATEs. We hereafter refer to this condition more succinctly as generalized monotonicity. We show further that this result remains true for any restriction that is in fact *stronger* than generalized monotonicity provided that these stronger restrictions do not restrict potential outcomes in a sense that we will make precise later. This feature of our results is remarkable because one might expect stronger restrictions, possibly with very complicated restrictions on potential treatments that are difficult to fully characterize, to reduce the identified set at least in some instances, but we show that this is not the case. Using this result, our analysis accommodates many examples of restrictions on potential treatments that have been previously considered in the literature. In particular, we show that encouragement designs, the types of monotonicity restrictions considered by Kline and Walters (2016), Kirkeboen et al. (2016), and Heckman and Pinto (2018), and certain additive random utility models, including some studied in Lee and Salanié (2023), all satisfy generalized monotonicity.

In establishing our results, we derive the identified sets for average potential outcomes under any such restriction and instrument exogeneity while maintaining the assumption that these restrictions are consistent with the distribution of the observable data. Our derivations reveal that the form of the resulting identified sets parallels the form of those derived by Balke and Pearl (1993, 1997) for a binary outcome, treatment, and instrument. An implication of the form of the identified sets is that average potential outcomes and ATEs are *only* identified under an identification-at-infinity-type condition when imposing instrument exogeneity and any such restriction. In our analysis, we also derive the identified sets for average potential outcomes and ATEs when imposing instrument exogeneity alone whenever the distribution of the observable data is consistent with instrument exogeneity and generalized monotonicity. As we explain in Example 4.1, this consistency is necessarily satisfied, for example, in the context of a multi-arm randomized controlled trial with one-sided non-compliance when defining the instrument to be random assignment to a given treatment arm.

[1] For settings with possibly non-binary outcomes, Kitagawa (2021) shows this phenomenon continues to hold for any parameter that is a function of the marginal distributions of potential outcomes.

Our results further provide necessary conditions on restrictions on potential treatments to help in identifying average potential outcomes and ATEs. See Theorem 3.3 and the subsequent discussion for details. We illustrate this phenomenon through a series of examples of models that need not satisfy generalized monotonicity and have identifying power for average potential outcomes and ATEs.

Our paper differs from the closely related literature that, in the context of a binary outcome, treatment, and instrument, considers the identifying power of the monotonicity condition of Imbens and Angrist (1994) and instrument exogeneity for the distribution (as opposed to the average) of potential outcomes, or considers the identifying power of these conditions when combined with additional restrictions on potential outcomes. In particular, Kamat (2019) shows that the monotonicity condition of Imbens and Angrist (1994) does have identifying power beyond instrument exogeneity for the (joint) distribution of potential outcomes. Machado et al. (2019) show that the monotonicity condition of Imbens and Angrist (1994) does have additional identifying power for the ATE beyond instrument exogeneity if one additionally imposes an assumption that requires potential outcomes to vary monotonically with the treatment. Thus, the phenomenon we explore is sensitive to both the choice of parameter and to whether one imposes assumptions on potential outcomes.

The remainder of the paper is organized as follows. Section 2 introduces our formal setup, notation and assumptions, including our generalized monotonicity condition. Our main identification results are presented in Section 3. In Section 4, we provide several examples of restrictions on potential treatments that imply our generalized monotonicity condition, and are thus examples of restrictions that have no identifying power beyond instrument exogeneity for average potential outcomes or ATEs. In contrast, in Section 5, we provide several examples of restrictions on potential treatments that imply that our generalized monotonicity condition does not hold, and further show that some of these restrictions in fact have identifying power beyond instrument exogeneity for average potential outcomes and ATEs. Proofs of all results can be found in the Appendix.

2 Setup and Notation

Denote by $Y \in \mathcal{Y}$ a multi-valued outcome of interest, by $D \in \mathcal{D}$ a multi-valued endogenous regressor, and by $Z \in \mathcal{Z}$ a multi-valued instrumental variable.[2] To rule out degenerate cases, we assume throughout that $2 \leq |\mathcal{Y}| < \infty$, $2 \leq |\mathcal{D}| < \infty$, and $2 \leq |\mathcal{Z}| < \infty$. Further denote by $Y_d \in \mathcal{Y}$ the potential outcome if $D = d \in \mathcal{D}$ and by $D_z \in \mathcal{D}$ the potential treatment if $Z = z \in \mathcal{Z}$. We impose the usual consistency assumption,

$$Y = \sum_{d \in \mathcal{D}} Y_d \mathbb{1}\{D = d\} \quad \text{and} \quad D = \sum_{z \in \mathcal{Z}} D_z \mathbb{1}\{Z = z\} \ . \tag{1}$$

Let P denote the distribution of (Y, D, Z) and Q denote the distribution of $((Y_d : d \in \mathcal{D}), (D_z : z \in \mathcal{Z}), Z)$. Note that (1) defines a mapping T through

$$(Y, D, Z) = T((Y_d : d \in \mathcal{D}), (D_z : z \in \mathcal{Z}), Z) \ ,$$

[2] Our restriction to a multi-valued Y facilitates exposition, but is not essential. At the expense of slightly more complicated arguments, we can accommodate more generally any real-valued Y.

and therefore $P = QT^{-1}$. In what follows, we will say that a given Q rationalizes a given P if $P = QT^{-1}$.

Below we will require that $Q \in \mathbf{Q}$, where $\mathbf{Q}$ is a class of distributions satisfying assumptions that we will specify. Different choices of $\mathbf{Q}$ represent different assumptions that we impose on the distribution of potential outcomes and potential treatments. In this sense, $\mathbf{Q}$ may be viewed as a model for potential outcomes and potential treatments.

Given P and a model $\mathbf{Q}$, the set of $Q \in \mathbf{Q}$ that can rationalize P is

$$\mathbf{Q}_0(P, \mathbf{Q}) = \{Q \in \mathbf{Q} : P = QT^{-1}\} ,$$

i.e., the pre-image of P under T. We say $\mathbf{Q}$ is consistent with P if and only if $\mathbf{Q}_0(P, \mathbf{Q}) \neq \emptyset$. We will start by considering models $\mathbf{Q}$ for which every $Q \in \mathbf{Q}$ satisfies

Assumption 2.1 (Instrument Exogeneity)**.** $((Y_d : d \in \mathcal{D}), (D_z : z \in \mathcal{Z})) \perp\!\!\!\perp Z$ under Q.

Our final result on the identifying power of generalized monotonicity will also apply to the weaker exogeneity restriction in Richardson and Robins (2013) that avoids the "cross-world" restrictions of Assumption 2.1; see Assumption 3.3 and Corollary 3.3 below in Section 3.2. If Q satisfies Assumption 2.1, then

$$p_{yd|z} := P\{Y = y, D = d \mid Z = z\} = Q\{Y_d = y, D_z = d \mid Z = z\} = Q\{Y_d = y, D_z = d\} . \tag{2}$$

Since the marginal distribution of Z under P and Q are the same, i.e., for all $z \in \mathcal{Z}$,

$$P\{Z = z\} = Q\{Z = z\} ,$$

$P = QT^{-1}$ if and only if (2) holds. Thus, if all $Q \in \mathbf{Q}$ satisfies Assumption 2.1, then $\mathbf{Q}_0(P, \mathbf{Q})$ can be simplified as

$$\mathbf{Q}_0(P, \mathbf{Q}) = \left\{ Q \in \mathbf{Q} : \ p_{yd|z} = Q\{Y_d = y, D_z = d\} \text{ for all } y \in \mathcal{Y}, d \in \mathcal{D}, z \in \mathcal{Z} \right\} . \tag{3}$$

Let $\theta(Q) = (\mathbb{E}_Q[Y_d] : d \in \mathcal{D})$ denote the vector of average potential outcomes. For fixed P and $\mathbf{Q}$, the identified set for $\theta(Q)$ under P relative to $\mathbf{Q}$ is given by

$$\Theta_0(P, \mathbf{Q}) := \{\theta(Q) : Q \in \mathbf{Q}_0(P, \mathbf{Q})\} .$$

$\Theta_0(P, \mathbf{Q})$ is nonempty whenever $\mathbf{Q}_0(P, \mathbf{Q})$ is nonempty. By construction, this set is "sharp" in the sense that for any value in the set there exists $Q \in \mathbf{Q}_0(P, \mathbf{Q})$ for which $\theta(Q)$ equals the prescribed value. The identified set for $\theta(Q)$ immediately implies that the identified set for any parameter $\lambda = \lambda(\theta)$ is given by $\lambda(\Theta_0(P, \mathbf{Q}))$. An important example is $\mathbb{E}_Q[Y_j] - \mathbb{E}_Q[Y_k]$, the ATE for treatment j versus treatment k.

In the next section, we consider identification in any model of potential treatments that implies the following restriction on all $Q \in \mathbf{Q}$:

Assumption 2.2 (Generalized Monotonicity)**.** For each $d \in \mathcal{D}$, there exists $z^* = z^*(d, Q) \in \mathcal{Z}$ such that

$$Q\{D_{z^*} \neq d,\ D_{z'} = d \text{ for some } z' \neq z^*\} = 0\ . \tag{4}$$

In what follows, we refer to Assumption 2.2 as generalized monotonicity. It states that under Q, for each treatment status $d \in \mathcal{D}$, there exists a value (possibly depending on d and Q) of the instrument $z^* \in \mathcal{Z}$ that maximally encourages all individuals to d. Here, by "maximally encourage", we mean that if an individual does not choose d when $Z = z^*$, then they never choose d for any other value of Z. Equivalently, if an individual chooses d when Z is equal to any value other than z^*, then they have to choose d when $Z = z^*$. When $\mathcal{D} = \mathcal{Z} = \{0, 1\}$, Assumption 2.2 is equivalent to the monotonicity assumption of Imbens and Angrist (1994).

We emphasize that Assumption 2.2 only requires, for each possible value of the treatment, that there exists a value of the instrument that maximally encourages that treatment; it does not require that the value of the instrument is unique. For a given distribution Q and given treatment $d \in \mathcal{D}$, let $\mathcal{Z}^*(d, Q)$ denote the set of z^* that satisfy (4). In this notation, Assumption 2.2 can be restated as $\mathcal{Z}^*(d, Q) \neq \emptyset$ for each $d \in \mathcal{D}$. In the statement of Assumption 2.2, $z^*(d, Q)$ is allowed to change across Q. The following lemma shows that $\mathcal{Z}^*(d, Q)$ is identified from P and is hence the same for all Q that rationalizes P and satisfies Assumptions 2.1 and 2.2. In what follows, we will therefore write $\mathcal{Z}^*(d)$ and $z^*(d)$ whenever the given distribution Q rationalizes P and satisfies Assumptions 2.1 and 2.2. This result generalizes the corresponding result in Imbens and Angrist (1994).

Lemma 2.1. *Suppose Q satisfies Assumptions 2.1 and 2.2 and $P = QT^{-1}$. Then, $z \in \mathcal{Z}^*(d, Q)$ if and only if*

$$P\{D = d \mid Z = z\} \geq P\{D = d \mid Z = z'\} \text{ for all } z' \in \mathcal{Z}\ . \tag{5}$$

Below we prove the necessity of (5); sufficiency is established in the appendix. Note that, for any $d \in \mathcal{D}$, $z' \in \mathcal{Z}$, and any $z \in \mathcal{Z}^*(d, Q)$,

$$\begin{aligned} P\{D = d \mid Z = z\} &= Q\{D_z = d\} \\ &= Q\{D_z = d, D_{z'} = d\} + Q\{D_z = d, D_{z'} \neq d\} \\ &= Q\{D_{z'} = d\} + Q\{D_z = d, D_{z'} \neq d\} \\ &\geq Q\{D_{z'} = d\} \\ &= P\{D = d \mid Z = z'\}\ , \end{aligned}$$

where the first and last equalities exploit Assumption 2.1, and the third equality uses Assumption 2.2.

3 Main Result

In order to describe our main result, we first introduce some further notation. Denote by $\mathbf{Q}^*_E$ (where E stands for exogeneity) the set of all distributions that satisfy Assumption 2.1 and by $\mathbf{Q}^*_{E,M}$ (where M stands for generalized monotonicity) the set of all distributions that satisfy Assumptions 2.1 and 2.2. We will further require that the model does not restrict potential outcomes in the following sense:

Assumption 3.1 (Unrestricted Potential Outcomes)**.** Let $Q \in \mathbf{Q}$ and $Q' \in \mathbf{Q}^*_E$. If the distributions of $(D_z : z \in \mathcal{Z})$ under Q and Q' are the same, then $Q' \in \mathbf{Q}$.

In terms of this notation, our main result can be stated as follows:

Theorem 3.1. *Suppose* $\mathbf{Q} \subseteq \mathbf{Q}^*_{E,M}$ *and* $\mathbf{Q}$ *satisfies Assumption 3.1. Then, for any* P *such that* $\mathbf{Q}_0(P, \mathbf{Q}) \neq \emptyset$*, we have* $\Theta_0(P, \mathbf{Q}) = \Theta_0(P, \mathbf{Q}^*_{E,M}) = \Theta_0(P, \mathbf{Q}^*_E)$.

Theorem 3.1 describes the sense in which restrictions on potential treatments stronger than generalized monotonicity have no identifying power for average potential outcomes and ATEs provided that these stronger restrictions do not restrict potential outcomes. This result is established through Theorems 3.2 and 3.3 below. Theorem 3.2, developed in Section 3.1, characterizes $\Theta_0(P, \mathbf{Q})$ for any model $\mathbf{Q}$ that is stronger than Assumptions 2.1 and 2.2, i.e., $\mathbf{Q} \subseteq \mathbf{Q}^*_{E,M}$, and does not restrict potential outcomes in the sense of Assumption 3.1. The result shows, in particular, that $\Theta_0(P, \mathbf{Q}) = \Theta_0(P, \mathbf{Q}^*_{E,M})$ for any such model $\mathbf{Q}$ whenever $\mathbf{Q}_0(P, \mathbf{Q}) \neq \emptyset$. Remarkably, this result holds even if the model $\mathbf{Q}$ is *strictly* more restrictive than Assumptions 2.1 and 2.2 in the sense that $\mathbf{Q} \subsetneq \mathbf{Q}^*_{E,M}$. On the other hand, Theorem 3.3 and Corollary 3.2, developed in Section 3.2, show that $\Theta_0(P, \mathbf{Q}^*_{E,M}) = \Theta_0(P, \mathbf{Q}^*_E)$ whenever $\mathbf{Q}_0(P, \mathbf{Q}^*_{E,M}) \neq \emptyset$. Together, these results immediately imply Theorem 3.1. In fact, Theorem 3.3 shows the stronger result that if a submodel of instrument exogeneity and generalized monotonicity is consistent with P, then any model sandwiched between this submodel and the model that only assumes mean independence leads to the same identified set for average potential outcomes and ATEs. This observation allows us to establish that generalized monotonicity also has no identifying power for average potential outcomes and ATEs beyond the weaker exogeneity restriction of Richardson and Robins (2013).

3.1 Identified Sets for $\mathbf{Q} \subseteq \mathbf{Q}^*_{E,M}$

For $d \in \mathcal{D}$ and $z \in \mathcal{Z}$, define $\beta_{d|z} = \mathbb{E}_P[Y \mathbb{1}\{D = d\} \mid Z = z]$. In addition, define $y^L = \min(\mathcal{Y})$ and $y^U = \max(\mathcal{Y})$. The following theorem derives the identified set for $\theta(Q)$, relative to any model that assumes instrument exogeneity and generalized monotonicity but does not restrict potential outcomes. Note in particular that the assumptions allow for $\mathbf{Q} \subsetneq \mathbf{Q}^*_{E,M}$, in which case the model assumes *strictly* more than instrument exogeneity and generalized monotonicity.

Theorem 3.2. *Suppose* $\mathbf{Q} \subseteq \mathbf{Q}^*_{E,M}$ *and* $\mathbf{Q}$ *satisfies Assumption 3.1. Then, for any* P *such that* $\mathbf{Q}_0(P, \mathbf{Q}) \neq \emptyset$*,*

$$\Theta_0(P, \mathbf{Q}) = \prod_{d \in \mathcal{D}} \left[\beta_{d|z^*(d)} + y^L\Big(1 - \sum_{y \in \mathcal{Y}} p_{yd|z^*(d)}\Big),\ \beta_{d|z^*(d)} + y^U\Big(1 - \sum_{y \in \mathcal{Y}} p_{yd|z^*(d)}\Big) \right] . \tag{6}$$

We now describe some intuition for Theorem 3.2. Note that the distribution of the data, P, only contains information on the distribution of Y_d for those individuals who would take treatment d for some value of the instrument. On the other hand, it contains no information on the distribution of Y_d for those individuals who would not take that treatment for any value of the instrument. Assumption 2.2 implies that individuals would take treatment d at some value of the instrument if and only if $D_{z^*(d)} = d$, i.e., when maximally encouraged to do so. Assumption 2.1 implies $\beta_{d|z^*(d)} = \mathbb{E}_Q[\mathbb{1}\{D_{z^*(d)} = d\}Y_d]$, and thus captures all the information from P relevant to $\mathbb{E}_Q[Y_d]$, which is the first part of the lower and upper bounds in (6). In contrast, Assumptions 2.1 and 2.2 imply that the probability that an individual would not take treatment d for any value of Z is identified from P to be $Q\{D_{z^*(d)} \neq d\} = 1 - \sum_{y \in \mathcal{Y}} p_{yd|z^*(d)}$, but P contains no information on the distribution of Y_d for such individuals. Furthermore, that $\mathbf{Q}$ satisfies Assumption 3.1 implies that the model does not restrict the distribution of Y_d for such individuals beyond $y \in \mathcal{Y}$, so that we can set Y_d to be any value between y^L and y^U for these individuals, which constitutes the second part of the upper and lower bounds in (6).

Remark 3.1. Under the instrument exogeneity and monotonicity assumptions of Imbens and Angrist (1994), Balke and Pearl (1993, 1997) found the same form of the identified set for $\theta(Q)$ as (6) when $\mathcal{Y} = \mathcal{D} = \mathcal{Z} = \{0, 1\}$. Theorem 3.2 therefore generalizes the result of Balke and Pearl (1993, 1997) to more than two treatment arms and instrument values, to outcomes taking more than two values, and, more surprisingly, to show that the same identified set holds when imposing possibly *stronger* restrictions on potential treatments than generalized monotonicity. ■

Theorem 3.2 immediately implies the following result on the identified sets for the ATE of treatment j versus k:

Corollary 3.1. *Under the assumptions of Theorem 3.2, the identified set for $\mathbb{E}_Q[Y_j - Y_k]$ is given by:*

$$\Bigg[(\beta_{j|z^*(j)} - \beta_{k|z^*(k)}) + (y^L - y^U) + y^U \sum_{y \in \mathcal{Y}} p_{yk|z^*(k)} - y^L \sum_{y \in \mathcal{Y}} p_{yj|z^*(j)} \ , \\ (\beta_{j|z^*(j)} - \beta_{k|z^*(k)}) + (y^U - y^L) + y^L \sum_{y \in \mathcal{Y}} p_{yk|z^*(k)} - y^U \sum_{y \in \mathcal{Y}} p_{yj|z^*(j)}\Bigg] \ . \quad (7)$$

e

Remark 3.2. From Corollary 3.1, the width of the identified set for $\mathbb{E}_Q[Y_j - Y_k]$ under the assumptions of Theorem 3.2 is given by

$$(y^U - y^L)\left(P\{D \neq j \mid Z = z^*(j)\} + P\{D \neq k \mid Z = z^*(k)\}\right) \ ,$$

which, following Lemma 2.1, equals

$$(y^U - y^L)\left(\min_{z \in \mathcal{Z}} P\{D \neq j \mid Z = z\} + \min_{z \in \mathcal{Z}} P\{D \neq k \mid Z = z\}\right) \ .$$

Therefore, when imposing generalized monotonicity, as well as when imposing any restriction implying generalized monotonicity, the ATE of j versus k is *only* identified "at infinity" (Heckman, 1990; Andrews

and Schafgans, 1998) in the sense that identification requires

$$\min_{z\in\mathcal{Z}} P\{D \neq j \mid Z = z\} = \min_{z\in\mathcal{Z}} P\{D \neq k \mid Z = z\} = 0 \ . \tag{8}$$

In other words, identification of the the ATE of j versus k under generalized monotonicity or under any restriction implying generalized monotonicity requires that there is some value of the instrument such that everyone takes treatment j at that value of the instrument, and some value of the instrument such that everyone takes treatment k at that value of the instrument. In contrast, by imposing restrictions on potential treatments that imply that generalized monotonicity is violated, the ATE can sometimes be identified without (8) even when potential outcomes are unrestricted; see Example 5.1 in Section 5 below. ■

3.2 Identifying Power of Generalized Monotonicity

Theorem 3.2 above establishes that, for possibly multi-valued Y, D, and Z, and any model $\mathbf{Q} \subseteq \mathbf{Q}^*_{E,M}$ such that $\mathbf{Q}$ does not restrict the potential outcomes, if $\mathbf{Q}_0(P, \mathbf{Q}) \neq \emptyset$, then $\Theta_0(P, \mathbf{Q})$ equals (6). We now show that $\Theta_0(P, \mathbf{Q}^*_{E,M}) = \Theta_0(P, \mathbf{Q}^*_E)$ as long as $\mathbf{Q}_0(P, \mathbf{Q}^*_{E,M}) \neq \emptyset$, so that the identified set for $\theta(Q)$ assuming instrument exogeneity and generalized monotonicity coincides with the identified set assuming instrument exogeneity alone, as long as both assumptions are consistent with the distribution of the observed data. Our result therefore generalizes Balke and Pearl (1993, 1997), which study the case of binary Y, Z, and D.

In order to do so, we consider a mean independence assumption even weaker than Assumption 2.1, and show the identified set under this even weaker assumption is also (6). A sandwich argument will then lead to our desired result. In particular, we first establish that the identified set for $\theta(Q)$ in (6) coincides with the identified set under the weaker mean independence assumption considered by Robins (1989) and Manski (1990):

Assumption 3.2 (Mean Independence)**.** $\mathbb{E}_Q[Y_d \mid Z = z] = \mathbb{E}_Q[Y_d]$ for all $d \in \mathcal{D}$ and $z \in \mathcal{Z}$.

Note Assumption 3.2 is weaker than instrument exogeneity in Assumption 2.1, and does not imply (2). Let $\mathbf{Q}^*_{MI}$ denote the set of all Q that satisfies Assumption 3.2 (where MI stands for mean independence). Following Robins (1989) and Manski (1990), the following lemma derives the identified set for $\theta(Q)$ under mean independence:

Lemma 3.1. *Suppose* $\mathbf{Q}_0(P, \mathbf{Q}^*_{MI}) \neq \emptyset$. *Then,*

$$\Theta_0(P, \mathbf{Q}^*_{MI}) = \prod_{d\in\mathcal{D}} \left[\max_{z\in\mathcal{Z}} \{\beta_{d|z} + y^L(1 - \sum_{y\in\mathcal{Y}} p_{yd|z})\}, \ \min_{z\in\mathcal{Z}} \{\beta_{d|z} + y^U(1 - \sum_{y\in\mathcal{Y}} p_{yd|z})\} \right] . \tag{9}$$

The following lemma, which relies on the observation in Lemma 2.1, establishes the equivalence between the identified sets in (6) and (9) when $\mathbf{Q}_0(P, \mathbf{Q}^*_{E,M}) \neq \emptyset$.

Lemma 3.2. *Suppose* $\mathbf{Q} \subseteq \mathbf{Q}^*_{E,M}$, $\mathbf{Q}$ *satisfies Assumption 3.1, and* P *is such that* $\mathbf{Q}_0(P, \mathbf{Q}) \neq \emptyset$. *Then, the sets in* (6) *and* (9) *coincide.*

Using Lemma 3.2, we are able to establish our desired result, which asserts that, maintaining Assumption 2.1, additionally imposing Assumption 2.2 either causes the identified set for $\theta(Q)$ to become empty (if those assumptions are not consistent with P) or leaves the identified set for $\theta(Q)$ unchanged (if those assumptions are consistent with P). In fact, we will establish a stronger result, that if a submodel of instrument exogeneity and generalized monotonicity is consistent with P, then any model sandwiched between this submodel and the model that only assumes mean independence leads to the same identified set for $\theta(Q)$.

Theorem 3.3. *Suppose* $\mathbf{Q} \subseteq \mathbf{Q}^*_{E,M}$ *and* $\mathbf{Q}$ *satisfies Assumption 3.1. Further suppose* $\mathbf{Q}'$ *satisfies*

$$\mathbf{Q} \subseteq \mathbf{Q}' \subseteq \mathbf{Q}^*_{MI} \ .$$

Then, for any P *such that* $\mathbf{Q}_0(P, \mathbf{Q}) \neq \emptyset$,

$$\Theta_0(P, \mathbf{Q}) = \Theta_0(P, \mathbf{Q}') = \Theta_0(P, \mathbf{Q}^*_{MI}) \ .$$

Remark 3.3. Theorem 3.3 implies that in order for a model to have identifying power for average potential outcomes, it has to be the case that the model does not contain a submodel of instrument exogeneity and generalized monotonicity that is consistent with P. In other words, the model has to contradict Assumption 2.1 or 2.2. We illustrate this observation in Example 5.1 below. ■

Corollary 3.2. *For any* P *such that* $\mathbf{Q}_0(P, \mathbf{Q}^*_{E,M}) \neq \emptyset$, $\Theta_0(P, \mathbf{Q}^*_{E,M}) = \Theta_0(P, \mathbf{Q}^*_E)$.

Remark 3.4. An implication of Corollary 3.2 is that the identified set for $\theta(Q)$ under Assumption 2.1 alone will be (6), regardless of whether Assumption 2.2 is imposed, as long as the distribution of the data is consistent with Assumptions 2.1 and 2.2. ■

Theorem 3.3 further implies that under the following weaker exogeneity assumption, generalized monotonicity also has no identifying power for average potential outcomes and ATEs:

Assumption 3.3 (Weak Instrument Exogeneity)**.** Under Q, $(Y_d, D_z) \perp\!\!\!\perp Z$ for all $d \in \mathcal{D}, z \in \mathcal{Z}$.

See Richardson and Robins (2013) for an analysis of alternative exogeneity restrictions, and in particular how Assumption 3.3 avoids the "cross-world" restrictions of the stronger joint independence in Assumption 2.1. Denote by $\mathbf{Q}^*_{WE}$ the set of all distributions that satisfy Assumption 3.3 and $\mathbf{Q}^*_{WE,M}$ the set of all distributions that satisfy Assumptions 3.3 and 2.2.

Corollary 3.3. *For any* P *such that* $\mathbf{Q}_0(P, \mathbf{Q}^*_{WE,M}) \neq \emptyset$, $\Theta_0(P, \mathbf{Q}^*_{WE,M}) = \Theta_0(P, \mathbf{Q}^*_{WE})$.

4 Examples of Models That Satisfy Assumption 2.2

We now consider some restrictions on potential treatments that have been considered previously in the literature. In each case, we show $\mathbf{Q} \subseteq \mathbf{Q}^*_{E,M}$; in particular, these restrictions satisfy generalized monotonicity. We emphasize that frequently $\mathbf{Q} \subsetneq \mathbf{Q}^*_{E,M}$. In what follows, it is implicitly understood that Assumption 3.1 is

satisfied, so that the model imposes no restriction on potential outcomes. Thus, in each example Theorem 3.1 applies and such restrictions do not provide any identifying power for average potential outcomes or ATEs. In Appendix C, we consider three additional examples: an RCT with a "close substitute" as considered in Kline and Walters (2016); the monotonicity and "irrelevance" assumptions considered in Kirkeboen et al. (2016); and an additive random utility model for a binary treatment.

Example 4.1. Consider a multi-arm randomized controlled trial (RCT) with noncompliance, where $Z = d$ denotes random assignment to treatment d, $D_d = d$ denotes that the subject would comply with assignment if assigned to treatment d, and $|\mathcal{D}| = |\mathcal{Z}|$. More generally, not necessarily in the context of an RCT, one can interpret $Z = d$ as encouragement to treatment d and interpret $D_d = d$ as the subject would take treatment d if encouraged to do so. In this example, Q satisfies Assumption 2.1 because Z is randomly assigned. We may generalize the "no-defier" restriction of Angrist et al. (1996) as: for each $d \in \mathcal{D}$,

$$Q\{D_d \neq d,\ D_{d'} = d \text{ for some } d' \neq d\} = 0 \ ,$$

i.e., there is zero probability that a subject would not take treatment d if assigned to (encouraged to take) d but would take d if assigned (encouraged) to some other treatment $d' \neq d$. If Q satisfies this generalized no-defier restriction, then Assumption 2.2 holds with $z^*(d) = d$ for all d. This no-defier restriction in particular holds in the context of an RCT with "one-sided non-compliance," where we assume

$$Q\{D_z \in \{0, z\}\} = 1 \ ,$$

for all $z \in \mathcal{Z}$. Here, non-compliance is one-sided because one can fall back to the control group if assigned to d but cannot choose d if assigned to the control group. ■

Remark 4.1. As explained in Example 4.1, in a multi-arm RCT with one-sided noncompliance, P will necessarily be consistent with Assumptions 2.1 and 2.2, i.e., $\mathbf{Q}(P, \mathbf{Q}_{E,M}) \neq \emptyset$. Following Remark 3.4, Theorem 3.3 therefore implies that the identified set for $\theta(Q)$ under Assumption 2.1 alone will be (6) for such a multi-arm RCT. A further implication is that the identified set under Assumption 2.1 on $\mathbb{E}[Y_j - Y_k]$ for a multi-arm RCT with one-sided noncompliance depends only on the treatment arms for random assignment to treatments j and k. ■

Example 4.2. Cheng and Small (2006) consider an RCT with noncompliance where $\mathcal{D} = \mathcal{Z} = \{0, 1, 2\}$. Assumption 2.1 continues to hold because Z is randomly assigned. In such a setting, they develop bounds on average effects within subgroups defined by potential treatments which, following the terminology of Frangakis and Rubin (2002), they call "principal strata." While Bai et al. (2025a) derives the identified sets for those parameters given their assumptions, we now use our analysis to consider instead identification of average potential outcomes and ATEs given their assumptions. Their "Monotonicity I" assumption is equivalent to one-sided noncompliance in the preceding example. Their "Monotonicity II" assumption states that subjects who would comply with assignment to treatment 2 would also comply with assignment to treatment 1, so that

$$Q\{D_1 = 1 \mid D_2 = 2\} = 1 \ .$$

They argue that such an assumption is plausible in a medical context when treatment 1 has fewer side effects than treatment 2, and in their application to treatments for alcohol dependence, in which complying with treatment 1 (compliance enhancement therapy) requires less effort by subjects than complying with treatment 2 (cognitive behavioural therapy). Because their Monotonicity I restriction implies our Assumption 2.2, so does imposing both their Monotonicity I and II restrictions. ■

Example 4.3. Suppose Q satisfies Assumption 2.1. Heckman and Pinto (2018) define "unordered monotonicity" as the assumption that, for any $d \in \mathcal{D}$, and any $z, z' \in \mathcal{Z}$,

$$Q\{\mathbb{1}\{D_z = d\} \geq \mathbb{1}\{D_{z'} = d\}\} = 1 \text{ or } Q\{\mathbb{1}\{D_z = d\} \leq \mathbb{1}\{D_{z'} = d\}\} = 1 \ . \tag{10}$$

Assumption 2.2 holds for any Q that satisfies (10). To see this, note that Assumption 2.2 can be expressed as the requirement that for each $d \in \mathcal{D}$, there exists $z^*(d) \in \mathcal{Z}$ such that

$$Q\{\mathbb{1}\{D_{z^*(d)} = d\} \geq \mathbb{1}\{D_z = d\}\} = 1 \text{ for all } z \in \mathcal{Z} \ ,$$

which is immediately implied by (10). Note, however, that $\mathcal{Z}^*(d)$ may not be a singleton unless some inequalities in (10) are strict. ■

Remark 4.2. Although unordered monotonicity implies Assumption 2.2, the converse is generally false. For example, suppose $\mathcal{Z} = \{0, 1, 2, 3\}$ and $\mathcal{D} = \{0, 1\}$. Suppose $\mathbb{1}\{D_3 = 1\} \geq \mathbb{1}\{D_z = 1\}$ w.p.1 under Q for $z \neq 3$ and $\mathbb{1}\{D_0 = 0\} \geq \mathbb{1}\{D_z = 0\}$ w.p.1 under Q for $z \neq 0$, but $Q\{D_1 = 1, D_2 = 0\}$ and $Q\{D_1 = 0, D_2 = 1\}$ are both strictly positive. Then Assumption 2.2 holds with $z^*(0) = 0$ and $z^*(1) = 3$, but unordered monotonicity fails. In particular, $\mathbb{1}\{D_1 = 1\}$ and $\mathbb{1}\{D_2 = 1\}$ are not ordered, thus violating (10). We thus conclude that if $\mathbf{Q}$ is defined as the set of distributions that satisfy instrument exogeneity and unordered monotonicity, then $\mathbf{Q} \subsetneqq \mathbf{Q}^*_{E,M}$. ■

Example 4.4. Suppose under Q, $(D_z : z \in \mathcal{Z})$ is determined by

$$D_z = \underset{d \in \mathcal{D}}{\text{argmax}} \ (g(z, d) + U_d) \ , \tag{11}$$

for $g : \mathcal{Z} \times \mathcal{D} \to \Re$ where $\Re$ is the set of real numbers and a random vector $(U_d : d \in \mathcal{D})$, whose distribution is absolutely continuous with respect to the Lebesgue measure on $\Re^{|\mathcal{D}|}$ and $Z \perp\!\!\!\perp ((U_d : d \in \mathcal{D}), (Y_d : d \in \mathcal{D}))$. Hence, Q satisfies Assumption 2.1 by construction. Let $\mathbf{Q}$ denote the set of distributions that are consistent with $(D_z : z \in \mathcal{Z})$ being determined by (11) for some g and $(U_d : d \in \mathcal{D})$ satisfying these requirements. The model $\mathbf{Q}$ is called an additive random utility model (ARUM). A sufficient condition for $Q \in \mathbf{Q}$ to satisfy Assumption 2.2 is that for each $d \in \mathcal{D}$ there exists $z^*(d) \in \mathcal{Z}$ such that

$$g(z^*(d), d) - g(z^*(d), d') > g(z, d) - g(z, d') \text{ for all } d' \neq d \text{ and } z \neq z^*(d) \ . \tag{12}$$

We refer to the requirement in (12) as uniform targeting of treatment d. The terminology is intended to reflect that there is a value of the instrument that maximizes the gains (in terms of g) of choosing treatment d versus any other treatment d' uniformly across these other possible values of the treatment. In this sense,

that value of the instrument targets treatment d uniformly. We now argue by contradiction that (12) implies (4); hence, if (12) holds for all $d \in \mathcal{D}$, then Q satisfies Assumption 2.2. To this end, suppose that, with positive probability, $D_{z^*(d)} = d' \neq d$ but $D_{z'} = d$ for $z' \neq z^*(d)$. Then,

$$g(z^*(d), d') + U_{d'} \geq g(z^*(d), d) + U_d \ ,$$
$$g(z', d) + U_d \geq g(z', d') + U_{d'} \ .$$

These two inequalities imply

$$g(z^*(d), d) - g(z^*(d), d') \leq g(z', d) - g(z', d') \ ,$$

which violates (12). A particular example when $|\mathcal{Z}| \geq |\mathcal{D}|$ that satisfies (12) with $z^*(d) = d$ after a suitable relabelling is

$$g(z, d) = \alpha_d + \beta_d \mathbb{1}\{z = d\} \ , \tag{13}$$

with $\beta_d > 0$, so that $Z = d$ strictly increases the latent value of treatment d while leaving the values of the remaining options unchanged. ■

Remark 4.3. The ARUM for a binary treatment is equivalent to the Heckman-Vytlacil nonparametric selection model for a binary treatment considered, e.g., in Heckman and Vytlacil (1999, 2005), and shown by Vytlacil (2002) to be equivalent to the monotonicity and exogeneity assumptions of Imbens and Angrist (1994). Heckman and Vytlacil (2001) show that the Heckman-Vytlacil nonparametric selection model for a binary treatment results in an identified set for the ATE of the form in (7) and that the model has no identifying power beyond instrument exogeneity for the ATE. Example C.1 in Appendix C shows that an ARUM for a binary treatment will satisfy Assumption 2.2 and therefore the results of this paper nest the results of Heckman and Vytlacil (2001). Example 4.4, on the other hand, extends their results to nonparametric selection models for a multi-valued treatment. For a partial identification analysis of a class of parameters that includes the ATE under a nonparametric selection model for a binary treatment and sometimes imposing additional restrictions, see, e.g., Mogstad et al. (2018), Han and Yang (2024), and Marx (2024). ■

Example 4.5. Lee and Salanié (2023) also consider the ARUM defined by (11) without imposing the uniform targeting of (12) for each treatment. Instead, they impose an assumption that they refer to as "strict one-to-one targeting", in which the set of treatments can be partitioned into a set of treatments $\mathcal{D}^\dagger$ that are "targeted" and a set of treatments $\mathcal{D} \setminus \mathcal{D}^\dagger$ that are "not targeted" such that

1. For $d \in \mathcal{D} \setminus \mathcal{D}^\dagger$, $g(z, d)$ is the same for all $z \in \mathcal{Z}$;
2. For $d \in \mathcal{D}^\dagger$, there exists $z^\dagger(d)$ such that $g(z^\dagger(d), d) > g(z', d)$ for all $z' \neq z^\dagger(d)$ and such that $g(z', d)$ takes the same value for all $z' \neq z^\dagger(d)$; additionally, $z^\dagger(d) \neq z^\dagger(d')$ for $d, d' \in \mathcal{D}^\dagger$, $d \neq d'$.

The terminology "one-to-one" stems from the second requirement above. They further impose that there exists a treatment that is known to be non-targeted. This class of ARUMs is equivalent to imposing (13) for

targeted treatments, imposing $g(z,d) = \alpha_d$ for non-targeted treatments, and imposing that there is at least one non-targeted treatment. In such a setting, Lee and Salanié (2023) analyze the identification of a particular class of average effects within subgroups defined by potential treatments. We now use our analysis to consider instead the identification of average potential outcomes and ATEs given their assumptions. Suppose strict one-to-one targeting holds, and additionally suppose that there is at least one targeted treatment. In Appendix B.1, we argue that Assumption 2.2 holds when $|\mathcal{Z}| > |\mathcal{D}^\dagger|$, so that there are more values of the instrument than there are targeted treatments. We further argue that (12) does not hold for some treatments unless $|\mathcal{D}| = |\mathcal{Z}| = 2$. Such models therefore provide another class of ARUMs, distinct from the one with uniform targeting described in Example 4.4, for which Assumption 2.2 holds. In Example 5.3 below, we show, however, that Assumption 2.2 does not hold when $|\mathcal{D}| \geq 3$, $|\mathcal{Z}| = |\mathcal{D}^\dagger|$, and the support of $(U_d : d \in \mathcal{D})$ is $\Re^{|\mathcal{D}|}$. ■

5 Examples of Models That Do Not Satisfy Assumption 2.2

We now consider models that do not satisfy generalized monotonicity (Assumption 2.2). For each model, we show that the identified sets for average potential outcomes are not given by (6). For the first two examples, we further show that they do in fact provide identifying power beyond instrument exogeneity.

Example 5.1. Suppose $\mathcal{Y} = \mathcal{D} = \mathcal{Z} = \{0,1\}$. Let $\mathbf{Q}$ denote all distributions Q that satisfy Assumption 2.1 and

$$Q\{D_0 = D_1\} = 0 \ ,$$

which, in the language of Angrist et al. (1996), is imposing that all individuals are either compliers or defiers. For any P such that $\mathbf{Q}_0(P,\mathbf{Q}) \neq \emptyset$, $\theta(Q)$ is identified relative to $\mathbf{Q}$, i.e., $\Theta_0(P,\mathbf{Q})$ is a singleton. To see this, note that for any $Q \in \mathbf{Q}_0(P,\mathbf{Q})$ and $y \in \mathcal{Y}$,

$$\begin{aligned} Q\{Y_1 = y\} &= Q\{Y_1 = y, D_0 = 0, D_1 = 1\} + Q\{Y_1 = y, D_0 = 1, D_1 = 0\} \\ &= Q\{Y_1 = y, D_1 = 1\} + Q\{Y_1 = y, D_0 = 1\} \\ &= Q\{Y_1 = y, D_1 = 1 \mid Z = 1\} + Q\{Y_1 = y, D_0 = 1 \mid Z = 0\} \\ &= P\{Y = y, D = 1 \mid Z = 1\} + P\{Y = y, D = 1 \mid Z = 0\} \ , \end{aligned}$$

where the first two equalities follows from $Q\{D_0 = D_1\} = 0$, the third equality follows from Assumption 2.1, and the final equality follows from $Q \in \mathbf{Q}_0(P,\mathbf{Q})$. A similar argument establishes identification of $Q\{Y_0 = y\}$. In contrast, we show in Appendix B.2 that there exists a P for which $\mathbf{Q}_0(P,\mathbf{Q}) \neq \emptyset$ and (6) is not a singleton. The identified set for $\theta(Q)$ is therefore not given by (6). We further show in Appendix B.2 that $\Theta_0(P,\mathbf{Q}^*_E)$ is not a singleton for the same P. Thus, the model $\mathbf{Q}$ does have identifying power beyond instrument exogeneity for $\theta(Q)$.

Furthermore, recall as discussed in Remark 3.3 that if $\mathbf{Q}$ has identifying power for $\theta(Q)$, then it cannot contain a submodel satisfying instrument exogeneity and generalized monotonicity that is consistent with P. In this example, the model $\mathbf{Q}$ contains such a submodel if and only if P satisfies either (i) $P\{D = 0 \mid$

$Z = 0\} = P\{D = 1 \mid Z = 1\} = 1$ (in which case all individuals are compliers) or (ii) $P\{D = 1 \mid Z = 0\} = P\{D = 0 \mid Z = 1\} = 1$ (in which case all individuals are defiers). In order for $\mathbf{Q}$ to have identifying power for $\theta(Q)$, it therefore must be the case that $0 < P\{D = 0 \mid Z = 0\}, P\{D = 0 \mid Z = 1\} < 1$, which is indeed satisfied by the counterexample in Appendix B.2. Further note that (6) is a singleton in this example if and only if P satisfies either (i) or (ii). ■

Example 5.2. Consider an ordered choice model for treatment. Suppose that $|\mathcal{Z}| \geq 3$, and let $\mathbf{Q}$ denote the set of all distributions that satisfy Assumption 2.1 and

$$Q\{D_j \geq D_k\} = 1 \text{ for all } j \geq k \,. \tag{14}$$

For example, D might represent quantity of some treatment, and Z might represent levels of subsidy for the treatment. The restriction in (14) is equivalent to the monotonicity assumption considered in Angrist and Imbens (1995). See Vytlacil (2006) for the connection between this restriction and ordered discrete-choice selection models. Without loss of generality, let $\mathcal{D} = \{0, \ldots, \bar{D}\}$. In this case, (4) is satisfied for $d \in \{0, \bar{D}\}$ for all $Q \in \mathbf{Q}_0(P, \mathbf{Q})$; to see this, take $z^*(0) = \min\{\mathcal{Z}\}$ and $z^*(\bar{D}) = \max\{\mathcal{Z}\}$. By a straightforward modification of the arguments underlying Theorem 3.2, one can show that the identified sets for $\mathbb{E}_Q[Y_0]$ and $\mathbb{E}_Q[Y_{\bar{D}}]$ are given by (6) for any P such that $\mathbf{Q}_0(P, \mathbf{Q}) \neq \emptyset$. The ordered monotonicity assumption in (14) therefore has no identifying power beyond instrument exogeneity for $\mathbb{E}_Q[Y_0]$ and $\mathbb{E}_Q[Y_{\bar{D}}]$. In contrast, (4) need not hold for $d \in \mathcal{D} \setminus \{0, \bar{D}\}$ and $Q \in \mathbf{Q}_0(P, \mathbf{Q})$. In Appendix B.3, we show there exists a P for which $\mathbf{Q}_0(P, \mathbf{Q}) \neq \emptyset$ and $\Theta_0(P, \mathbf{Q})$ is not given by (6). We further show $\Theta_0(P, \mathbf{Q}) \subsetneqq \Theta_0(P, \mathbf{Q}^*_E)$ for the same P. Thus, the ordered monotonicity assumption in (14) does have identifying power beyond instrument exogeneity for $\mathbb{E}_Q[Y_d]$ for $d \in \mathcal{D} \setminus \{0, \bar{D}\}$. ■

Example 5.3. In Example 4.5, we considered ARUMs satisfying the strict one-to-one targeting assumption of Lee and Salanié (2023). As discussed there, if $|\mathcal{D}| = 2$ or if $|\mathcal{D}| \geq 3$ and $|\mathcal{Z}| > |\mathcal{D}^\dagger|$, then Assumption 2.2 holds and the results in Section 3 are applicable. Now consider the case in which $|\mathcal{D}| \geq 3$ and $|\mathcal{Z}| = |\mathcal{D}^\dagger|$. Denote by $\mathbf{Q}$ the ARUM model defined by (11) under the additional assumption that the support of $(U_d : d \in \mathcal{D})$ is $\Re^{|\mathcal{D}|}$. Then, as we show in Appendix B.4, while (4) will hold for targeted treatments, (4) cannot hold for any non-targeted treatment, and thus Assumption 2.2 is violated. We show that, while the identified set for $\mathbb{E}_Q[Y_d]$ is given by (6) for targeted treatments when $\mathbf{Q}_0(P, \mathbf{Q}) \neq \emptyset$, there exists a P for which $\mathbf{Q}_0(P, \mathbf{Q}) \neq \emptyset$ and the identified set for $\mathbb{E}_Q[Y_d]$ is not given by (6) for non-targeted treatments. ■

Disclosure Statement

The authors report there are no competing interests to declare.

A Proofs of Main Results

A.1 Proof of Lemma 2.1

The necessity of (5) has been proved in the main text. On the other hand, fix a $d \in \mathcal{D}$, suppose (5) holds for some $z \in \mathcal{Z}$. Fix a particular $z^* \in \mathcal{Z}^*(d, Q)$ that satisfies (4). If $z = z^*$ then of course Assumption 2.2 holds for z. Suppose $z \neq z^*$. Then,

$$\begin{aligned} 0 &\leq P\{D = d \mid Z = z\} - P\{D = d \mid Z = z^*\} \\ &= Q\{D_z = d\} - Q\{D_{z^*} = d\} \\ &= Q\{D_{z^*} \neq d, D_z = d\} - Q\{D_z \neq d, D_{z^*} = d\} \\ &= -Q\{D_z \neq d, D_{z^*} = d\} \ , \end{aligned}$$

where the first inequality is using that z satisfies (5), the second line is using Assumption 2.1, and the last line is using that z^* satisfies (4). Thus $Q\{D_z \neq d, D_{z^*} = d\} = 0$. Assumption 2.2 implies

$$\begin{aligned} &Q\{D_z \neq d, D_{z'} = d \text{ for some } z' \in \mathcal{Z}\} \\ &= Q\{D_z \neq d, D_{z^*} = d\} + Q\{D_z \neq d, D_{z^*} \neq d, D_{z'} = d \text{ for some } z' \in \mathcal{Z} \setminus \{z^*\}\} \\ &= 0 \ , \end{aligned}$$

and z satisfies Assumption 2.2 as well. ■

A.2 Proof of Theorem 3.1

The desired result follows immediately from Theorems 3.2 and 3.3. ■

A.3 Proof of Theorem 3.2

This section is organized as follows. In Section A.3.1, we introduce additional notation that is helpful in formally proving our result, including defining subgroups of individuals, called treatment response types, who are defined by what treatment they would take at each value of the instrument. Because all variables are discrete, we will directly work with the probability mass function. We derive a lemma that characterizes a sufficient condition for a given distribution of potential outcomes and treatments Q to rationalize the distribution of the data P. The lemma states that whether a Q that satisfies Assumption 2.1 rationalizes P depends on the probability of each treatment response type and on the probability of that type's potential outcomes corresponding to treatments they would choose for some value of the instrument (so that they "comply with" this treatment at least for some values of the instrument), but does not depend on the probability of that type's potential outcomes corresponding to treatments they would not choose for any value of the instrument (so that they are "never-takers" of this treatment). If Q satisfies Assumptions 2.1 and 2.2, then the set of treatments for which a treatment response type is a "never-taker" are precisely

the set of treatments that they would not take even when maximally encouraged to do so. Therefore, the implication of the lemma is that if Q satisfies Assumptions 2.1 and 2.2 and rationalizes P, then any other Q^* satisfying Assumption 2.1 and 2.2 will also rationalize P if Q and Q^* differ *only* in the probability of potential outcomes corresponding to treatments that a given response type would not take even when maximally encouraged to do so.

In Section A.3.2, we use the notation and lemma introduced in Section A.3.1 to prove the theorem. Let $\mathbf{Q}$ satisfy the assumptions of the theorem. We first show that $\Theta_0(P, \mathbf{Q})$ is a subset of the bounds in (6). We then show that the bounds in (6) are a subset of $\Theta_0(P, \mathbf{Q})$ using the following proof strategy. By assumption, $\Theta_0(P, \mathbf{Q})$ is non-empty, so that there exists a distribution $Q \in \mathbf{Q}$ that rationalizes P. For each value θ_0 in the bounds of (6), we construct an alternative distribution $Q^* \in \mathbf{Q}$ such that $\theta(Q^*) = \theta_0$ with Q and Q^* differing only in the probability of outcomes corresponding to treatments that a given response type would not take even when maximally encouraged to do so. That the constructed Q^* lies in $\mathbf{Q}$ follows from the assumption that $\mathbf{Q}$ satisfies Assumption 3.1 and that Q and Q^* have the same distribution of potential treatment choices with $Q \in \mathbf{Q}$. That the constructed Q^* rationalizes P follows from $Q \in \mathbf{Q}$ and the previously described lemma. That we are able to construct such a Q^* with $\theta(Q^*) = \theta_0$ for every θ_0 in the bounds of (6) establishes that the bounds (6) are a subset of $\Theta_0(P, \mathbf{Q})$, completing the proof.

A.3.1 Auxillary Results

To present the proof of Theorem 3.2, we first introduce some further notation. Because all variables are discrete, we will directly work with the probability mass function. Recall from the discussion in Section 2 that if Q satisfies Assumption 2.1, then $P = QT^{-1}$ if and only if

$$p_{yd|z} = Q\{Y_d = y, D_z = d\} \ .$$

Following Heckman and Pinto (2018), we define a treatment response type as a vector $r^t \in \mathcal{D}^{|\mathcal{Z}|}$,

$$r^t = (d_z : z \in \mathcal{Z}) \in \mathcal{D}^{|\mathcal{Z}|} \ .$$

Treatment response types are also called principal strata (Frangakis and Rubin, 2002). We analogously define an outcome response type as a vector $r^o \in \mathcal{Y}^{|\mathcal{D}|}$,

$$r^o = (y_d : d \in \mathcal{D}) \in \mathcal{Y}^{|\mathcal{D}|} \ .$$

Because all variables are discrete, we define the probability mass function as

$$q\big(r^o, r^t\big) = Q\{(Y_d : d \in \mathcal{D}) = r^o, (D_z : z \in \mathcal{Z}) = r^t\} \ .$$

For the rest of the proof, without loss of generality, we suppose $\mathcal{D} = \{0, 1, \ldots, |\mathcal{D}|-1\}$ and $\mathcal{Z} = \{0, 1, \ldots, |\mathcal{Z}|-1\}$. Let r^o_j denote the $(j+1)$th entry of r^o and r^t_j denote the $(j+1)$th entry of r^t. In other words, r^o_j denotes the value of the potential outcome for the outcome response type r^o when the treatment equals j, and r^t_j

denotes the value of the potential treatment for the treatment response type r^t when the instrument equals j. In this notation, if Q satisfies Assumption 2.1, then it follows from (2) that $P = QT^{-1}$ if and only if

$$p_{yd|z} = \sum_{(r^o,r^t):r^o_d=y,r^t_z=d} q(r^o, r^t) \qquad \forall\ y \in \mathcal{Y},\ d \in \mathcal{D},\ z \in \mathcal{Z}\ . \tag{15}$$

Below we derive a lemma that simplifies determining whether $q(r^o, r^t)$ satisfies (15) and will be used subsequently to derive our characterization of the identified set. To this end, we require some further notation. Let

$$\mathcal{N}(r^t) = \{d \in \mathcal{D}\ :\ r^t_z \neq d \text{ for all } z \in \mathcal{Z}\}\ ,$$
$$\mathcal{N}(r^t)^c = \{d \in \mathcal{D}\ :\ r^t_z = d \text{ for some } z \in \mathcal{Z}\}\ ,$$

For a given treatment response type r^t, $\mathcal{N}(r^t)$ is the set of treatments for which that treatment response type is a "never-taker," and $\mathcal{N}(r^t)^c$ is the set of treatments for which that treatment response type will "comply with" the treatment for some value of z. Using this notation, partition outcome and treatment response types (r^o, r^t) as $(r^o_n(r^t), r^o_c(r^t), r^t)$ where

$$r^o_n(r^t) = (r^o_d : d \in \mathcal{N}(r^t))\ ,$$
$$r^o_c(r^t) = (r^o_d : d \in \mathcal{N}(r^t)^c)\ .$$

For a given treatment response type r^t, $r^o_n(r^t)$ are those outcomes that are never observed for that response type, and $r^o_c(r^t)$ are the remaining outcomes that are observed given some value of Z. Here, the subscripts n and c stand for "never-taker" and "complier."

Remark A.1. Here we illustrate how our notation specializes under Assumption 2.2. Note Assumption 2.2 can be expressed as restricting $q(r^o, r^t) = 0$ unless the treatment response type r^t satisfies the condition therein; in other words, it restricts the support of the treatment response type. In particular, if for some $d \in \mathcal{D}$, $r^t_{z^*(d)} \neq d$ while $r^t_{z'} = d$ for some $z' \neq z^*(d)$, then $q(r^o, r^t) = 0$ for all r^o. For any r^t in the support,

$$\mathcal{N}(r^t) = \{d \in \mathcal{D}\ :\ r^t_{z^*(d)} \neq d\}\ ,$$
$$\mathcal{N}(r^t)^c = \{d \in \mathcal{D}\ :\ r^t_{z^*(d)} = d\}\ ,$$

and

$$r^o_n(r^t) = (r^o_d : d \in \mathcal{D},\ r^t_{z^*(d)} \neq d)\ ,$$
$$r^o_c(r^t) = (r^o_d : d \in \mathcal{D},\ r^t_{z^*(d)} = d)\ .$$

Indeed, $z^*(d)$ is the instrument that maximally encourages to treatment d, so if $r^t_{z^*(d)} \neq d$, then $r^t_z \neq d$ for all $z \in \mathcal{Z}$. As a result, someone with that treatment response type r^t never takes d, and hence $d \in \mathcal{N}(r^t)$. Otherwise, $d \in \mathcal{N}(r^t)^c$, or this person is a "complier" for treatment d at least when $z = z^*(d)$. The outcome

response type r^o is then partitioned into $r_n^o(r^t)$ and $r_c^o(r^t)$ according to whether $d \in \mathcal{N}(r^t)$ or not. ■

For notational convenience, we further define the probability mass $q(r_c^o(r^t), r^t)$ as $q(r_n^o(r^t), r_c^o(r^t), r^t)$ summed over $r_n^o(r^t)$:

$$q(r_c^o(r^t), r^t) = \begin{cases} q(r^o, r^t) & \text{if } \mathcal{N}(r^t) = \emptyset \text{ so that } r_c^o(r^t) = r^o \\ \sum_{r_n^o(r^t) \in \mathcal{Y}^{|\mathcal{N}(r^t)|}} q(r_n^o(r^t), r_c^o(r^t), r^t) & \text{if } \mathcal{N}(r^t) \neq \emptyset \text{ so that } r_c^o(r^t) \neq r^o \end{cases} .$$

In defining the probability mass $q(r_c^o(r^t), r^t)$, we sum over all possible values of $r_n^o(r^t)$, because these are the outcomes of treatments that are never taken by the treatment response type r^t, and hence will not be relevant for the observed data. Using this notation, we have the following lemma that asserts whether $q(r^o, r^t)$ satisfies (15) depends only on $q(r_c^o(r^t), r^t)$. This lemma implies that whether a distribution of potential outcomes and treatments Q that satisfies Assumption 2.1 rationalizes the distribution of the data P depends only on the probability of each treatment response type and the probability of that type's potential outcomes that would be observed for some value of the instrument.

Lemma A.1. *Suppose q satisfies (15). Then, q^* satisfies (15) if, for each $r^t \in \mathcal{D}^{|\mathcal{Z}|}$,*

$$q^*(r_c^o(r^t), r^t) = q(r_c^o(r^t), r^t) \quad \forall\ r_c^o(r^t) \ . \tag{16}$$

Proof. We can rewrite (15) as

$$\begin{aligned} p_{yd|z} &= \sum_{r^t : r_z^t = d} \ \sum_{r^o : r_d^o = y} q(r^o, r^t) \\ &= \sum_{r^t : r_z^t = d} \ \sum_{r_c^o(r^t) : r_d^o = y} \left(\sum_{r_n^o(r^t)} q(r_n^o(r^t), r_c^o(r_t), r^t) \right) \\ &= \sum_{r^t : r_z^t = d} \ \sum_{r_c^o(r^t) : r_d^o = y} q(r_c^o(r_t), r^t) \ , \end{aligned}$$

where the second equality uses that $r_c^o(r^t)$ is nonempty because $r_z^t = d$ and that r_d^o is an element of $r_c^o(r^t)$ for r^t such that $r_z^t = d$. The result now follows. ■

A.3.2 Proof of the Theorem

$\underline{\Theta_0(P, \mathbf{Q}) \subseteq (6)}$

We first show that (6) provides valid bounds on $\theta(Q)$ under the stated assumptions, that is, $\Theta_0(P, \mathbf{Q})$ is a subset of the bounds of (6). Suppose $Q \in \mathbf{Q}_0(P, \mathbf{Q})$. For each $d \in \mathcal{D}$,

$$\begin{aligned} \mathbb{E}_Q[Y_d] &= \mathbb{E}_Q[Y_d \mathbb{1}\{D_{z^*(d)} = d\}] + \mathbb{E}_Q[Y_d \mathbb{1}\{D_{z^*(d)} \neq d\}] \\ &= \beta_{d|z^*(d)} + \mathbb{E}_Q[Y_d \mathbb{1}\{D_{z^*(d)} \neq d\}] \ , \end{aligned}$$

where the second equality is using Assumption 2.1. We have

$$\mathbb{E}[Y_d \mathbb{1}\{D_{z^*(d)} \neq d\}] \in [y^L Q\{D_{z^*(d)} \neq d\},\ y^U Q\{D_{z^*(d)} \neq d\}] \ ,$$

while Assumption 2.1 implies that

$$Q\{D_{z^*(d)} \neq d\} = 1 - \sum_{y \in \mathcal{Y}} p_{yd|z^*(d)} \ .$$

We thus have that

$$\mathbb{E}[Y_d] \in [\beta_{d|z^*(d)} + y^L(1 - \sum_{y \in \mathcal{Y}} p_{yd|z^*(d)}),\ \beta_{d|z^*(d)} + y^U(1 - \sum_{y \in \mathcal{Y}} p_{yd|z^*(d)})] \ ,$$

for each $d \in \mathcal{D}$, and thus (6) provides valid bounds on $\theta(Q)$ under the stated assumptions.

(6) $\subseteq \Theta_0(P, \mathbf{Q})$

We now show that the bounds of (6) are the identified set for $\theta(Q)$, that is, the bounds of (6) are a subset of $\Theta_0(P, \mathbf{Q})$. Let q denote latent probabilities corresponding to a fixed $Q \in \mathbf{Q}_0(P, \mathbf{Q})$. There exists such a q by the assumption that $\mathbf{Q}_0(P, \mathbf{Q})$ is non-empty. We show that for each θ_0 in the right-hand side of (6), we can construct an alternative distribution of potential outcomes and treatments Q^* that is contained in $\mathbf{Q}_0(P, \mathbf{Q})$ and for which $\theta(Q^*) = (\mathbb{E}_{Q^*}[Y_j] : j \in \mathcal{D})$ is equal to θ_0. In particular, for each θ_0 in the right-hand side of (6) we will construct q^* corresponding to Q^* that

(a) satisfies $\sum_{r^o} q^*(r^o, r^t) = \sum_{r^o} q(r^o, r^t)$ and hence $Q^* \in \mathbf{Q}$ because the distribution of $(D_z : z \in \mathcal{Z})$ is unchanged and by assumption that $\mathbf{Q}$ satisfies Assumption 3.1,

(b) satisfies (16) and hence $P = Q^* T^{-1}$ due to Lemma A.1, and

(c) satisfies $\theta(Q^*) = \theta_0$.

Properties (a) and (b) allow us to conclude that $Q^* \in \mathbf{Q}_0(P, \mathbf{Q})$, that is, the constructed distribution is consistent with P and the model $\mathbf{Q}$. These properties will follow from our iterative construction of q^*, which preserves the marginal distribution of potential treatments but modifies the marginal distributions of potential outcomes for outcomes that are never observed for a given treatment response type, that is, correspond to a never-taken treatment for a given treatment response type. Because the marginal distribution of potential treatments is preserved, property (a) follows. Because only the marginal distributions of potential outcomes for never-taken treatments are modified, property (b) follows. Property (c) follows from being able to flexibly modify the marginal distributions of potential outcomes for never-taken treatments, so that any θ can be achieved.

Part 1: construct Q^*

We now construct an alternative q^* as follows. Fix some vector of weights $\alpha = (\alpha_0, \alpha_1, ..., \alpha_{|\mathcal{D}|-1})' \in [0,1]^{|\mathcal{D}|}$ to be specified below. For each treatment response type r^t, let $q_0^*(r^o, r^t) = q(r^o, r^t)$ for all r^o.

Let $K(r^t) = |\mathcal{N}(r^t)|$ be the number of treatments for which treatment response type r^t is a never-taker. Note that if $K(r^t) = 0$, then $\mathcal{N}(r^t) = \emptyset$, so $r^t_{z^*(d)} = d$ for all $d \in \mathcal{D}$ under r^t. For such an r^t, we set $q^*(r^o, r^t) = q^*_0(r^o, r^t) = q(r^o, r^t)$ for all $r^o \in \mathcal{Y}^{|\mathcal{D}|}$.

If $K(r^t) \geq 1$, enumerate the set of never-taken treatments $\mathcal{N}(r^t)$ as $\{j[1], ..., j[K(r^t)]\}$, and for $k = 1$ to $K(r^t)$, define q^*_k iteratively as follows:

$$\begin{aligned} q^*_k((r^o_{-j[k]}, r^o_{j[k]} = y^L), r^t) &= (1 - \alpha_{j[k]}) \sum_{r^o_{j[k]} \in \mathcal{Y}} q^*_{k-1}((r^o_{-j[k]}, r^o_{j[k]}), r^t) \\ q^*_k((r^o_{-j[k]}, r^o_{j[k]} = y), r^t) &= 0 \qquad \text{for } y \notin \{y^L, y^U\} \\ q^*_k((r^o_{-j[k]}, r^o_{j[k]} = y^U), r^t) &= \alpha_{j[k]} \sum_{r^o_{j[k]} \in \mathcal{Y}} q^*_{k-1}((r^o_{-j[k]}, r^o_{j[k]}), r^t) \ , \end{aligned} \tag{17}$$

for all $r^o_{-j[k]}$, where we partition $r^o = (r^o_{-j[k]}, r^o_{j[k]})$. Intuitively, in step k, for each $r^o_{-j[k]}$ and r^t we reassign the probabilities of all outcome responses to never-taken treatment $j[k]$ to outcome responses y^L and y^U, splitting between y^L and y^U according to weight $\alpha_{j[k]}$.

With this construction, the marginal distribution of $Y_{j[k]}$ for treatment response type r^t is only modified in step k. This statement implies that for each fixed k, for step $\ell \leq k - 1$, and for any outcome $y \in \mathcal{Y}$,

$$\sum_{r^o_{-j[k]}} q^*_\ell((r^o_{-j[k]}, r^o_{j[k]} = y), r^t) = \sum_{r^o_{-j[k]}} q^*_0((r^o_{-j[k]}, r^o_{j[k]} = y), r^t) = \sum_{r^o_{-j[k]}} q((r^o_{-j[k]}, r^o_{j[k]} = y), r^t) \ . \tag{18}$$

On the other hand, for each fixed k, because the marginal distribution of $Y_{j[k]}$ for treatment response type r^t is not further modified after step k, (17) and (18) imply

$$\begin{aligned} \sum_{r^o_{-j[k]}} q^*_{K(r^t)}((r^o_{-j[k]}, r^o_{j[k]} = y^L), r^t) &= \sum_{r^o_{-j[k]}} q^*_k((r^o_{-j[k]}, r^o_{j[k]} = y^L), r^t) \\ &= (1 - \alpha_{j[k]}) \sum_{r^o_{j[k]} \in \mathcal{Y}} \sum_{r^o_{-j[k]}} q((r^o_{-j[k]}, r^o_{j[k]}), r^t) \\ &= (1 - \alpha_{j[k]}) \sum_{r^o} q(r^o, r^t) \ , \\ \sum_{r^o_{-j[k]}} q^*_{K(r^t)}((r^o_{-j[k]}, r^o_{j[k]} = y), r^t) &= 0 \qquad \text{for } y \notin \{y^L, y^U\} \ , \\ \sum_{r^o_{-j[k]}} q^*_{K(r^t)}((r^o_{-j[k]}, r^o_{j[k]} = y^U), r^t) &= \sum_{r^o_{-j[k]}} q^*_k((r^o_{-j[k]}, r^o_{j[k]} = y^U), r^t) \\ &= \alpha_{j[k]} \sum_{r^o_{j[k]} \in \mathcal{Y}} \sum_{r^o_{-j[k]}} q((r^o_{-j[k]}, r^o_{j[k]}), r^t) \\ &= \alpha_{j[k]} \sum_{r^o} q(r^o, r^t) \ . \end{aligned} \tag{19}$$

These equations state that under $q^*_{K(r^t)}$, the constructed distribution in the final step $K(r^t)$, the probability that $r^o_{j[k]} = y$ for each r^t is zero if y is not y^L or y^U, is $\alpha_{j[k]}$ times the true probability of r^t under q if $y = y^U$, and is $1 - \alpha_{j[k]}$ times the true probability of r^t under q if $y = y^L$.

Finally, set

$$q^*(r^o, r^t) = q^*_{K(r^t)}(r^o, r^t) \qquad \forall\ r^o\ .$$

Part 2: verify property (a), $Q^* \in \mathbf{Q}$

With this construction, $q^*(r^o, r^t)$ is non-negative and from (19) we have

$$\sum_{r^o} q^*(r^o, r^t) = \sum_{r^o} q(r^o, r^t) \quad \text{for all} \quad r^t\ .$$

Because we assume $\mathbf{Q}$ satisfies Assumption 3.1 and $Q \in \mathbf{Q}$, this implies that $Q^* \in \mathbf{Q}$.

Part 3: verify property (b), $P = Q^* T^{-1}$

Furthermore, for each r^t and for all $r^o_c(r^t)$, the construction of q^* in (17) implies

$$\begin{aligned}
q^*(r^o_c(r^t), r^t) &= \sum_{r^o_n(r^t) \in \mathcal{Y}^{|\mathcal{N}(r^t)|}} q^*(r^o_n(r^t), r^o_c(r^t), r^t) \\
&= \sum_{r^o_{j[1]} \in \mathcal{Y}} \cdots \sum_{r^o_{j[K(r^t)]} \in \mathcal{Y}} q^*_{K(r^t)}((r^o_{j[1]}, \ldots, r^o_{j[K(r^t)]}), r^o_c(r^t), r^t) \\
&= \sum_{r^o_{j[1]} \in \mathcal{Y}} \cdots \sum_{r^o_{j[K(r^t)]} \in \{y^L, y^U\}} q^*_{K(r^t)}(r^o_{-j[K(r^t)]}, r^o_{j[K(r^t)]}, r^t) \\
&= \sum_{r^o_{j[1]} \in \mathcal{Y}} \cdots \sum_{r^o_{j[K(r^t)]} \in \mathcal{Y}} q^*_{K(r^t)-1}(r^o_{-j[K(r^t)]}, r^o_{j[K(r^t)]}, r^t) \\
&= \sum_{r^o_{j[1]} \in \mathcal{Y}} \cdots \sum_{r^o_{j[K(r^t)]} \in \mathcal{Y}} q^*_0((r^o_{j[1]}, \ldots, r^o_{j[K(r^t)]}), r^o_c(r^t), r^t) \\
&= \sum_{r^o_n(r^t) \in \mathcal{Y}^{|\mathcal{N}(r^t)|}} q(r^o_n(r^t), r^o_c(r^t), r^t) \\
&= q(r^o_c(r^t), r^t)\ .
\end{aligned}$$

Therefore for each r^t, $q^*(r^o_c(r^t), r^t) = q(r^o_c(r^t), r^t) \ \ \forall\ r^o_c(r^t)$, and hence by Lemma A.1, q^* satisfies (15) so that $P = Q^* T^{-1}$. Thus $Q^* \in \mathbf{Q}_0(P, \mathbf{Q})$.

Part 4: verify property (c), $\theta(Q^*) = \theta_0$

Note for each $d \in \mathcal{D}$,

$$\mathbb{E}_{Q^*}[Y_d \mathbb{1}\{D_{z^*(d)} = d\}] = \mathbb{E}_P[Y \mathbb{1}\{D = d\} \mid Z = z^*(d)] = \beta_{d|z^*(d)}\ .$$

Further note since Assumption 2.2 holds for Q, we have that for each $d \in \mathcal{D}$ and for each r^t if $r^t_{z^*(d)} = d$ then r^o_d is a component of $r^o_c(r^t)$ and $d \in \mathcal{N}(r^t)^c$, while if $r^t_{z^*(d)} \neq d$ then r^o_d is a component of $r^o_n(r^t)$ and $d \in \mathcal{N}(r^t)$. Then we also have that for each $d \in \mathcal{D}$,

$$\mathbb{E}_{Q^*}[Y_d \mathbb{1}\{D_{z^*(d)} \neq d\}] = \sum_{y \in \mathcal{Y}} \sum_{r^t : r^t_{z^*(d)} \neq d} \sum_{r^o : r^o_d = y} y\ q^*(r^o, r^t)$$

$$
\begin{aligned}
&= \sum_{y\in\mathcal{Y}} \sum_{r^t: r^t_{z^*(d)}\neq d} \sum_{r^o_{-d}} y\ q^*((r^o_{-d}, r^o_d = y), r^t) \\
&= (\alpha_d y^U + (1-\alpha_d) y^L) \sum_{r^t: r^t_{z^*(d)}\neq d} \sum_{r^o} q(r^o, r^t) \\
&= (\alpha_d y^U + (1-\alpha_d) y^L)\ Q\{D_{z^*(d)} \neq d\} \\
&= (\alpha_d y^U + (1-\alpha_d) y^L)(1 - \sum_y p_{yd|z^*(d)})\ ,
\end{aligned}
$$

where the third equality is using that (19) holds for r^t such that $r^t_{z^*(d)} \neq d$, so that $d = j[k']$ for some k' in constructing $q^*(\cdot, r^t)$ for that r^t, and the last equality is using that Q satisfies (15). Thus, for each $d \in \mathcal{D}$,

$$
\begin{aligned}
\mathbb{E}_{Q^*}[Y_d] &= \mathbb{E}_{Q^*}[Y_d \mathbb{1}\{D_{z^*(d)} = d\}] + \mathbb{E}_{Q^*}[Y_d \mathbb{1}\{D_{z^*(d)} \neq d\}] \\
&= \beta_{d|z^*(d)} + (\alpha_d y^U + (1-\alpha_d) y^L)(1 - \sum_y p_{yd|z^*(d)})\ .
\end{aligned}
$$

For any θ_0 contained in (6), we can thus choose $\alpha = (\alpha_0, \alpha_1, ..., \alpha_{|\mathcal{D}|-1})' \in [0,1]^{|\mathcal{D}|}$ such that $\theta(Q^*) = \theta_0$. ■

A.4 Proof of Corollary 3.1

The results follows immediately from Theorem 3.2 because $\mathbb{E}[Y_j] - \mathbb{E}[Y_k]$ is a function of $\theta(Q)$. ■

A.5 Proof of Lemma 3.1

To see it, note for any $d \in \mathcal{D}$ and $z \in \mathcal{Z}$,

$$
\begin{aligned}
\mathbb{E}_Q[Y_d] = \mathbb{E}_Q[Y_d \mid Z = z] &= \mathbb{E}_Q[Y_d \mathbb{1}\{D = d\} \mid Z = z] + \mathbb{E}_Q[Y_d \mathbb{1}\{D \neq d\} \mid Z = z] \\
&= \mathbb{E}_Q[Y \mathbb{1}\{D = d\} \mid Z = z] + \mathbb{E}_Q[Y_d \mathbb{1}\{D \neq d\} \mid Z = z] \\
&= \beta_{d|z} + \mathbb{E}_Q[Y_d \mathbb{1}\{D \neq d\} \mid Z = z] \\
&\leq \beta_{d|z} + y^U P\{D \neq d \mid Z = z\} \\
&= \beta_{d|z} + y^U (1 - \sum_{y\in\mathcal{Y}} p_{yd|z})\ .
\end{aligned}
$$

Because the inequality holds for all $z \in \mathcal{Z}$, the upper end for each $d \in \mathcal{D}$ of (9) is a valid upper bound for $\mathbb{E}[Y_d]$. On the other hand, they can be simultaneously attained for all $d \in \mathcal{D}$ by setting $Y_d = y^U$ whenever $D \neq d$ and $Z = z$, without affecting the distribution of (Y, D, Z). A similar argument can be applied to the lower ends. In addition, any values in between can also be attained simultaneously for all $d \in \mathcal{D}$ by setting Y_d to be a convex combination of y^L and y^U whenever $D \neq d$ and $Z = z$ without affecting the distribution of (Y, D, Z), and therefore (9) is indeed the identified set for $\theta(Q)$ under mean independence. ■

A.6 Proof of Lemma 3.2

Suppose $Q \in \mathbf{Q}_0(P, \mathbf{Q}) \neq \emptyset$. Consider the upper endpoints of (6) and (9). For any $d \in \mathcal{D}$, $z \in \mathcal{Z}$,

$$\begin{aligned}
&\mathbb{E}_P[Y\mathbb{1}\{D=d\} \mid Z = z^*(d)] + y^U \mathbb{E}_P[1 - \mathbb{1}\{D=d\} \mid Z = z^*(d)] \\
&\quad - \mathbb{E}_P[Y\mathbb{1}\{D=d\} \mid Z = z] - y^U \mathbb{E}_P[1 - \mathbb{1}\{D=d\} \mid Z = z] \\
&= \mathbb{E}_Q[Y_d \mathbb{1}\{D_{z^*(d)} = d\}] + y^U \mathbb{E}_Q[1 - \mathbb{1}\{D_{z^*(d)} = d\}] \\
&\quad - \mathbb{E}_Q[Y_d \mathbb{1}\{D_z = d\}] - y^U \mathbb{E}_Q[1 - \mathbb{1}\{D_z = d\}] \\
&= \mathbb{E}_Q[(Y_d - y^U)(\mathbb{1}\{D_{z^*(d)} = d\} - \mathbb{1}\{D_z = d\})] \\
&= \mathbb{E}_Q[(Y_d - y^U)(\mathbb{1}\{D_{z^*(d)} = d, D_z \neq d\} - \mathbb{1}\{D_{z^*(d)} \neq d, D_z = d\})] \\
&= \mathbb{E}_Q[(Y_d - y^U)\mathbb{1}\{D_{z^*(d)} = d, D_z \neq d\}] \\
&\leq 0 \ ,
\end{aligned}$$

where the first equality uses Assumption 2.1 and the fourth equality uses that $Q\{D_{z^*(d)} \neq d, D_z = d\} = 0$ for all Q satisfying Assumption 2.2. Since this inequality holds for all $z \in \mathcal{Z}$, we have that the upper endpoint of the interval in (6) is weakly smaller than the upper endpoint of the interval in (9). Conversely, the upper endpoint of (6) is contained in the set over which the upper endpoint of (9) is minimizing over, and thus the upper endpoint of (6) is weakly larger than the upper endpoint of (9). We conclude that the upper endpoints are the same. Parallel arguments show the equivalence of the lower endpoints. ■

A.7 Proof of Theorem 3.3

Because by assumption $\mathbf{Q} \subseteq \mathbf{Q}' \subseteq \mathbf{Q}^*_{MI}$, we have

$$\Theta_0(P, \mathbf{Q}) \subseteq \Theta_0(P, \mathbf{Q}') \subseteq \Theta_0(P, \mathbf{Q}^*_{MI}) \ .$$

By Lemma 3.2, if $\mathbf{Q} \subseteq \mathbf{Q}^*_{E,M}$, $\mathbf{Q}$ satisfies Assumption 3.1, and $\mathbf{Q}_0(P, \mathbf{Q}) \neq \emptyset$, then we have

$$\Theta_0(P, \mathbf{Q}) = \Theta_0(P, \mathbf{Q}^*_{MI}) \ .$$

The result now follows by a sandwich argument. ■

A.8 Proof of Corollary 3.2

The result follows by taking $\mathbf{Q} = \mathbf{Q}^*_{E,M}$ and $\mathbf{Q}' = \mathbf{Q}^*_E$, noting that $\mathbf{Q}^*_E \subseteq \mathbf{Q}^*_{MI}$ because Assumption 2.1 implies Assumption 3.2. ■

A.9 Proof of Corollary 3.3

To begin, note by assumption $\mathbf{Q}_0(P, \mathbf{Q}^*_{WE,M}) \neq \emptyset$, so there exists $Q \in \mathbf{Q}^*_{WE,M}$ such that $P = QT^{-1}$. We then have $Q\{Z = z\} = P\{Z = z\}$ for all $z \in \mathcal{Z}$, and because Q satisfies Assumption 3.3, we know (2) holds. Let Q_1 denote the marginal distribution of $((Y_d : d \in \mathcal{D}), (D_z : z \in \mathcal{Z}))$ under Q, and Q_Z denote the marginal distribution of Z under Q, and define $\tilde{Q} = Q_1 \times Q_Z$. Then, $\tilde{Q}$ satisfies Assumption 2.1 by construction, and it satisfies Assumption 2.2 because the marginal distribution of $(D_z : z \in \mathcal{Z})$ under $\tilde{Q}$ is the same as that under Q. In summary, $\tilde{Q} \in \mathbf{Q}^*_{E,M}$. Furthermore, $P = \tilde{Q}T^{-1}$ because (1) $\tilde{Q}\{Z = z\} = Q\{Z = z\} = P\{Z = z\}$ for all $z \in \mathcal{Z}$; and (2) $\tilde{Q}\{Y_d = y, D_z = d\} = Q\{Y_d = y, D_z = d\}$ for all $d \in \mathcal{D}$ and $z \in \mathcal{Z}$, and hence (2) is still satisfied. As a result, we know $\mathbf{Q}_0(P, \mathbf{Q}^*_{E,M}) \neq \emptyset$.

Next, take $\mathbf{Q} = \mathbf{Q}^*_{E,M}$ and $\mathbf{Q}' = \mathbf{Q}^*_{WE,M}$, and note that $\mathbf{Q}^*_{E,M} \subseteq \mathbf{Q}^*_{WE,M} \subseteq \mathbf{Q}^*_{MI}$ because Assumption 2.1 implies Assumption 3.3, which in turn implies Assumption 3.2. Because we know $\mathbf{Q}_0(P, \mathbf{Q}^*_{E,M}) \neq \emptyset$ from the previous paragraph, we then obtain from Theorem 3.3 that

$$\Theta_0(P, \mathbf{Q}^*_{E,M}) = \Theta_0(P, \mathbf{Q}^*_{WE,M}) = \Theta_0(P, \mathbf{Q}^*_{MI}) \ .$$

Similarly, taking $\mathbf{Q}' = \mathbf{Q}^*_{WE}$, we have

$$\Theta_0(P, \mathbf{Q}^*_{E,M}) = \Theta_0(P, \mathbf{Q}^*_{WE}) = \Theta_0(P, \mathbf{Q}^*_{MI}) \ .$$

The desired conclusion now follows. ■

B Details of Examples

B.1 Details of Example 4.5

We show that, except in the special case where the treatment and the instrument are both binary, the strict one-to-one targeting assumption of Lee and Salanié (2023) with one or more targeted treatments implies that (12) does not hold for some treatments. To see this, suppose that the strict one-to-one targeting assumption of Lee and Salanié (2023) holds with $|\mathcal{D}^\dagger| \geq 1$. For each $d \in \mathcal{D}^\dagger$, their assumptions include that there exists some $z^\dagger(d)$ and some $\overline{U}(d)$, $\underline{U}(d)$ with $\overline{U}(d) > \underline{U}(d)$ such that

$$g(z, d) = \begin{cases} \overline{U}(d) & \text{if } z = z^\dagger(d) \\ \underline{U}(d) & \text{if } z \neq z^\dagger(d) \ . \end{cases} \tag{20}$$

On the other hand, for each $d \in \mathcal{D} \setminus \mathcal{D}^\dagger$, they impose that $g(z, d) = \underline{U}(d)$ for all $z \in \mathcal{Z}$, and they impose that there is at least one such non-targeted treatment. For any $d \in \mathcal{D} \setminus \mathcal{D}^\dagger$, (12) requires that there exists $z^*(d) \in \mathcal{Z}$ such that

$$g(z, d') > g(z^*(d), d') \quad \text{for all } d' \neq d \text{ and } z \neq z^*(d) \ . \tag{21}$$

Suppose $|\mathcal{Z}| \geq 3$, and fix some targeted treatment $d' \in \mathcal{D}^\dagger$. Suppose $z^*(d) \neq z^\dagger(d')$. Then, for $z \in \mathcal{Z} \setminus \{z^*(d), z^\dagger(d')\}$, (21) requires $\underline{U}(d') > \underline{U}(d')$, a contradiction. Now suppose $z^*(d) = z^\dagger(d')$. Then, for $z \in \mathcal{Z} \setminus \{z^*(d)\}$, (21) requires $\underline{U}(d') > \overline{U}(d')$, a contradiction. Thus, $|\mathcal{Z}| \geq 3$ implies that (12) does not hold for non-targeted treatments.

Now suppose $|\mathcal{Z}| = 2$, which we label as $\mathcal{Z} = \{0, 1\}$, and suppose $|\mathcal{D}| \geq 3$. Without loss of generality suppose $1 \in \mathcal{D}^\dagger$ and $z^\dagger(1) = 1$. If $z^*(d) = 1$, then (21) requires $\underline{U}(1) > \overline{U}(1)$, a contradiction. Now suppose $z^*(d) = 0$. Then (21) requires $g(1, d') > g(0, d')$ for all $d' \neq d$. Consider the following two cases:

- If $|\mathcal{D}^\dagger| = 1$, then $g(1, d') > g(0, d')$ holding for any $d' \in (\mathcal{D} \setminus \mathcal{D}^\dagger) \setminus \{d\}$ requires $\underline{U}(d') > \underline{U}(d')$, a contradiction.
- If $|\mathcal{D}^\dagger| > 1$, then there exists $d'' \in D^\dagger \setminus \{1\}$. By assumption $z^\dagger(d'') \neq z^\dagger(1)$ so that $z^\dagger(d'') = 0$. Then $g(1, d') > g(0, d')$ holding for $d' = d''$ requires $\underline{U}(d'') > \overline{U}(d'')$, again a contradiction.

Thus, $|\mathcal{Z}| = 2$ with $|\mathcal{D}| \geq 3$ implies that (12) does not hold for some treatments.

We have shown (12) does not hold for some treatments when either Z or D takes at least three values. Now suppose $|\mathcal{D}| = |\mathcal{Z}| = 2$. Let $D = 0$ denote the nontargeted treatment and $D = 1$ the targeted treatment, and let $z^\dagger(1) = 1$. Consider $z^*(0) = 0$ and $z^*(1) = 1$. Then evaluating (12) at either $d = 0$ or $d = 1$ results in $\overline{U}(1) > \underline{U}(1)$, and thus (12) holds when $|\mathcal{D}| = |\mathcal{Z}| = 2$. We conclude that the strict one-to-one targeting of Lee and Salanié (2023) implies that (12) does not hold for some $d \in \mathcal{D}$ except in the special case where $|\mathcal{D}| = |\mathcal{Z}| = 2$.

We now show that the strict one-to-one targeting of Lee and Salanié (2023) implies that Assumption 2.2 holds when $|\mathcal{Z}| > |\mathcal{D}^\dagger|$. Let $\mathcal{Z}^\dagger \subseteq \mathcal{Z}$ denote the set of instruments that target some treatment,

$$\mathcal{Z}^\dagger = \{z \in \mathcal{Z} : z = z^\dagger(d) \text{ for some } d \in \mathcal{D}^\dagger\} .$$

Their strict one-to-one targeting assumption combined with $|\mathcal{Z}| > |\mathcal{D}^\dagger|$ implies that there are values of the instrument that do not target any treatment; in other words, $\mathcal{Z}^\dagger \subsetneq \mathcal{Z}$. Following Lee and Salanié (2023), we label the treatment that is known not to be targeted as treatment 0, so that $g(z, 0) = \underline{U}(0)$ for all $z \in \mathcal{Z}$, and impose their normalization that $\underline{U}(0) = 0$. Consider (4) for $d = 0$. Note that

$$Q\{D_{z^*(0)} \neq 0,\ D_{z'} = 0 \text{ for some } z' \neq z^*(0)\} = Q\left\{\bigcup_{d^* \neq 0, z' \neq z^*(0)} D_{z^*(0)} = d^*,\ D_{z'} = 0\right\} .$$

We wish to investigate whether there exists some $z^*(0) \in \mathcal{Z}$ such that the above probability is zero. Consider $z^*(0)$ equal to any value in $\mathcal{Z} \setminus \mathcal{Z}^\dagger$, i.e., any value of the instrument that does not target any treatment. For any fixed $d^* \neq 0, z' \neq z^*(0)$, consider the event $\{D_{z^*(0)} = d^*,\ D_{z'} = 0\}$. Since $z^*(0)$ does not target any treatment and thus does not target d^*, $D_{z^*(0)} = d^*$ implies

$$U_0 - U_{d^*} \leq \underline{U}(d^*) . \tag{22}$$

If z' targets d^*, then $D_{z'} = 0$ implies

$$U_0 - U_{d^*} \geq \overline{U}(d^*) \ . \tag{23}$$

If z' does not target d^*, then $D_{z'} = 0$ implies

$$U_0 - U_{d^*} \geq \underline{U}(d^*) \ . \tag{24}$$

Thus, the event $\{D_{z^*(0)} = d^*,\ D_{z'} = 0\}$ either requires (22) and (23) to jointly hold, which is a contradiction since $\overline{U}(d^*) > \underline{U}(d^*)$, or requires (22) and (24) to jointly hold, which is a zero probability event given our assumption that the distribution of $(U_d : d \in \mathcal{D})$ is absolutely continuous w.r.t. Lebesgue measure. Thus $Q\{D_{z^*(0)} \neq 0,\ D_{z'} = 0 \text{ for some } z' \neq z^*(0)\}$ is a probability of a finite union of zero probability events, and thus, by Boole's inequality, equals zero so that (4) holds for $d = 0$. A parallel argument shows that (4) holds for any non-targeted treatment, and related argument shows that (4) holds for any targeted treatment. Thus, under the strict one-to-one targeting of Lee and Salanié (2023), when there are more values of the instrument than targeted treatments, Assumption 2.2 holds even though (12) is violated for some $d \in \mathcal{D}$.

B.2 Details of Example 5.1

Let $\mathbf{Q}$ denote all distributions Q for which Assumption 2.1 holds and such that $Q\{D_0 = D_1\} = 0$. Then for $Q \in \mathbf{Q}_0(P, \mathbf{Q})$,

$$\begin{aligned}
p_{y1|1} &= Q\{Y_1 = y, D_1 = 1, D_0 = 0\} \\
p_{y0|0} &= Q\{Y_0 = y, D_1 = 1, D_0 = 0\} \\
p_{y0|1} &= Q\{Y_0 = y, D_1 = 0, D_0 = 1\} \\
p_{y1|0} &= Q\{Y_1 = y, D_1 = 0, D_0 = 1\}
\end{aligned}$$

and

$$\begin{aligned}
Q\{Y_0 = 1\} &= Q\{Y_0 = 1, D_1 = 1, D_0 = 0\} + Q\{Y_0 = 1, D_1 = 0, D_0 = 1\} \\
&= p_{10|0} + p_{10|1} \\
Q\{Y_1 = 1\} &= Q\{Y_1 = 1, D_1 = 1, D_0 = 0\} + Q\{Y_1 = 1, D_1 = 0, D_0 = 1\} \\
&= p_{11|1} + p_{11|0} \ .
\end{aligned}$$

Therefore, if $\mathbf{Q}$ is consistent with P, then $\theta(Q)$ is identified as

$$\Theta_0(P, \mathbf{Q}) = \left\{ \begin{pmatrix} p_{10|0} + p_{10|1} \\ p_{11|0} + p_{11|1} \end{pmatrix} \right\} \ . \tag{25}$$

In contrast, the identified set that follows from imposing Assumption 2.1 alone, $\Theta_0(P, \mathbf{Q}^*_E)$, is shown by

Balke and Pearl (1997) to be

$$\max\left\{\begin{array}{c} p_{10|1} \\ p_{10|0} \\ p_{10|0}+p_{11|0}-p_{00|1}-p_{11|1} \\ p_{01|0}+p_{10|0}-p_{00|1}-p_{01|1} \end{array}\right\} \leq Q\{Y_0=1\} \leq \min\left\{\begin{array}{c} 1-p_{00|1} \\ 1-p_{00|0} \\ p_{01|0}+p_{10|0}+p_{10|1}+p_{11|1} \\ p_{10|0}+p_{11|0}+p_{01|1}+p_{10|1} \end{array}\right\} \tag{26}$$

and

$$\max\left\{\begin{array}{c} p_{11|0} \\ p_{11|1} \\ -p_{00|0}-p_{01|0}+p_{00|1}+p_{11|1} \\ -p_{01|0}-p_{10|0}+p_{10|1}+p_{11|1} \end{array}\right\} \leq Q\{Y_1=1\} \leq \min\left\{\begin{array}{c} 1-p_{01|1} \\ 1-p_{01|0} \\ p_{00|0}+p_{11|0}+p_{10|1}+p_{11|1} \\ p_{10|0}+p_{11|0}+p_{00|1}+p_{11|1} \end{array}\right\}. \tag{27}$$

It follows from $\mathbf{Q} \subseteq \mathbf{Q}_E^*$ that $\Theta_0(P,\mathbf{Q}) \subseteq \Theta_0(P,\mathbf{Q}_E^*)$, and thus (25) is contained in (26)–(27).

Next, we show that there exists a P for which $\mathbf{Q}_0(P,\mathbf{Q}) \neq \emptyset$, $\Theta_0(P,\mathbf{Q})$ is not given by (6) and $\Theta_0(P,\mathbf{Q}) \subsetneqq \Theta_0(P,\mathbf{Q}_E^*)$. We do so by providing a numerical example. Consider the P specified in Table 1 and the Q specified in Table 2, where we write $q(y_0y_1,d_0d_1) = Q\{Y_d=y_d, D_z=d_z,\ (d,z)\in\mathcal{D}\times\mathcal{Z}\}$ and omit any $q(\cdot)=0$. One can check that $Q\in\mathbf{Q}$ and rationalizes P, so that $Q\in\mathbf{Q}_0(P,\mathbf{Q})\neq\emptyset$. In this example, $Q\{Y_0=1\}=0.4274$, and thus the identified set for $Q\{Y_0=1\}$ relative to $\mathbf{Q}$ is the singleton $\{0.4274\}$. In contrast, evaluating (26) at P gives the identified set for $Q\{Y_0=1\}$ relative to $\mathbf{Q}_E^*$ as $[0.3336, 0.5212]$. We thus conclude that $\Theta_0(P,\mathbf{Q}) \subsetneqq \Theta_0(P,\mathbf{Q}_E^*)$ for some P that can be rationalized by $Q\in\mathbf{Q}$. Now consider evaluating the bounds of (6) at the same P. Doing so results in bounds on $Q\{Y_0=1\}$ given by $[0.1618, 0.5445]$ if setting $z^*(0)=0$ and given by $[0.2656, 0.8829]$ if setting $z^*(0)=1$. Therefore, no matter $z^*(0)=0$ or 1, the bounds (6) is not the identified set for $Q\{Y_0=1\}$ relative to either $\mathbf{Q}_E^*$ or $\mathbf{Q}$.

$p_{00\|0}$ 0.4555	$p_{10\|0}$ 0.1618	$p_{01\|0}$ 0.3077	$p_{11\|0}$ 0.0750
$p_{00\|1}$ 0.1171	$p_{10\|1}$ 0.2656	$p_{01\|1}$ 0.0188	$p_{11\|1}$ 0.5985

Table 1: Distribution P in Appendix B.2.

$q(00,01)$ 0.0039	$q(00,10)$ 0.0428	$q(01,01)$ 0.4516	$q(01,10)$ 0.0743
$q(10,01)$ 0.0149	$q(10,10)$ 0.2649	$q(11,01)$ 0.1469	$q(11,10)$ 0.0007

Table 2: Distribution Q.

B.3 Details of Example 5.2

Suppose $\mathcal{Y}=\{0,1\}$, $\mathcal{D}=\{0,1,2\}$, and $\mathcal{Z}=\{0,1,2\}$. Then, the linear program approach in Balke and Pearl (1993, 1997) leads to the following identified set for $\mathbb{E}_Q[Y_1]=Q\{Y_1=1\}$ relative to $\mathbf{Q}$ being defined as in

Example 5.2:

$$\left[\max\left\{\begin{array}{c} p_{11|0} \\ p_{11|1} \\ p_{11|2} \\ p_{11|0}-p_{11|1}+p_{11|2}\end{array}\right\},\quad \min\left\{\begin{array}{c} 1-p_{01|2} \\ 1-p_{01|1} \\ 1-p_{01|0} \\ 1-p_{01|0}+p_{01|1}-p_{01|2}\end{array}\right\}\right] . \tag{28}$$

We will show that for some P such that $\mathbf{Q}_0(P,\mathbf{Q}) \neq \emptyset$, we have $\Theta_0(P,\mathbf{Q})$ strictly smaller than (6) and $\Theta_0(P,\mathbf{Q}_1^*)$. For this purpose we are only concerned with the validity of (28) instead of its sharpness. For the lower bounds, first note for $z \in \mathcal{Z}$,

$$Q\{Y_1=1\} = Q\{Y_1=1 \mid Z=z\} \geq Q\{Y_1=1, D=1 \mid Z=z\} = Q\{Y=1, D=1 \mid Z=z\} ,$$

and therefore the first three rows follow. To show the last row, note it's equivalent to

$$Q\{Y_1=1, D_1=1\} + Q\{Y_1=1, D_0=0\} + Q\{Y_1=1, D_0=2\} \geq Q\{Y_1=1, D_2=1\} .$$

It therefore suffices to show that

$$\{D_2=1\} \implies \{D_1=1\} \cup \{D_0=0\} \cup \{D_0=2\} . \tag{29}$$

Suppose $D_2=1$ but $D_0 \neq 0$ and $D_0 \neq 2$. Then $D_0=1$. But $D_0 \leq D_1 \leq D_2$, so $D_1=1$. (29) now follows. The lower bounds in (28) have all been shown to hold, and the upper bounds can be proved similarly.

Next, we show that there exists a P for which $\mathbf{Q}_0(P,\mathbf{Q}) \neq \emptyset$, $\Theta_0(P,\mathbf{Q})$ is not given by (6) and $\Theta_0(P,\mathbf{Q}) \subsetneqq \Theta_0(P,\mathbf{Q}_E^*)$. We do so by providing a numerical example. Consider the P specified in Table 3 and the four Q distributions specified in Table 4, 5, 6 and 7, which we denote as $Q_{\text{ex,min}}$, $Q_{\text{ex,max}}$, $Q_{\text{ex,om,min}}$ and $Q_{\text{ex,om,max}}$ respectively, where we write $q(y_0y_1y_2, d_0d_1d_2) = Q\{Y_d=y_d, D_z=d_z,\ (d,z) \in \mathcal{D}\times\mathcal{Z}\}$ and omit any $q(\cdot)=0$. One can check that all the four Qs are in $\mathbf{Q}_0(P,\mathbf{Q}_E^*)$, i.e., they all rationalize P and satisfy Assumption 2.1. Moreover, $Q_{\text{ex,om,min}} \in \mathbf{Q}_0(P,\mathbf{Q})$ and $Q_{\text{ex,om,max}} \in \mathbf{Q}_0(P,\mathbf{Q})$ so that $\mathbf{Q}_0(P,\mathbf{Q}) \neq \emptyset$. Evaluating (28) at P gives $[0.2117, 0.8205] =: I_{\text{ex,om}}$. In contrast, if one evaluates (6) by setting $z^*(1) = 0, 1, 2$ at the same P, the resulting bounds for $\mathbb{E}_Q[Y_1]$ are $[0.1664, 0.9255] =: I_{(6),0}$, $[0.0712, 0.9311] =: I_{(6),1}$ and $[0.1165, 0.8261] =: I_{(6),2}$ respectively. In all cases, we see $I_{\text{ex,om}} \subsetneqq I_{(6),z^*(1)}$ for all $z^*(1) \in \{0,1,2\}$ so $\Theta_0(P,\mathbf{Q})$ is not given by (6). Furthermore, $Q_{\text{ex,min}} \notin \mathbf{Q}_0(P,\mathbf{Q})$ and $Q_{\text{ex,max}} \notin \mathbf{Q}_0(P,\mathbf{Q})$ because, for example, $q_{\text{ex,min}}(000, 021) > 0$ and $q_{\text{ex,max}}(000, 210) > 0$. At the same time, $\mathbb{E}_{Q_{\text{ex,min}}}[Y_1] = 0.1664 \notin I_{\text{ex,om}}$ and $\mathbb{E}_{Q_{\text{ex,max}}}[Y_1] = 0.8261 \notin I_{\text{ex,om}}$. Therefore, $\Theta_0(P,\mathbf{Q}) \subsetneqq \Theta_0(P,\mathbf{Q}_E^*)$.

$p_{00\|0}$ 0.3808	$p_{10\|0}$ 0.2427	$p_{01\|0}$ 0.0745	$p_{11\|0}$ 0.1664	$p_{02\|0}$ 0.0345	$p_{12\|0}$ 0.1011
$p_{00\|1}$ 0.2830	$p_{10\|1}$ 0.1947	$p_{01\|1}$ 0.0689	$p_{11\|1}$ 0.0712	$p_{02\|1}$ 0.2014	$p_{12\|1}$ 0.1808
$p_{00\|2}$ 0.0802	$p_{10\|2}$ 0.0976	$p_{01\|2}$ 0.1739	$p_{11\|2}$ 0.1165	$p_{02\|2}$ 0.2444	$p_{12\|2}$ 0.2874

Table 3: Distribution P in Appendix B.3.

$q(000,002)$ 0.0139	$q(000,011)$ 0.0304	$q(000,021)$ 0.0054	$q(001,002)$ 0.2238	$q(001,020)$ 0.0802	$q(001,021)$ 0.0271	$q(001,121)$ 0.0735	$q(001,211)$ 0.0375
$q(010,101)$ 0.0453	$q(010,111)$ 0.0712	$q(010,122)$ 0.0499	$q(100,000)$ 0.0966	$q(100,022)$ 0.1461	$q(100,110)$ 0.0010	$q(100,202)$ 0.0345	$q(101,202)$ 0.0636

Table 4: Distribution $Q_{\text{ex,min}}$.

$q(001,021)$ 0.0305	$q(001,121)$ 0.0745	$q(001,211)$ 0.0689	$q(010,000)$ 0.0220	$q(010,001)$ 0.0064	$q(010,002)$ 0.0085	$q(010,010)$ 0.0260	$q(010,122)$ 0.1664
$q(010,202)$ 0.0345	$q(011,002)$ 0.2116	$q(011,022)$ 0.0758	$q(011,210)$ 0.0322	$q(110,000)$ 0.0976	$q(110,001)$ 0.0971	$q(110,011)$ 0.0130	$q(110,022)$ 0.0350

Table 5: Distribution $Q_{\text{ex,max}}$.

$q(000,000)$ 0.0802	$q(000,001)$ 0.0079	$q(000,002)$ 0.0430	$q(000,022)$ 0.0181	$q(000,111)$ 0.0209	$q(000,122)$ 0.0536	$q(000,222)$ 0.0345	$q(001,002)$ 0.1066
$q(001,022)$ 0.0797	$q(001,222)$ 0.1011	$q(010,001)$ 0.0453	$q(010,111)$ 0.0712	$q(010,122)$ 0.0952	$q(100,000)$ 0.0976	$q(100,001)$ 0.0971	$q(100,011)$ 0.0480

Table 6: Distribution $Q_{\text{ex,om,min}}$.

$q(000,001)$ 0.0079	$q(000,111)$ 0.0689	$q(000,122)$ 0.0056	$q(010,000)$ 0.0802	$q(010,001)$ 0.1114	$q(010,012)$ 0.0430	$q(010,111)$ 0.0051	$q(010,122)$ 0.1613
$q(010,222)$ 0.0345	$q(011,002)$ 0.0835	$q(011,022)$ 0.0548	$q(011,222)$ 0.1011	$q(100,001)$ 0.0971	$q(110,000)$ 0.0976	$q(111,012)$ 0.0231	$q(111,022)$ 0.0249

Table 7: Distribution $Q_{\text{ex,om,max}}$.

B.4 Details of Example 5.3

Consider the ARUM defined by (11) with strict one-to-one targeting and $|\mathcal{D}| = 3$, $|\mathcal{Z}| = |\mathcal{D}^\dagger| = 2$. In this case, there are two targeted treatments and one non-targeted treatment. Following Lee and Salanié (2023), label that non-targeted treatment as treatment 0 and impose the normalization that $g(z,0) = 0$ for all $z \in \mathcal{Z}$. Label $Z = 0$ as the instrument value that targets treatment 1 and label $Z = 1$ as the instrument value that targets treatment 2, so that (20) holds for $d = 1, 2$ for some $\overline{U}(d), \underline{U}(d)$ with $\overline{U}(d) > \underline{U}(d)$ and with $z^\dagger(1) = 0$, $z^\dagger(2) = 1$. Let $\mathbf{Q}$ denote the set of all distributions for which $(D_z : z \in \mathcal{Z})$ is determined by (11) with these restrictions and additionally imposing that the support of (U_0, U_1, U_2) is $\Re^3$. Recall $(U_0, U_1, U_2) \perp\!\!\!\perp Z$ by assumption.

Let $U_{10} = U_1 - U_0$ and $U_{20} = U_2 - U_0$. In this model, the treatment value is completely determined by the vector of realizations (U_{10}, U_{20}). For instance, $D_z = 2$ if and only if

$$U_{20} \geq -g(z,2)$$
$$U_{20} - U_{10} \geq g(z,1) - g(z,2) \ ,$$

and a similar characterization holds for $D_z = 1$. See Figure 1, which is taken from Figure 1 in Lee and Salanié (2023).

We first show that (4) holds for the targeted treatments by verifying that (12) holds for the targeted

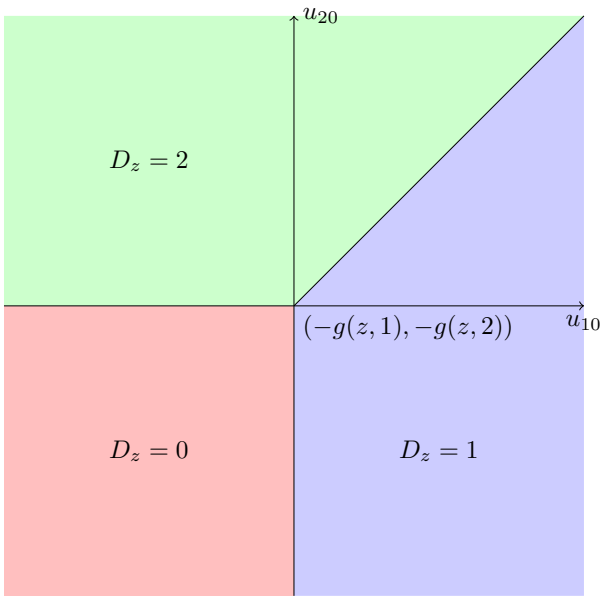

Figure 1: Treatment under each value of (u_{10}, u_{20}) for a given z.

treatments. Consider (12) for $d = 1$. It holds with $z^*(1) = z^\dagger(1) = 0$ because

$$\begin{aligned} \overline{U}(1) &> \underline{U}(1) \ , \\ \overline{U}(1) - \underline{U}(2) &> \underline{U}(1) - \overline{U}(2) \ , \end{aligned}$$

which in turn holds because $\overline{U}(d) > \underline{U}(d)$ for $d \in \{1, 2\}$. Thus (12) holds for $d = 1$, which, as shown in Example 4.4, implies that (4) holds for $d = 1$. By a parallel argument, (4) holds for $d = 2$.

We now show that there does not exist a value of $z^*(0)$ such that (4) holds for the non-targeted treatment, treatment 0. Suppose $z^*(0) = 0$. Then

$$\begin{aligned} Q\{D_0 \neq 0,\ D_1 = 0\} &\geq Q\{D_0 = 1,\ D_1 = 0\} \\ &= Q\{-\underline{U}(1) \geq U_{10} \geq -\overline{U}(1), U_{20} \leq -\overline{U}(2), U_{10} - U_{20} \geq \underline{U}(2) - \overline{U}(1)\} \\ &> 0 \ , \end{aligned}$$

where the last line is using that the support of the distribution of $(U_{10}, U_{20}) = \Re^2$ by assumption and that strict targeting of treatment 1 requires $-\underline{U}(1) > -\overline{U}(1)$. Thus (4) cannot hold for $d = 0$ with $z^*(0) = 0$. A parallel argument shows that (4) cannot hold for $d = 0$ with $z^*(0) = 1$.

We conclude that, when $|\mathcal{D}| = 3$ and $|\mathcal{D}^\dagger| = |\mathcal{Z}| = 2$, one-to-one strict targeting with the regularity condition that the support of (U_0, U_1, U_2) is $\Re^3$ implies that (4) holds for the targeted treatments but not for the non-targeted treatments, and thus Assumption 2.2 cannot hold. This argument can be adapted for any ARUM with $|\mathcal{D}| \geq 3$ and $|\mathcal{Z}| = |\mathcal{D}^\dagger|$ to show that, while (4) holds for the targeted treatments, there does not exist a value of $z^*(d)$ such that (4) holds for any non-targeted treatment d, and thus that Assumption 2.2 cannot hold.

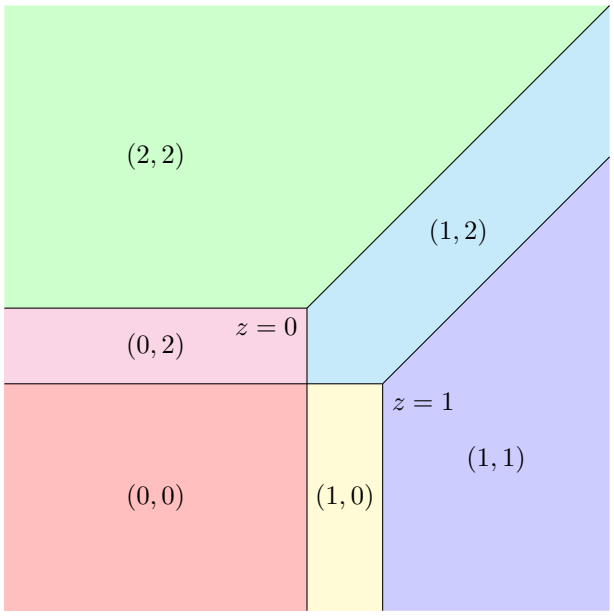

Figure 2: Values of (D_0, D_1) for each value of (u_{10}, u_{20}).

Next, consider the identified sets for the average potential outcomes. Since (4) is satisfied for the targeted treatments, a straightforward modification of the arguments underlying Theorem 3.2 show that the identified sets for $\mathbb{E}_Q[Y_d]$ for $d \in \{1, 2\}$ is given by (6) for any P such that $\mathbf{Q}_0(P, \mathbf{Q}) \neq \emptyset$.

We now derive the identified set for the average potential outcome of the non-targeted treatment. First, note that $-g(0,1) = -\overline{U}(1) < -\underline{U}(1) = -g(1,1)$ and $-g(0,2) = -\underline{U}(2) > -\overline{U}(2) = -g(1,2)$. Therefore, it can be verified from Figure 2 that for all $Q \in \mathbf{Q}$,

$$Q\{(D_0, D_1) \in \{(0,0), (1,0), (1,1), (0,2), (1,2), (2,2)\}\} = 1 \ . \tag{30}$$

Let $\mathbf{Q}'$ denote the set of all distributions that satisfies (30). Note that all $Q \in \mathbf{Q}$ satisfies (30), so $\mathbf{Q} \subseteq \mathbf{Q}'$. On the other hand, by assigning appropriate probabilities to each set in the partition in Figure 2, we immediately see that each $Q \in \mathbf{Q}'$ can be rationalized by a $Q \in \mathbf{Q}$. Therefore, $\mathbf{Q} = \mathbf{Q}'$. Using linear programming as in Balke and Pearl (1993, 1997), we obtain the following identified set for $\mathbb{E}_Q[Y_0] = Q\{Y_0 = 1\}$ relative to $\mathbf{Q}$:

$$\left[\max \begin{Bmatrix} p_{10|0} \\ p_{10|1} \end{Bmatrix}, \quad \min \begin{Bmatrix} 1 - p_{00|1} \\ 1 - p_{00|0} \end{Bmatrix} \right] \ . \tag{31}$$

The identified set in (31) equals (9) for $d = 0$ with Y and Z binary. Thus, the identified set for $\mathbb{E}_Q[Y_0]$ relative to $\mathbf{Q}$ corresponds to the identified set relative to $\mathbf{Q}_3^*$, the set of distributions that satisfy mean independence, 3.2. By the same sandwich argument used to prove Theorem 3.3, the identified set for $\mathbb{E}_Q[Y_0]$ relative to $\mathbf{Q}$ corresponds to the identified set relative to $\mathbf{Q}_1^*$, and thus imposing this ARUM has no identifying power for $\mathbb{E}_Q[Y_0]$ beyond instrument exogeneity.

Finally, we show that there exists a P for which $\mathbf{Q}_0(P, \mathbf{Q}) \neq \emptyset$ and (31) is strictly smaller than (6), so

that $\Theta_0(P, \mathbf{Q})$ is not given by (6). We do so by providing a numerical example. Consider the P specified in Table 8 and the $Q_{\text{arum,min}}$ and $Q_{\text{arum,max}}$ specified in Tables 9 and 10 respectively, where we write $q(y_0y_1y_2, d_0d_1) = Q\{Y_d = y_d, D_z = d_z,\ (d,z) \in \mathcal{D} \times \mathcal{Z}\}$ and omit any $q(\cdot) = 0$. One can check that both $Q_{\text{arum,min}}$ and $Q_{\text{arum,max}}$ rationalize P and satisfy Assumption 2.1. One can further check that both $Q_{\text{arum,min}}$ and $Q_{\text{arum,max}}$ satisfy the restriction in (30), so that $\mathbf{Q}_0(P, \mathbf{Q}) \neq \emptyset$. Evaluating (31) at P gives the identified set for $\mathbb{E}_Q[Y_0]$ relative to $\mathbf{Q}$ as $[0.2518, 0.8167]$. One can further check that the two endpoints are attained by $\mathbb{E}_{Q_{\text{arum,min}}}[Y_0] = 0.2518$ and $\mathbb{E}_{Q_{\text{arum,max}}}[Y_0] = 0.8167$. On the other hand, if one evaluates (6) by setting $z^*(0) = 0, 1$ at the same P, the resulting bounds for $\mathbb{E}_Q[Y_0]$ equal $[p_{10|0}, 1 - p_{00|0}] = [0.2518, 0.8937]$ and $[p_{10|1}, 1 - p_{00|1}] = [0.2372, 0.8167]$ respectively. In both cases, (31) is strictly contained in (6). Hence, $\Theta_0(P, \mathbf{Q})$ is not given by (6).

$p_{00\|0}$	$p_{10\|0}$	$p_{01\|0}$	$p_{11\|0}$	$p_{02\|0}$	$p_{12\|0}$
0.1063	0.2518	0.2946	0.3183	0.0020	0.0270
$p_{00\|1}$	$p_{10\|1}$	$p_{01\|1}$	$p_{11\|1}$	$p_{02\|1}$	$p_{12\|1}$
0.1833	0.2372	0.0140	0.1399	0.1701	0.2555

Table 8: Distribution P in Appendix B.4.

$q(000, 10)$	$q(000, 11)$	$q(000, 12)$	$q(000, 22)$	$q(001, 02)$	$q(001, 12)$
0.0049	0.0140	0.1535	0.0020	0.1063	0.1222
$q(001, 22)$	$q(010, 10)$	$q(010, 11)$	$q(100, 00)$	$q(100, 02)$	
0.0270	0.1784	0.1399	0.2372	0.0146	

Table 9: Distribution $Q_{\text{arum,min}}$.

$q(000, 00)$	$q(010, 10)$	$q(100, 00)$	$q(100, 10)$	$q(100, 11)$	$q(100, 02)$
0.1063	0.0770	0.0837	0.0521	0.0140	0.1681
$q(100, 22)$	$q(101, 12)$	$q(101, 22)$	$q(110, 10)$	$q(110, 11)$	
0.0020	0.2285	0.0270	0.1014	0.1399	

Table 10: Distribution $Q_{\text{arum,max}}$.

C Additional Examples of Models That Satisfy Assumption 2.2

In Section 4, we considered examples of restrictions on potential treatments previously considered in the literature that satisfy generalized monotonicity. We now consider three additional such examples.

Example C.1. Consider the ARUM of Example 4.4 when $|\mathcal{D}| = 2$, and let $\mathbf{Q}$ denote the set of distributions defined in that example. Then, Assumption 2.2 holds for all $Q \in \mathbf{Q}$. To see this, consider $Q \in \mathbf{Q}$. Label $\mathcal{D} = \{0, 1\}$, and let $g_{10}(z) = g(z, 1) - g(z, 0)$ and $U_{10} = U_1 - U_0$. The assumptions of Example 4.4 on (U_1, U_0) imply that the distribution of U_{10} is absolutely continuous with respect to Lebesgue measure and that $U_{10} \perp\!\!\!\perp Z$. Ignoring ties that occur with probability zero, (11) can be rewritten as

$$D_z = \mathbb{1}\{g_{10}(z) + U_{10} \geq 0\} \ . \tag{32}$$

Let $\overline{\mathcal{Z}} = \text{argmax}_{z \in \mathcal{Z}}\{g_{10}(z)\}$, and let $\underline{\mathcal{Z}} = \text{argmin}_{z \in \mathcal{Z}}\{g_{10}(z)\}$. Then, Q satisfies Assumption 2.2 with

$\mathcal{Z}^*(1) = \overline{\mathcal{Z}}$ and $\mathcal{Z}^*(0) = \underline{\mathcal{Z}}$. To contrast with Example 4.4, note that (12) holds if and only if $\overline{\mathcal{Z}}$ and $\underline{\mathcal{Z}}$ are both singletons. ■

Example C.2. Kline and Walters (2016) considers an RCT with a "close substitute" to study the effects of preschooling on educational outcomes. In their setting, $D \in \mathcal{D} = \{0, 1, 2\}$, where $D = 0$ denotes home care (no preschool), $D = 2$ denotes a preschool program called Head Start, and $D = 1$ denotes preschools other than Head Start, namely the close substitute. Let $Z \in \mathcal{Z} = \{0, 1\}$ denote an indicator variable for an offer to attend Head Start. Assumption 2.1 holds because Z is randomly assigned. Kline and Walters (2016) impose the restriction that

$$Q\{D_1 = 2 \mid D_0 \neq D_1\} = 1 \ . \tag{33}$$

The condition in (33) states that if the choice of a family changes upon receiving a Head Start offer, then they must choose Head Start when receiving the offer. In other words, it cannot be the case that upon receiving a Head Start offer, a family switches from no preschool to preschools other than Head Start, or the other way around. Assumption 2.2 then holds with $z^*(0) = z^*(1) = 0$ and $z^*(2) = 1$. To see this, note that (33) implies $Q\{D_0 \neq D_1, D_1 \neq 2\} = 0$ and thus

$$\begin{aligned}
Q\{D_0 \neq 0, D_1 = 0\} &= Q\{D_0 \neq D_1, D_1 = 0\} = 0 \ , \\
Q\{D_0 \neq 1, D_1 = 1\} &= Q\{D_0 \neq D_1, D_1 = 1\} = 0 \ , \\
Q\{D_1 \neq 2, D_0 = 2\} &\leq Q\{D_0 \neq D_1, D_1 \neq 2\} = 0 \ .
\end{aligned}$$

Note in this example Assumption 2.2 still holds although $|\mathcal{Z}| < |\mathcal{D}|$. See Bai et al. (2025b) for results on the sharp testable implications of the assumptions for this example and Example C.3. ■

Example C.3. Kirkeboen et al. (2016) study the effects of fields of study on earnings. In their setting, $\mathcal{D} = \{0, 1, 2\}$ represent three fields of study, ordered by their (soft) admission cutoffs from the lowest to the highest. The instrument is $Z \in \{0, 1, 2\}$, with $Z = 1$ when the student crosses the (soft) admission cutoff for field 1, $Z = 2$ when the student crosses the (soft) admission cutoff for field 2, and $Z = 0$ otherwise. The authors assume that Z is exogenous in the sense that Q satisfies Assumption 2.1 and impose the following monotonicity conditions:

$$Q\{D_1 = 1 \mid D_0 = 1\} = 1 \ , \tag{34}$$

$$Q\{D_2 = 2 \mid D_0 = 2\} = 1 \ . \tag{35}$$

The conditions in (34)–(35) require that crossing the cutoff for field 1 or 2 weakly encourages them towards that field. They further impose the following "irrelevance" conditions:

$$Q\{\mathbb{1}\{D_1 = 2\} = \mathbb{1}\{D_0 = 2\} \mid D_0 \neq 1, D_1 \neq 1\} = 1 \ , \tag{36}$$

$$Q\{\mathbb{1}\{D_2 = 1\} = \mathbb{1}\{D_0 = 1\} \mid D_0 \neq 2, D_2 \neq 2\} = 1 \ . \tag{37}$$

The condition in (36) states that if crossing the cutoff for field 1 does not cause the student to switch to field 1, then it does not cause them to switch to or away from field 2. A similar interpretation applies to

(37). Lee and Salanié (2023) show the set of all distributions that satisfy (34)–(37) are equivalent to a strict one-to-one targeting model with $|\mathcal{Z}| = 3$ and $|\mathcal{D}^\dagger| = 2$; it therefore follows from Remark 4.5 that any Q that satisfies (34)–(37) also satisfies Assumption 2.2. Here, we establish directly that (34)–(37) imply Assumption 2.2 with $z^*(0) = 0$, $z^*(1) = 1$, and $z^*(2) = 2$. To show $z^*(0) = 0$, we prove by contradiction that

$$Q\{D_0 \neq 0, D_1 = 0\} = 0 \ .$$

Suppose with positive probability that $D_0 \neq 0$ but $D_1 = 0$. On this event, (34) implies $D_0 \neq 1$, so $D_0 = 2$. But $D_1 = 0$, which contradicts (36). Similarly,

$$Q\{D_0 \neq 0, D_2 = 0\} = 0 \ ,$$

and therefore $z^*(0) = 0$. To show $z^*(1) = 1$, first note (34) implies

$$Q\{D_1 \neq 1, D_0 = 1\} = 0 \ .$$

It therefore remains to argue by contradiction that

$$Q\{D_1 \neq 1, D_2 = 1\} = 0 \ . \tag{38}$$

Suppose with positive probability that $D_1 \neq 1$ but $D_2 = 1$. On this event, (34) implies $D_0 \neq 1$. If $D_0 = 2$, then (35) implies $D_2 = 2$, a contradiction to $D_2 = 1$; if instead $D_0 = 0$, then because we assume $D_2 = 1$, (37) implies $D_2 \neq 1$, another contradiction. Therefore, (38) holds, and $z^*(1) = 1$. $z^*(2) = 2$ can be established following similar arguments. ■

References

Andrews, D. W. and Schafgans, M. M. (1998). Semiparametric estimation of the intercept of a sample selection model. *The Review of Economic Studies*, **65** 497–517.

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

Angrist, J. D., Imbens, G. W. and Rubin, D. B. (1996). Identification of causal effects using instrumental variables. *Journal of the American Statistical Association*, **91** 444–455.

Bai, Y., Huang, S., Moon, S., Santos, A., Shaikh, A. M. and Vytlacil, E. J. (2025a). Inference for Treatment Effects Conditional on Generalized Principal Strata using Instrumental Variables. ArXiv:2411.05220 [econ], URL http://arxiv.org/abs/2411.05220.

Bai, Y., Huang, S. and Tabord-Meehan, M. (2025b). Sharp Testable Implications of Encouragement Designs. ArXiv:2411.09808 [econ], URL http://arxiv.org/abs/2411.09808.

Balke, A. and Pearl, J. (1993). Nonparametric bounds on causal effects from partial compliance data. Technical Report R-199, UCLA.

Balke, A. and Pearl, J. (1997). Bounds on treatment effects from studies with imperfect compliance. *Journal of the American Statistical Association*, **92** 1171–1176.

Cheng, J. and Small, D. S. (2006). Bounds on causal effects in three-arm trials with non-compliance. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **68** 815–836.

Frangakis, C. E. and Rubin, D. B. (2002). Principal stratification in causal inference. *Biometrics*, **58** 21–29.

Han, S. and Yang, S. (2024). A computational approach to identification of treatment effects for policy evaluation. *Journal of Econometrics*, **240** 105680.

Heckman, J. (1990). Varieties of selection bias. *The American Economic Review*, **80** 313–318.

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

Heckman, J. J. and Vytlacil, E. (2005). Structural equations, treatment effects, and econometric policy evaluation. *Econometrica*, **73** 669–738.

Heckman, J. J. and Vytlacil, E. J. (1999). Local instrumental variables and latent variable models for identifying and bounding treatment effects. *Proceedings of the national Academy of Sciences*, **96** 4730–4734.

Heckman, J. J. and Vytlacil, E. J. (2001). Instrumental variables, selection models, and tight bounds on the average treatment effect. In *Econometric Evaluation of Labour Market Policies* (M. Lechner and F. Pfeiffer, eds.). Physica-Verlag HD, 1–15.

Imbens, G. W. and Angrist, J. D. (1994). Identification and estimation of local average treatment effects. *Econometrica*, **62** 467–475.

Kamat, V. (2019). On the identifying content of instrument monotonicity. ArXiv:1807.01661 [econ], URL http://arxiv.org/abs/1807.01661.

Kirkeboen, L. J., Leuven, E. and Mogstad, M. (2016). Field of Study, Earnings, and Self-Selection. *The Quarterly Journal of Economics*, **131** 1057–1111. URL https://doi.org/10.1093/qje/qjw019.

Kitagawa, T. (2021). The identification region of the potential outcome distributions under instrument independence. *Journal of Econometrics*, **225** 231–253.

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

Lee, S. and Salanié, B. (2023). Treatment Effects with Targeting Instruments. ArXiv:2007.10432 [econ, stat], URL http://arxiv.org/abs/2007.10432.

Machado, C., Shaikh, A. M. and Vytlacil, E. J. (2019). Instrumental variables and the sign of the average treatment effect. *Journal of Econometrics*, **212** 522–555. URL http://www.sciencedirect.com/science/article/pii/S0304407619301381.

Manski, C. F. (1990). Nonparametric bounds on treatment effects. *The American Economic Review*, **80** 319–323.

Marx, P. (2024). Sharp bounds in the latent index selection model. *Journal of Econometrics*, **238** 105561.

Mogstad, M., Santos, A. and Torgovitsky, A. (2018). Using instrumental variables for inference about policy relevant treatment parameters. *Econometrica*, **86** 1589–1619.

Richardson, T. S. and Robins, J. M. (2013). Single world intervention graphs (swigs): A unification of the counterfactual and graphical approaches to causality. *Center for the Statistics and the Social Sciences, University of Washington Series. Working Paper*, **128** 2013.

Robins, J. M. (1989). The analysis of randomized and non-randomized aids treatment trials using a new approach to causal inference in longitudinal studies. *Health service research methodology: a focus on AIDS* 113–159.

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

Vytlacil, E. (2006). Ordered discrete-choice selection models and local average treatment effect assumptions: Equivalence, nonequivalence, and representation results. *The Review of Economics and Statistics*, **88** 578–581.