

# Ordered, Unordered and Minimal Monotonicity Criteria

Manu Navjeevan and Rodrigo Pinto\*

Revised July 11, 2022

## Abstract

This paper investigates the relationship among monotonicity conditions in IV models with multiple choices and categorical instruments. The comparison between monotonicity conditions of ordered and unordered choice models is central to our analysis. We show that these seemingly unrelated conditions exhibit non-trivial symmetries that can be traced back to a weaker condition called *Minimal Monotonicity*. This novel condition captures an essential property for identifying causal parameters while being necessary for ascribing causal interpretation to certain Two-Stage Least Squares (2SLS) estimands. We show that minimal monotonicity naturally arises from a notion of rationality in revealed preference analysis. The condition enables researchers to describe non-standard choice behaviors and serves as a building block for a wide range of economically-justified monotonicity conditions that do not fit the narrative dictated by either ordered or unordered choice models.

*Keywords:* Monotonicity, Instrumental Variables, Discrete Choice, Selection bias, Roy Model, Identification, Discrete Mixture Model.

*JEL codes:* I21, C93, J15, V16.

---

\*Rodrigo Pinto is an Assistant Professor of the Economics Department at University of California at Los Angeles (UCLA), and Manu Navjeevan is a PhD candidate at UCLA. We are most grateful for the interaction with Professors James Heckman, Steven Durlauf, Moshe Buchinsky, Andres Santos and Rosa Matzkin. We also benefit from useful conversations with Stephane Bonhomme, Azeem Shaikh, Maurizio Mazzocco, Juan Pantano, and Parag Pathak. We also thank useful comments on presentations at University of California at Los Angeles, University of Chicago, Science Po (Paris), and University of Geneva, All errors are our own.

## 1 Introduction

Economists have long used instrumental variables (IV) to identify the causal effect of an endogenous treatment choice on outcomes of interest. An IV is characterized by two core properties: it is an exogenous variable, and it affects the outcome only through its impact on the treatment choice. These properties, however, are insufficient to identify treatment effects. The early literature on IV models ensures identification by imposing strong functional form restrictions (Theil, 1953; Heckman and Robb, 1985). Unfortunately, these assumption often prevent the possibility of heterogeneous treatment effects across agents (Heckman et al., 2006).

In an influential paper, Imbens and Angrist (1994) investigate weaker assumptions that secure the nonparametric identification of a causal effect in a binary choice model with heterogenous agents. They introduce the *monotonicity condition*, which states that a change in the instrument must induce all agents to shift their choices towards the same treatment status. The monotonicity condition has several desirable properties: it is simple, intuitive, and renders the identification of the Local Average Treatment Effect (LATE) via the Two-Stage Least Squares (2SLS) estimand. The condition gives rise to notions such as the marginal treatment effect and response functions (Heckman and Vytlacil, 1999, 2005). Finally, the monotonicity condition of Imbens and Angrist (1994) is frequently regarded as the minimal criteria necessary to ascribe causal interpretation to the 2SLS regression.<sup>1</sup>

Contrary to what one might expect, identification in IV models with multiple treatment choices is not a straightforward extension of the binary choice case. Multiple choices allow for a variety of distinct monotonicity conditions that collapse into the same condition when the treatment is limited to only two choices. Notably, Angrist and Imbens (1995) apply the monotonicity of Imbens and Angrist (1994) to the case of multiple treatments. The condition is equivalent to assuming an ordered choice model (Vytlacil, 2006), hence we term the condition as *ordered monotonicity*. In contrast, the *unordered monotonicity* of Heckman and Pinto (2018) applies to settings where treatment choices are not ordered. Both conditions imply choice restrictions that identify causal parameters that can be estimated by 2SLS regressions. However, neither of the conditions implies nor is implied by the other. Despite their different formulations, both conditions are equivalent to the Imbens and Angrist (1994) monotonicity in the case of binary choices.

The properties of the ordered monotonicity of Angrist and Imbens (1995) and the unordered monotonicity of Heckman and Pinto (2018) raise several questions: Do these conditions share some of the choice restrictions they imply? Is there a weaker monotonicity condition underlying both conditions? If this weaker condition exists, is it equivalent to Imbens and Angrist (1994) in the binary choice model? Are there some criteria that enable us to characterize the similarities and differences among monotonicity conditions of IV models with multiple choices? What is the economic rationale

---

<sup>1</sup>Huber and Mellace (2012) is an example of a paper that considers identification of the LATE via 2SLS under alternate conditions

that justifies the differences among these monotonicity conditions?

This paper uses the ordered and unordered monotonicity conditions of Angrist and Imbens (1995) and Heckman and Pinto (2018) as leading examples to address some of these questions. Specifically, we address three intertwined inquiries. First, we investigate the relationship between ordered and unordered monotonicity. To do so, we update the equivalence results of Vytlacil (2006) and Heckman and Pinto (2018) to provide symmetric characterizations for both monotonicity conditions. These characterizations enable us to study common properties of ordered and unordered monotonicity and set the stage for joint analysis of the two conditions.

Second, we consider whether a weaker and more inclusive condition exists that subsumes the ordered and the unordered monotonicity conditions. We leverage the symmetric characterizations of ordered and unordered monotonicity conditions to show that both conditions share a common property termed *minimal monotonicity*. We show that the minimal monotonicity condition is both interpretable and informative in conducting causal inference. The condition implies the monotonicity of Imbens and Angrist (1994) in the case of binary choices. It is a necessary condition to assign causal interpretation for the 2SLS estimand for any two values of the IV. We provide a simple verification criterion for determining whether the condition holds. We use this analysis to develop formal criteria for comparing arbitrary monotonicity conditions.

Finally, we study the economic content of various monotonicity conditions. We demonstrate that minimal monotonicity can be justified by a basic notion of choice rationality. The condition serves as a foundation for a variety of economically-justified monotonicity conditions. We present examples of common economic settings in which minimal monotonicity holds but ordered and unordered monotonicity fail. More generally, we show that a broad range of monotonicity conditions emerge by combining choice incentives with the notion of choice rationality that induces minimal monotonicity.

This paper contributes to the theoretical literature that examines equivalence properties of ordered and unordered choice models (Vytlacil, 2002, 2006; Heckman and Pinto, 2018). It also adds to the recent IV literature that explore extensions of standard monotonicity conditions (Kamat, 2021; Mogstad et al., 2018; Mogstad and Torgovitsky, 2018; Hull, 2018). Finally, it relates to the empirical literature that uses non-standard monotonicity conditions to evaluation of treatment effects (Pinto, 2021; Kline and Walters, 2016; Mountjoy, 2021; Feller et al., 2016; Brinch et al., 2017; Kirkeboen et al., 2016).

This paper proceeds as follows. Section 2 reviews the prior literature on monotonicity conditions. Section 3 describes the IV model and introduces our notation. Section 4 discusses the content of ordered and unordered monotonicity conditions and revisits the equivalence results for ordered and unordered choice models. It explores the symmetry of equivalence results between these two models to motivate a novel monotonicity condition. Section 5 discusses the properties of the Minimal Monotonicity Condition. Section 6 discusses the economic content of the minimal monotonicity condition. Section 7 discuss some applications of monotonicity criteria that are economically justified.

Section 8 concludes.

## 2 Literature Review

Historically, the traditional approach to evaluating IV models has been to use structural equations to describe the agent's choice (Goldberger, 1972; Heckman, 1976, 1979). Imbens and Angrist (1994) departed from this traditional approach. They describe the IV model using the language of potential outcomes (Rubin, 1974, 1978a) instead of structural equations. They introduced the notion of *monotonicity*, which formalizes an intuitive assumption stating that an IV change induces all agents toward choosing the same treatment choice.<sup>2</sup>

Angrist and Imbens (1995) extend the monotonicity condition of Imbens and Angrist (1994) to the case of multiple choices. They show that the monotonicity condition provides a causal interpretation of the conventional Two-Stage Square Least Squares (2SLS) estimand in models with endogenous choices and heterogeneous treatment responses. Their work spiked a substantial literature on both empirical and theoretical aspects of monotonicity conditions (Angrist et al., 2000; Barua and Lang, 2016; Dahl et al., 2017; Huber and Mellace, 2012, 2015; Imbens and Rubin, 1997; Klein, 2010; Small and Tan, 2007; Aliprantis, 2012; de Chaisemartin, 2017).<sup>3</sup>

Vytlacil (2002, 2006) bridge the gap between IV models that rely on monotonicity conditions and the previous literature that invokes structural equations. Vytlacil (2002) shows that the monotonicity condition of Imbens and Angrist (1994) is equivalent to assume that the binary choice is determined by a latent threshold-crossing indicator Heckman and Vytlacil (1999, 2005, 2007a). In a similar fashion, Vytlacil (2006) shows the monotonicity criterion of Angrist and Imbens (1995) is equivalent to an ordered choice model with random thresholds. This ordered choice model is examined by Cameron and Heckman (1998) and further studied by Carneiro et al. (2003); Cunha et al. (2007).

Unordered choice models have been studied mostly by the IV literature on structural equations. A common approach is to assume that the equations that govern the treatment are generated by additively separable threshold-crossing models. Examples of this literature are Heckman and Vytlacil (2007b); Heckman et al. (2006, 2008). A substantial contribution to this literature is due to Lee and Salanié (2018), who studied the identification of causal effects for choice models defined by an arbitrary set of threshold-crossing rules. Heckman and Pinto (2018) connect the structural and monotonicity approaches. They present an economically motivated condition termed *unordered monotonicity* which applies to treatment values that do not have a natural order. They build upon Vytlacil (2002) to show that unordered monotonicity can be equivalently expressed as a multivariate choice model with a latent crossing thresholds.

Little is known about the similarities between ordered and unordered choice models. The economic

---

<sup>2</sup>See also Angrist et al. (1996).

<sup>3</sup>Huber et al. (2017) consider weaker assumptions at the principal strata level, which are also employed by Frölich (2007).

rationale that motives ordered choice models differ considerably from the reasoning that justifies unordered choice models. Not surprisingly, each model often carries distinct mathematical formalizations. A rare example of a comparative discussion between ordered and unordered choice models is Heckman et al. (2006). Their ordered choice model employs a partition of the real line by non-stochastic thresholds. The treatment choice indicates the interval that the latent stochastic index lies in this partition. In contrast, their unordered choice model employs a set of latent indexes that are additive in the observed and unobserved characteristics of the agent.

As mentioned, we perform a comparative analysis between ordered and unordered monotonicities. To do so, we revisit the monotonicity condition of Angrist and Imbens (1995) using new tools of analysis developed in Heckman and Pinto (2018).

### 3 Setup

Our IV model consists of three observed variables: a categorical instrument  $Z$  that takes  $N_Z$  values in  $\mathcal{Z} = \{z_1, \dots, z_{N_Z}\}$ ; a multiple treatment choice  $T$  that takes  $N_T$  values in  $\mathcal{T} = \{t_1, \dots, t_{N_T}\}$ ; and a real-valued outcome  $Y \in \mathbb{R}$ .<sup>4</sup> These variables belong to the probability space  $(\mathcal{I}, \mathcal{F}, P)$  where  $i \in \mathcal{I}$  denotes an individual. Most of the IV literature describes model assumptions via the potential outcome framework (Rubin, 1978b; Holland, 1986), where  $Y_i(t, z)$  denotes the potential outcome for individual  $i$  when  $Z_i, T_i$  take values  $z, t$  and  $T_i(z)$  denotes the potential choice for  $i$  when the instrument  $Z_i$  is set to the value  $z$ . The assumptions that characterize the core properties of the IV are:

$$\text{Exogeneity Condition: } Z_i \perp\!\!\!\perp (T_i(z), Y_i(t)) \text{ for all } (z, t, i) \in \mathcal{Z} \times \mathcal{T} \times \mathcal{I} \quad (1)$$

$$\text{Exclusion Restriction: } Y_i(t, z) = Y_i(t, z') \text{ for all } z, z' \in \mathcal{Z}, \text{ and } (t, i) \in \mathcal{T} \times \mathcal{I} \quad (2)$$

$$\text{IV Relevance: } T \not\perp Z \text{ (Not statistically independent)} \quad (3)$$

The exogeneity assumption means that the instrument  $Z$  is as good as random. The exclusion restriction implies that  $Z$  does not directly cause  $Y$  and the IV relevance states that  $T$  and  $Z$  correlate.<sup>5</sup> The IV model can be equivalently described by structural equations. The structural approach has the advantage of enabling us to express the individual's unobserved characteristics that generate selection bias as a random vector  $\mathbf{V}$ . The IV model is described by the following equations:

$$\text{Choice Equation : } T = f_T(Z, \mathbf{V}), \quad (4)$$

$$\text{Outcome Equation : } Y = f_Y(T, \mathbf{V}, \epsilon), \quad (5)$$

$$\text{Independence Condition: } Z, \mathbf{V}, \epsilon \text{ are statistically independent} \quad (6)$$

<sup>4</sup>We suppress pre-treatment variables  $X$  from the model for the sake of notational simplicity. The analysis can be understood as conditioned on these variables.

<sup>5</sup>More specifically, we have that  $P(T = t|Z = z) > 0$  for all  $(t, z) \in \mathcal{T} \times \mathcal{Z}$  and that  $E(\mathbf{D}_T' \mathbf{D}_Z)$  is full-rank where  $\mathbf{D}_T = [\mathbf{1}[T = t_1], \dots, \mathbf{1}[T = t_{N_T}]]'$  and  $\mathbf{D}_Z = [\mathbf{1}[Z = z_1], \dots, \mathbf{1}[Z = z_{N_Z}]]'$  are vectors that indicate the treatment and IV values.

The choice equation (4) means that  $T$  is caused by the instrument  $Z$  and unobserved characteristics  $\mathbf{V}$ . The outcome equation (5) states that  $Y$  is caused by choice  $T$ , unobserved characteristics  $\mathbf{V}$  and an unobserved error term  $\epsilon$  that is exogenous.<sup>6</sup> Functions  $f_T(\cdot)$  and  $f_Y(\cdot)$  are not observed and can take arbitrary functional forms. The counterfactual (or potential) choice when the instrument were fixed to a value  $z \in \mathcal{Z}$  is given by  $T(z) \equiv f_T(z, \mathbf{V})$  and the counterfactual outcome generated by fixing  $T$  to a value  $t$ , that is  $Y(t) \equiv f_Y(t, \mathbf{V}, \epsilon)$ .<sup>7</sup> The independence condition (6) states that  $Z$  is statistically independent of the individual’s unobserved characteristics  $\mathbf{V}$  and the error term  $\epsilon$ . The condition implies the exogeneity condition (1) of the potential outcome framework.<sup>8</sup> It also implies a matching (or unconfoundedness) condition:<sup>9</sup>

$$\text{Matching Condition: } Y(t) \perp\!\!\!\perp T | \mathbf{V} \text{ for all } t \in \mathcal{T}. \quad (7)$$

Condition (7) states that  $Y(t)$  becomes statistically independent of  $T$  when conditioning for  $\mathbf{V}$ . If  $\mathbf{V}$  were observable, we would be able to evaluate the counterfactual outcome  $Y(t)$  by conditioning the outcome  $Y$  on  $T = t$  and  $\mathbf{V}$ . The primary consequence of the matching condition (7) is that the identification of causal effects hinges on controlling for the unobserved characteristics  $\mathbf{V}$ . Identification strategies seek to exploit the exogenous variation of  $Z$  to control for  $\mathbf{V}$ .

Controlling for unobserved characteristics is a daunting task since  $\mathbf{V}$  can have an arbitrary dimension. The *response vector*  $\mathbf{S}$  in (8) facilitates this task:

$$\mathbf{S} = [T(z_1), \dots, T(z_{N_Z})]^\top, \quad \text{supp}(\mathbf{S}) \equiv \{\mathbf{s}_1, \dots, \mathbf{s}_{N_S}\} \quad (8)$$

The response vector  $\mathbf{S}$  is the  $N_Z \times 1$  unobserved random vector that stacks the counterfactual choices  $T(z)$  across the IV-values  $z$  in  $\mathcal{Z}$ . Elements of the support of the response vector,  $\mathbf{s} \in \text{supp}(\mathbf{S})$ , are called *response-types*. Each element  $T(z)$  in  $\mathbf{S}$  may take any of the  $N_T$  values in  $\mathcal{T}$ . Thus, the number of possible response-types totals  $N_T^{N_Z}$ .

The LATE model of [Imbens and Angrist \(1994\)](#) investigates the case of a binary instrument  $\mathcal{Z} = \{z_0, z_1\}$  and a binary treatment  $\mathcal{T} = \{t_0, t_1\}$ . The response vector  $\mathbf{S} = [T(z_0), T(z_1)]^\top$ , admits four possible response-types: never-takers  $\mathbf{s}_{nt} = [t_0, t_0]^\top$ , compliers  $\mathbf{s}_c = [t_0, t_1]^\top$ , always-takers  $\mathbf{s}_{at} = [t_1, t_1]^\top$ , and defiers  $\mathbf{s}_d = [t_1, t_0]^\top$ . Note that the choice  $T$  is fully determined by the instrument  $Z$  for any given response-type  $\mathbf{s}$ . For instance, compliers  $\mathbf{s}_c$  choose  $T = t_1$  for  $Z = z_1$  and  $T = t_0$  for  $Z = z_0$ . We can express the treatment choice as a function of  $\mathbf{S}$  and  $Z$  as  $T = [\mathbf{1}[Z = z_1], \dots, \mathbf{1}[Z = z_{N_Z}]] \cdot \mathbf{S}$ , where  $\mathbf{1}[\cdot]$  denotes then indicator function.

The response vector  $\mathbf{S}$  simplifies the identification problem by playing the role of a balancing score

<sup>6</sup>Error term  $\epsilon$  is used so that  $Y$  conditioned on  $\mathbf{V}$  and  $Z$  is not deterministic.

<sup>7</sup>See [Heckman and Pinto \(2014\)](#) and [Pinto and Heckman \(2021\)](#) for a discussion on causal models and the fixing operator.

<sup>8</sup>Indeed, the counterfactuals  $Y(t), T(z)$  are a function of  $\mathbf{V}$  and  $\epsilon$ , which, according to (6), are statistically independent of  $Z$ . The exclusion restriction (2) arises because  $Z$  is not an argument in the outcome equation (6). IV relevance (3) corresponds to the assumption that  $Z$  causes  $T$  in the choice equation (4).

<sup>9</sup>The independence condition (6) implies that  $\epsilon \perp\!\!\!\perp Z | \mathbf{V}$ . Thereby  $f_Y(t, \mathbf{V}, \epsilon) \perp\!\!\!\perp f_T(Z, \mathbf{V}) | \mathbf{V} \Rightarrow Y(t) \perp\!\!\!\perp T | \mathbf{V}$ .

for  $\mathbf{V}$ . That is,  $\mathbf{S}$  is a function of  $\mathbf{V}$  that preserves the matching property (7), that is,  $Y(t) \perp\!\!\!\perp T | \mathbf{S}$  holds.<sup>10</sup> This matching property enables us to express the expected values of observed data in terms of unobserved counterfactuals:<sup>11</sup>

$$\underbrace{E(Y|T = t, Z = z)P(T = t|Z = z)}_{\text{Observed}} = \sum_{\mathbf{s} \in \text{supp}(\mathbf{S})} \underbrace{\mathbf{1}[T = t | \mathbf{S} = \mathbf{s}, Z = z]}_{\text{Known}} \cdot \underbrace{E(Y(t) | \mathbf{S} = \mathbf{s})P(\mathbf{S} = \mathbf{s})}_{\text{Unobserved}}. \quad (9)$$

Equation (9) characterizes the identification problem of IV models as the solution of a system of linear equations. We seek to identify the unobserved outcome counterfactuals and response-type probabilities on the right-hand side of equation (9) using the observed outcome expectations and propensity scores on the left-hand side. The general solution to this problem requires some matrix notation.

The response matrix  $\mathbf{R}$  is central to our analysis. It organizes the response-types in  $\text{supp}(\mathbf{S})$  into a  $N_Z \times N_S$  array where each column displays a response-type and each row corresponds to an instrument value:

$$\mathbf{R} \equiv [\mathbf{s}_1, \dots, \mathbf{s}_{N_S}] \in \mathcal{T}^{N_Z \times N_S}. \quad (10)$$

We use  $\mathbf{B}_t = \mathbf{1}[\mathbf{R} = t]$  for the  $N_Z \times N_S$  binary matrix that takes value one if the respective entry in  $\mathbf{R}$  is  $t$  and zero otherwise. The value in the  $z$ -th row and  $\mathbf{s}$ -th column of  $\mathbf{B}_t$  is given by  $\mathbf{B}_t[z, \mathbf{s}] = \mathbf{1}[T = t | \mathbf{S} = \mathbf{s}, Z = z]$ . The response matrix in the LATE model is given by:

$$\mathbf{R} = \begin{bmatrix} \mathbf{s}_{nt} & \mathbf{s}_c & \mathbf{s}_{at} & \mathbf{s}_d \\ t_0 & t_0 & t_1 & t_1 \\ t_0 & t_1 & t_1 & t_0 \end{bmatrix} \begin{matrix} z_0 \\ \vdots \\ z_1 \end{matrix} \cdot \mathbf{B}_{t_0} = \begin{bmatrix} \mathbf{s}_{nt} & \mathbf{s}_c & \mathbf{s}_{at} & \mathbf{s}_d \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix}, \mathbf{B}_{t_1} = \begin{bmatrix} \mathbf{s}_{nt} & \mathbf{s}_c & \mathbf{s}_{at} & \mathbf{s}_d \\ 0 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 \end{bmatrix}. \quad (11)$$

In this notation, we can rewrite equation (9) as:

$$\mathbf{Q}_Z(t) \odot \mathbf{P}_Z(t) = \mathbf{B}_t \cdot (\mathbf{Q}_S(t) \odot \mathbf{P}_S) \quad \text{for all } t \in \mathcal{T}, \quad (12)$$

where

$$\begin{aligned} \mathbf{Q}_Z(t) &= [E(Y|T = t, Z = z_1), \dots, E(Y|T = t, Z = z_{N_Z})], \\ \mathbf{P}_Z(t) &= [P(T = t|Z = z_1), \dots, P(T = t|Z = z_{N_Z})] \end{aligned}$$

are the observed vectors of outcome expectations and propensity scores and

$$\begin{aligned} \mathbf{Q}_S(t) &= [E(Y(t)|\mathbf{S} = \mathbf{s}_1), \dots, E(Y(t)|\mathbf{S} = \mathbf{s}_{N_S})], \\ \mathbf{P}_S &= [P(\mathbf{S} = \mathbf{s}_1), \dots, P(\mathbf{S} = \mathbf{s}_{N_S})] \end{aligned}$$

<sup>10</sup>Indeed,  $\mathbf{S}$  is a function of  $\mathbf{V}$  because  $T(z); z \in \mathcal{Z}$  are functions of  $\mathbf{V}$ . Moreover,  $Y(t) \perp\!\!\!\perp T | \mathbf{S}$  holds because, given  $\mathbf{S}$ , the treatment decision  $T$  depends only on the instrument  $Z$  which is independent of the potential outcome  $Y(t)$ .

<sup>11</sup>See Heckman and Pinto (2018) for a proof.

are the unobserved vectors of counterfactual outcomes and response-type probabilities of interest. The symbol  $\odot$  denotes element-wise (Hadamard) multiplication.

Equation (12) serves two purposes in our analysis. Firstly, it establishes that the identification of causal parameters stems from the rank properties of the binary matrices  $\mathbf{B}_t$ . If  $\mathbf{B}_t$  were invertible, we would be able to identify all counterfactual outcomes by simply inverting  $\mathbf{B}_t$ .<sup>12</sup> However, in general the number of response types will be larger than the number of instruments and so  $\mathbf{B}_t$  will not be invertible. More generally then, Heckman and Pinto (2018) show that, for any given subset of response-types  $\mathcal{S} \subset \text{supp}(\mathbf{S})$ ,

$$E(Y(t)|\mathbf{S} \in \mathcal{S}) \text{ is identified if and only if } \mathbf{b}(\mathcal{S})'(\mathbf{I} - \mathbf{B}_t^+ \mathbf{B}_t)\mathbf{b}(\mathcal{S}) = 0, \quad (13)$$

where  $\mathbf{I}$  is the identity matrix,  $\mathbf{B}_t^+$  is the Moore-Penrose pseudo-inverse of  $\mathbf{B}_t$  and  $\mathbf{b}(\mathcal{S}) = [\mathbf{1}[s_1 \in \mathcal{S}], \dots, \mathbf{1}[s_{N_S} \in \mathcal{S}]]'$  is a binary vector that indicates which response-type belongs to  $\mathcal{S}$ .<sup>13</sup>

Equation (12) can help understand how monotonicity conditions ensure the identification of causal parameters. The equation illustrates a fundamental identification problem. The number of known parameters in the left-hand side of (12) totals  $N_Z \cdot N_T$ . The number of unknown parameters in the right-hand side of (12) is proportional to the number response-types in  $\text{supp}(\mathbf{S})$ , which totals  $N_T^{N_Z}$ . An identification problem arises since the number of known parameters grows linearly in  $N_Z$  while the number of unknowns grows exponentially in  $N_Z$ . Monotonicity conditions solve this problem by imposing choice restrictions that systematically eliminate potential response-types in  $\text{supp}(\mathbf{S})$ . Monotonicity conditions effectively equate the growth rate of the known parameters in the right-hand side of (12) and unknown parameters of the left-hand side of (12). For instance, the monotonicity condition of Imbens and Angrist (1994) reduces the number of response-types of the binary choice model from  $2^{N_Z}$  to  $N_Z + 1$ .

## 4 Ordered and Unordered Monotonicity

Monotonicity conditions are choice restrictions that eliminate response-types systematically. As discussed, the ordered monotonicity of Angrist and Imbens (1995) and unordered monotonicity of Heckman and Pinto (2018) provide monotonicity criteria for ordered and unordered choice models, respectively.

**Ordered Monotonicity (OM):** For any  $z, z' \in \mathcal{Z}$  either,

$$\begin{aligned} T_i(z) &\geq T_i(z') \text{ for all } i \in \mathcal{I} \\ \text{or } T_i(z) &\leq T_i(z') \text{ for all } i \in \mathcal{I}. \end{aligned} \quad (14)$$

**Unordered Monotonicity (UM):** For any  $z, z' \in \mathcal{Z}$  and any  $t \in \mathcal{T}$  either,

<sup>12</sup>That is,  $\mathbf{Q}_S(t) = (\mathbf{B}_t^{-1} \mathbf{Q}_Z(t) \odot \mathbf{P}_Z(t)) \div (\mathbf{B}_t^{-1} \mathbf{P}_Z(t))$ , where  $\div$  denotes element-wise division.

<sup>13</sup>If  $E(Y(t)|\mathbf{S} \in \mathcal{S})$  is identified, then it can be evaluated by the expression  $E(Y(t)|\mathbf{S} \in \mathcal{S}) = \frac{\mathbf{b}(\mathcal{S})' \mathbf{B}_t^+ (\mathbf{Q}_Z(t) \odot \mathbf{P}_Z(t))}{\mathbf{b}(\mathcal{S})' \mathbf{B}_t^+ \mathbf{P}_Z(t)}$ .



$$\begin{aligned} & \mathbf{1}[T_i(z) = t] \geq \mathbf{1}[T_i(z') = t] \text{ for all } i \in \mathcal{I} \\ \text{or } & \mathbf{1}[T_i(z) = t] \leq \mathbf{1}[T_i(z) = t] \text{ for all } i \in \mathcal{I} \end{aligned} \tag{15}$$

OM (14) captures the notion that a change in instrumental values produces incentives that either move all agents towards weakly “higher” treatment values or move all agents towards weakly “lower” treatment values. The condition can be understood as stating that an instrumental change that induces one agent to increase their treatment choice cannot cause another agent to decrease their treatment choice. As such, the condition requires an ordinal treatment, such as years of schooling.

UM (15) states that for each treatment, each instrumental change must either move all agents weakly towards that treatment or weakly away from the treatment. This differs from OM (14) as it compares *the indicator function* of the treatment instead of the treatment value itself. Because of this, UM (15) does not require ordered treatments, making it relevant for settings where the treatment has no natural ordering such as analysis of college major choice or neighborhood effects.<sup>14</sup>

Importantly, both OM (14) and UM (15) enable the researcher to identify a mixture of Local Average Treatment Effects (LATEs) with identifiable weights and both conditions ascribe causal interpretations to the estimands of Two-Stage Least Squares (2SLS) regressions.

#### 4.1 Expressing Monotonicities as Sequences of Counterfactual Choices

Because the definition of OM (14) compares treatment values, it requires that  $\mathcal{T}$  be an ordered set. We propose a slightly more inclusive definition of ordered monotonicity that does not require an ordered treatment. The central property of ordered monotonicity is a mapping between a sequence of IV values and some sequence of treatment values in which higher rankings of  $Z$  correspond to higher rankings of  $Z$ . The following formula expresses this criterion:

**OM Sequence:** There exist a sequencing of  $\mathcal{Z}$ ,  $(z_1, \dots, z_{N_Z})$ , and a strict ordering on  $\mathcal{T}$  such that:<sup>15</sup>

$$(T_i(z_1), \dots, T_i(z_{N_Z})) \text{ is an increasing sequence in } \mathcal{T} \text{ for any } i \in \mathcal{I}. \tag{16}$$

The OM sequential criteria (16) generates the OM condition (14) whenever the ordering  $\mathcal{T}$  is assumed, however it does not require a specific ordering on  $\mathcal{T}$  a priori. In Section 7 we will demonstrate the usefulness of this more inclusive definition with a plausible research design that generates OM-Sequence (16) on a treatment space that has no natural ordering.

We can also characterize the UM condition in (15) in terms of a sequence of counterfactual choices:

---

<sup>14</sup>Additionally, Heckman and Pinto (2018) show that UM occurs naturally in economic settings where choice incentives weakly increase among all treatment choices as the instrument varies. Buchinsky and Pinto (2021) use revealed preference analysis to show how choice incentives induced by the instrumental variable generate a range of monotonicity conditions.

<sup>15</sup>A strict ordering is one such that for any  $t, t' \in \mathcal{T}$  with  $t \neq t'$  exactly one of  $(t' \geq t)$  or  $(t \geq t')$  is true.

**UM Sequence:** For each  $t \in \mathcal{T}$  there exists a sequencing of  $\mathcal{Z}$ ,  $(z_1^{(t)}, \dots, z_{N_Z}^{(t)})$  such that:

$$(\mathbf{1}[T_i(z_1^{(t)}) = t], \dots, \mathbf{1}[T_i(z_{N_Z}^{(t)}) = t]) \text{ is weakly increasing for any } i \in \mathcal{I}. \quad (17)$$

UM Sequence (17) differs from OM Sequence (16) in two significant ways. First, the sequence of IV values in the unordered case can differ across treatment values while the IV sequence of ordered case remains the same for all  $t \in \mathcal{T}$ . Second, UM Sequence (17) utilizes treatment indicators, while the OM Sequence (16) employs the treatment values themselves.

It is easy to see that the OM and UM sequence characterizations in (16) and (17) are equivalent for a binary treatment. However, this equivalence between ordered and unordered monotonicity breaks down for choice models with three or more treatment choices. In general, ordered monotonicity does not imply unordered monotonicity nor vice versa. To partially demonstrate why this is the case, consider an ordering on the treatments where  $t_1 \leq t_2 \leq t_3$  and the following two pairs of treatment response patterns:

$$\begin{array}{cc} \mathbf{s}_a & \mathbf{s}_b \\ \begin{pmatrix} t_1 & t_2 \\ t_2 & t_3 \end{pmatrix} & z \\ & z' \end{array} \quad \begin{array}{cc} \mathbf{s}_c & \mathbf{s}_d \\ \begin{pmatrix} t_1 & t_3 \\ t_2 & t_2 \end{pmatrix} & z \\ & z' \end{array} \quad (18)$$

The treatment response patterns displayed by agents  $\mathbf{s}_a$  and  $\mathbf{s}_b$  are natural under ordered monotonicity, they could be rationalized by a research design where  $z'$  provides uniformly greater treatment incentives than  $z$ . However, they cannot both exist in a response matrix that satisfies unordered monotonicity. By examining UM Sequence (17) we can see that there would be no possible sequencing of the instruments that would allow the sequence of  $t_2$  indicators to be weakly increasing for all agents.

Similarly, the treatment response patterns displayed by agents  $\mathbf{s}_c$  and  $\mathbf{s}_d$  in (18) are not prohibited by unordered monotonicity; they can both be rationalized by a research design where instrument  $z'$  explicitly incentivizes treatment  $t_2$ . Despite this, they cannot both be present in a response matrix that satisfies ordered monotonicity under the ordering on the treatments given above. Since the switch from instrument  $z$  to instrument  $z'$  induces agent  $\mathbf{s}_c$  to switch to a “lower” treatment while inducing agent  $\mathbf{s}_d$  to switch to a “higher” treatment, there can be no sequencing on the instruments satisfying OM Sequence (16).

The two conditions are not mutually exclusive, namely, a response matrix can satisfy both ordered and unordered monotonicity, even when the treatment is multi-valued. Moreover, the lack of nesting between ordered and unordered monotonicity does not change if we consider all possible orderings on the treatment space. For a complete example of this and more in depth discussion, refer to Appendix C.

## 4.2 Symmetric Characterizations of Unordered and Ordered Monotonicity

The symmetry between the sequential characterizations of ordered and unordered monotonicity in (14)-(15) suggest that other properties of ordered and unordered monotonicity may be analyzed symmetrically.

Table 1 displays two equivalence results that emphasize the symmetry between the ordered and the unordered monotonicity conditions.

Theorem 1 builds upon the unordered monotonicity equivalence of Heckman and Pinto (2018). The first two items of Theorem 1 reflect the discussion in Section 4 relating UM-Sequence (17) to the classical definition of unordered monotonicity introduced in Heckman and Pinto (2018) and restated in (14). Item (iii) resembles the discussion above and states that unordered monotonicity can be verified by individually checking each  $2 \times 2$  submatrix of  $\mathbf{R}$ . Unfortunately, this requirement may be difficult to verify in practice, since the number of  $2 \times 2$  submatrices is growing exponentially with the dimensions of the response matrix. Item (iv) of the theorem provides a practical method of verifying the condition using matrix algebra. Item (v) is familiar to the literature and provides an equivalence between unordered monotonicity and separability conditions such as in Vytlačil (2002). For further discussion, Heckman and Pinto (2018) describe other useful aspects of each of the equivalent statements, (iii) and (v) above.

Theorem 2 extends Vytlačil (2006) in a fashion that enables us to compare these monotonicity conditions. The first and second items of the ordered equivalence result reconcile the two notions of ordered monotonicity presented above. It shows that if OM-Sequence (16) holds, we can find an ordering on  $\mathcal{T}$  that satisfies the typical definition of ordered monotonicity and vice versa; if there is an ordering on  $\mathcal{T}$  that satisfies ordered monotonicity we can find a sequence on  $\mathcal{Z}$  to satisfy OM-Sequence (16).<sup>16</sup> Item (iii) of Theorem 2 provides a similar insight to Item (iii) of Theorem 1, namely that ordered monotonicity can be verified simply by looking at the  $2 \times 2$  submatrices of the response matrix  $\mathbf{R}$ . Item (iv) provides a tractable method for verifying this property.

The final item of the theorem restates the equivalence result of Vytlačil (2006) and shows that assuming ordered monotonicity is equivalent to taking an ordered choice behavioral model. While this result is familiar to the literature, we provide an alternative proof in Appendix A using properties of *lonesum binary matrices*; a concept we borrow from the information theory literature (Ryser, 1957).

### *Symmetries between Ordered and Unordered Monotonicity*

The characterizations of ordered monotonicity in Theorem 2 are symmetric to those of unordered monotonicity in Theorem 1. We have already discussed the usefulness of some of these specific symmetries above. For example, the symmetry between the sequential characterizations of ordered and unordered monotonicities provides an easy way of seeing that ordered and unordered monotonicity

---

<sup>16</sup>In particular we can take the sequence that orders  $z'$  after  $z$  if  $T_i(z') \geq T_i(z)$  for all  $i \in \mathcal{I}$ .

Table 1: Equivalence and Symmetry of Ordered and Unordered Monotonicity Conditions

**Theorem 1** (Unordered Equivalence). *The following statements are equivalent:*

- (i). *For each  $t \in \mathcal{T}$  there is a sequence of instruments  $(z_1^{(t)}, \dots, z_{N_T}^{(t)})$  such that UM Sequence (17) holds.*
- (ii). *Given any  $t \in \mathcal{T}$  and any  $k \in \{1, \dots, N_Z - 1\}$ , we have that*

$$\mathbf{1}[T_i(z_{k+1}^{(t)}) = t] \geq \mathbf{1}[T_i(z_k^{(t)}) = t] \text{ for all } i \in \mathcal{I}..$$

- (iii). *For any  $t \in \mathcal{T}$  and  $t', t'' \neq t$  there are no  $2 \times 2$  submatrices in  $\mathbf{R}$  of the form:*

$$\begin{pmatrix} t & t'' \\ t' & t \end{pmatrix} \text{ or } \begin{pmatrix} t' & t \\ t & t'' \end{pmatrix}. \quad (19)$$

- (iv). *For the unordered verification matrix  $\Psi_U$  defined below,  $\|\Psi_U\| = 0$ ;*

$$\Psi_U \equiv ((\mathbf{1} - \mathbf{U})^T \mathbf{U}) \odot ((\mathbf{1} - \mathbf{U})^T \mathbf{U})^T. \quad (20)$$

where  $\mathbf{1}$  denotes a matrix of all ones,  $\odot$  denotes the Hadamard (element-wise) product and:

$$\mathbf{U} \equiv \begin{bmatrix} \mathbf{B}_1 & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{1} & \mathbf{B}_2 & \mathbf{0} & \dots & \mathbf{0} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \mathbf{1} & \mathbf{1} & \mathbf{1} & \dots & \mathbf{B}_{N_T} \end{bmatrix}. \quad (21)$$

- (v). *For each  $t \in \mathcal{T}$ , there are real-valued functions  $\varphi(\cdot, t)$  and  $\zeta(\cdot, t)$  such that the treatment choice  $T$  can be rationalized by:*

$$\mathbf{1}[T = t] = \mathbf{1}[\zeta(Z, t) \geq \varphi(\mathbf{V}, t)],$$

where  $\zeta(z_{k+1}^{(t)}, t) > \zeta(z_k^{(t)}, t)$  for  $k = 1, \dots, N_Z - 1$  and any  $t$ . □

*Proof.* See Appendix A.

**Theorem 2** (Ordered Equivalence). *The following statements are equivalent:*

- (i). *There is a sequence on  $\mathcal{Z}$ ,  $(z_1, \dots, z_{N_Z})$  and a strict ordering on  $\mathcal{T}$  that satisfies the requirement of OM-Sequence (16).*
- (ii). *There is a strict ordering on  $\mathcal{T}$  such that for any  $k \in \{1, \dots, N_Z - 1\}$  and any  $t$ :*

$$\mathbf{1}[T_i(z_{k+1}) \geq t] \geq \mathbf{1}[T_i(z_k) \geq t] \text{ for all } i \in \mathcal{I}.$$

- (iii). *There is a strict ordering on  $\mathcal{T}$  such that for any  $t < t''$  and  $t' > t'''$  there are no  $2 \times 2$  submatrices of  $\mathbf{R}$  of the form either*

$$\begin{pmatrix} t & t' \\ t'' & t''' \end{pmatrix} \text{ or } \begin{pmatrix} t' & t \\ t''' & t'' \end{pmatrix}; \quad (22)$$

- (iv). *There is a strict ordering on  $\mathcal{T}$  such that for the ordered verification matrix  $\Psi_O$  defined below,  $\|\Psi_O\| = 0$ ;*

$$\Psi_O \equiv ((\mathbf{1} - \mathbf{O})^T \mathbf{O}) \odot ((\mathbf{1} - \mathbf{O})^T \mathbf{O})^T, \quad (23)$$

where  $\mathbf{1}$  indicates a matrix of all ones,  $\odot$  represents the Hadamard (element-wise) product, and:

$$\mathbf{O} \equiv [\mathbf{B}_{t_1}^*, \dots, \mathbf{B}_{t_{N_T}}^*];$$

- (v). *There is a strict ordering on  $\mathcal{T}$  such that for some real-valued functions  $\varphi(\cdot, t)$  and  $\zeta(\cdot, t)$  the treatment choice can be rationalized by*

$$\mathbf{1}[T \geq t] = \mathbf{1}[\zeta(Z, t) \geq \varphi(\mathbf{V}, t)],$$

where  $\zeta(z_{k+1}, t) > \zeta(z_k, t)$  for  $k = 1, \dots, N_Z - 1$  and any  $t$ . □

*Proof.* See Appendix A

are equivalent in the case of a binary treatment. Other symmetries are new to our discussion and are worth briefly mentioning. The symmetric matrix verification characterizations provides an easy way to verify if a response matrix satisfies both the ordered and unordered monotonicity conditions by checking if  $\|\Psi_U\| + \|\Psi_O\| = 0$ . Verifying this allows researchers to take advantage of both sets of identification results.

The symmetry between the restricted  $2 \times 2$  submatrices in ordered and unordered monotonicity provides insight on how a response matrix could satisfy ordered monotonicity but not unordered monotonicity and vice versa. More importantly, however, the similarity between the two restricted submatrices in Theorems 1 and 2 suggests a common condition shared by both criteria. In particular, note that both restricted patterns in (19) and (22) prevent any two agents from having exactly opposite treatment responses for the same instrumental variable shift. We term this common restriction the *Minimal Monotonicity Condition* and analyze its properties in Section 5.

## 5 The Minimal Monotonicity Condition

The minimal monotonicity (MM) condition (24) is a weak criteria shared by both ordered and unordered conditions. It is determined by a symmetric restriction that is common to Theorems 1 and 2. Indeed, it turns out that minimal monotonicity is the core common property of ordered and unordered monotonicity that enables the 2SLS estimand to achieve causal interpretability.

**Minimal Monotonicity (MM):** For any pair of instruments  $z, z' \in \mathcal{Z}$  and any pair of treatments  $t, t' \in \mathcal{T}$  either

$$\begin{aligned} & \mathbf{1}[T_i(z) = t] \mathbf{1}[T_i(z') = t'] \geq \mathbf{1}[T_i(z) = t'] \mathbf{1}[T_i(z') = t] \forall i \in \mathcal{I} \\ \text{or } & \mathbf{1}[T_i(z) = t] \mathbf{1}[T_i(z') = t'] \leq \mathbf{1}[T_i(z) = t'] \mathbf{1}[T_i(z') = t] \forall i \in \mathcal{I}. \end{aligned} \quad (24)$$

The first row in (24) states that an instrumental change from  $z$  to  $z'$  incentivizes *all* agents to shift their choice away from  $t$  and towards  $t'$ . The second row in (24) describes the opposite behavior. In summary, the MM condition states that an intrumental change that induces an agent to switch their choice from  $t$  to  $t'$  cannot induce another another agent to switch their choice from  $t'$  to  $t$ . Lemma 1 provides an equivalent characterization of the MM condition in terms of response-types.

**Lemma 1.** *Minimal monotonicity MM holds if and only if for all distinct instruments  $z, z' \in \mathcal{Z}$  and all distinct treatments  $t, t' \in \mathcal{T}$ , there are no response-types  $\mathbf{s}, \mathbf{s}' \in \text{supp}(\mathbf{S})$  such that*

$$\begin{pmatrix} \mathbf{s}[z] & \mathbf{s}'[z] \\ \mathbf{s}[z'] & \mathbf{s}'[z'] \end{pmatrix} = \begin{pmatrix} \mathbf{s} & \mathbf{s}' \\ t & t' \\ t' & t \end{pmatrix} \begin{matrix} z \\ z' \end{matrix} \quad (25)$$

*Proof.* See Appendix A □

Lemma 1 presents the prohibited pattern of  $2 \times 2$  submatrices of the response matrix  $\mathbf{R}$  induced by MM. The pattern is the common intersection between the submatrix characterizations in item (iii) of Theorems 1 and 2. Lemma 2 establishes that MM is *strictly* weaker than MM and OM.

**Lemma 2.** *The following relationships are true of ordered, unordered, and minimal monotonicity:*

1.  $UM \Rightarrow MM$ , but  $MM \not\Rightarrow UM$
2.  $OM \Rightarrow MM$ , but  $MM \not\Rightarrow OM$

*Proof.* See Appendix A □

To see why minimal monotonicity is crucial for the interpretability of 2SLS, it is useful to quickly define and discuss interpretable causal parameters.

### 5.1 Interpretable Causal Parameters

We follow an established literature that defines a meaningful causal parameter  $\tau$  as a weighted average of local average treatment effects with positive weights:<sup>17</sup>

$$\tau = \sum_{\{t,t'\}, t \neq t'} \omega_{t,t'} E[Y(t) - Y(t') \mid \mathbf{S} \in \mathcal{S}_{t,t'}] \quad \text{with } \omega_{t,t'} = 0 \text{ or } \omega_{t',t} = 0. \quad (26)$$

Here  $\mathcal{S}_{t,t'}$  denotes a set of response types may vary according to the treatments being compared and  $\omega_{t,t'} \geq 0$  are positive weights. The defining idea is that each treatment pair is only represented once, so we cannot have a positive weight on both  $E[Y(t) - Y(t') \mid \mathbf{S} \in \mathcal{S}_{t,t'}]$  and  $E[Y(t') - Y(t) \mid \mathbf{S} \in \mathcal{S}_{t',t}]$ . The absence of negative weights allows this causal parameter to give us meaningful insight into the direction of the treatment effects.

Angrist and Imbens (1995) demonstrate that, under ordered monotonicity, the 2SLS estimand identifies such a meaningful causal parameter using a binary instrument with multiple treatments. Heckman and Pinto (2018) show a similar result for unordered monotonicity using comparisons of the outcome  $Y$  for any two instruments  $z, z' \in \mathcal{Z}$ . The equivalence result for minimal monotonicity in Theorem 3 establishes that it is indeed MM that is the driving force behind both of these identification results.

Lemma 2 provides some intuition for why this is the case. Consider the difference in average outcome between two values of the instrument, as in the numerator of a 2SLS estimand. This difference always has a unique decomposition into a weighted average of treatment effects among all

---

<sup>17</sup>For examples of works that adopt this criteria, see Angrist and Imbens (1995); Heckman and Urzúa (2010); Kirkeboen et al. (2016); Mogstad et al. (2021).

the (ordered) pairs of possible treatment values.<sup>18</sup> The restriction on the response matrix imposed by minimal monotonicity (25) means that if one pair of treatment values,  $(t, t')$ , is represented in this weighted sum, the opposite pair,  $(t', t)$ , cannot also be represented. So, minimal monotonicity is sufficient for this difference to satisfy the condition for an interpretable causal parameter (26). Moreover, because this decomposition is unique, we can show that minimal monotonicity is necessary for interpretability as well.

## 5.2 Equivalence Results

We now provide a set of equivalent characterizations of the minimal monotonicity condition in the spirit of the results for unordered and ordered monotonicity in Theorems 1 and 2.

**Theorem 3** (Minimal Monotonicity Equivalence). *The following statements are equivalent:*

(i). *For any distinct pair of instruments  $z, z' \in \mathcal{Z}$  and any pair of treatments,  $t, t' \in \mathcal{T}$ , we have either:*

$$\begin{aligned} \mathbf{1}[T_i(z) = t]\mathbf{1}[T_i(z') = t'] &\geq \mathbf{1}[T_i(z) = t']\mathbf{1}[T_i(z') = t] \forall i \in \mathcal{I} \\ \text{or } \mathbf{1}[T_i(z) = t]\mathbf{1}[T_i(z') = t'] &\leq \mathbf{1}[T_i(z) = t']\mathbf{1}[T_i(z') = t] \forall i \in \mathcal{I}. \end{aligned} \quad (27)$$

(ii). *There are no  $2 \times 2$  submatrices of  $\mathbf{R}$  of the form in (25).*

(iii). *For the matrix  $\Psi_M$  defined below,  $\|\Psi_M\| = 0$*

$$\Psi_M \equiv \sum_{t \neq t'} (\mathbf{B}_t^\top \mathbf{B}_{t'}) \odot (\mathbf{B}_t^\top \mathbf{B}_{t'})^\top. \quad (28)$$

where  $\odot$  represents the Hadamard (element-wise) multiplication.<sup>19</sup>

(iv). *For any pair of instruments  $z, z'$  the 2SLS type estimand*

$$\beta_{z, z'} = E[Y | Z = z] - E[Y | Z = z']$$

*identifies an interpretable causal parameter as described in (26).*

*Proof.* See Appendix A □

Many features of this equivalence result are symmetric to the unordered and ordered equivalence results of Theorems 1 and 2. Item (i) defines the complete version of the MM condition. Items (ii) and (iii) of Theorem 3 provide ways of verify the MM condition symmetric to counterparts for unordered and ordered monotonicity in Theorems 1 and 2. Item (iv) presents a general response

<sup>18</sup>See Appendix B for a discussion of the exact forms of this decomposition as well as the 2SLS estimands for ordered and unordered monotonicity mentioned above.

<sup>19</sup>We use the short-hand notation  $\sum_{t \neq t'} \xi(t, t') \equiv \sum_{t \in \mathcal{T}} \sum_{t' \in \mathcal{T} \setminus \{t\}} \xi(t, t')$ .

matrix condition. It states that no  $2 \times 2$  submatrix of the response-matrix  $\mathbf{R}$  presents the prohibited pattern in (25). Item (iii) provides a tractable method of verifying the MM condition. The verification requires an order of  $\mathcal{T}^2$  matrix operations.

The last item of Theorem 3 is the empirically relevant feature of the MM condition. It provides a solution to our initial inquiry on a weak monotonicity criteria that ensures interpretable causal parameters for the widely used method of 2SLS. Indeed, there can be no weaker monotonicity criterion that guarantees such causal interpretability.

### 5.3 Relationship Between Monotonicity Criterion

The three monotonicity conditions are equivalent in the case of a binary treatment. In this special case the definition of MM (24) reduces to:<sup>20</sup>

$$\begin{aligned} \mathbf{1}[T_i(z) = t] &\geq \mathbf{1}[T_i(z') = t] \quad \text{for all } i \in \mathcal{I} \\ \text{or } \mathbf{1}[T_i(z) = t] &\leq \mathbf{1}[T_i(z') = t] \quad \text{for all } i \in \mathcal{I}, \end{aligned}$$

which is exactly the requirement imposed by both ordered and unordered monotonicity. However, as demonstrated by Lemma 2, when there are multiple treatments the three monotonicity criterion are distinct.

We can gain further interpretation of the monotonicity restrictions by examining the relation between the verification matrices  $\Psi_U$ ,  $\Psi_O$ , and  $\Psi_M$  of Theorems 1, 2 and 3. We express the verification matrices in terms of a primitive component defined by:

$$\Psi(t, t', t'', t''') \equiv (\mathbf{B}_t^\top \mathbf{B}_{t'}) \odot (\mathbf{B}_{t''}^\top \mathbf{B}_{t'''}). \quad (29)$$

$\Psi(t, t', t'', t''')$  is a function of four binary matrices  $(\mathbf{B}_t, \mathbf{B}_{t'}, \mathbf{B}_{t''}, \mathbf{B}_{t'''})$  that returns a primitive verification matrix of dimension  $N_Z \times N_S$  whose elements are either zeros or natural numbers.<sup>21</sup> Under this notation, the verification matrix  $\Psi_M$  can be expressed as:

$$\Psi_M = \sum_{t \neq t'} \Psi(t, t', t', t). \quad (30)$$

Equation (30) explains the content of the verification matrix  $\Psi_M$ . Theorem 3 states that MM (24) holds if and only if  $\|\Psi_M\| = 0$ . By definition the matrix  $\Psi_M$  is the sum of the primitive verification matrices  $\Psi(t, t', t', t)$  across all  $N_T \cdot (N_T - 1)$  binary combinations of two distinct treatment choices  $t, t' \in \mathcal{T}$ . The elements of the primitive verification matrices are weakly positive and so  $\|\Psi_M\| = 0$  if and only if  $\|\Psi(t, t', t', t)\| = 0$  for all distinct treatment values  $t$  and  $t'$ . Thus a necessary and

<sup>20</sup>This is done by replacing  $\mathbf{1}[T_i(z') = t']$  with  $(1 - \mathbf{1}[T_i(z) = t'])$  on the left hand side and  $\mathbf{1}[T_i(z) = t']$  with  $(1 - \mathbf{1}[T_i(z) = t])$  on the right hand side. Afterwards, distribute and simplify.

<sup>21</sup>The output of function  $\Psi$  remains the same if first two inputs can be switched by the last two inputs  $\Psi(t, t', t'', t''') = \Psi(t'', t''', t, t')$ . We also have that  $\Psi(t, t', t'', t''')^\top = \Psi(t', t, t''', t'')$ , namely, the transpose of  $\Psi(t, t', t'', t''')$  is equal to the matrix  $\Psi(t', t, t''', t'')$  in which we switch  $t$  by  $t'$  and  $t''$  by  $t'''$ .



sufficient condition for MM to hold is that each primitive verification matrix  $\Psi(t, t', t', t)$  contains only zero elements for all  $t, t' \in \mathcal{T}$  such that  $t \neq t'$ . Indeed,  $\|\Psi(t, t', t', t)\| = 0$  is a equivalent to the nonexistence of any  $2 \times 2$  submatrix of the response matrix  $\mathbf{R}$  is of the form  $\begin{pmatrix} t & t' \\ t' & t \end{pmatrix}$ .

Theorem 4 relates the UM verification matrix to the primitive verification matrices and the MM verification matrix.

**Theorem 4** (Decomposing Unordered Verification). *The following relation holds for any response matrix  $\mathbf{R} \in \mathcal{T}^{N_Z \times N_S}$  :*

$$\|\Psi_{\mathbf{U}}\| = 0 \quad \Leftrightarrow \quad \|\Psi_{\mathbf{M}}\| + \|\Psi_{\mathbf{U} \setminus \mathbf{M}}\| = 0, \quad (31)$$

$$\text{where } \Psi_{\mathbf{U} \setminus \mathbf{M}} \equiv \sum_{t \neq t' \neq t''} \Psi(t, t', t'', t). \quad (32)$$

*Proof.* See Appendix A □

Lemma 2 explains that UM imposes extra constraints in addition to those required for MM to hold. Theorem 4 clarifies these additional constraints. Equation (31) decomposes the UM verification ( $\|\Psi_{\mathbf{U}}\| = 0$ ) into two verification requirements. The first verification criterion,  $\|\Psi_{\mathbf{M}}\| = 0$ , means that MM must hold. The additional constraint,  $\|\Psi_{\mathbf{U} \setminus \mathbf{M}}\| = 0$ , means that the elements of the matrix  $\Psi(t, t', t'', t)$  must be zero for any selection of three distinct treatment choices  $t, t', t'' \in \mathcal{T}$ . This additional criterion rules out violations of the prohibited pattern in item (iii) of Theorem 1 that involve three distinct treatment values.

Theorem 4 offers a combinatorial interpretation of monotonicity conditions MM and UM. MM imposes  $\binom{N_T}{2}$  constraints on the primitive verification matrix  $\Psi(t, t', t', t)$  across the combination of treatment choices taken *two* at a time. UM imposes an additional  $\binom{N_T}{3}$  constraints on  $\Psi(t, t', t'', t)$  across the combination of treatment choices taken *three* at a time.

Theorem 5 decomposes the verification criterion of ordered monotonicity ( $\|\Psi_{\mathbf{O}}\| = 0$ ) into the verification criterion of the MM condition,  $\|\Psi_{\mathbf{M}}\| = 0$ , and two other criteria corresponding to matrices  $\Psi_{\mathbf{O} \setminus \mathbf{M}}^{(1)}$  and  $\Psi_{\mathbf{O} \setminus \mathbf{M}}^{(2)}$ . The requirement that  $\|\Psi_{\mathbf{M}}\| = 0$  means that OM implies MM, as in Lemma 2. Matrix  $\Psi_{\mathbf{O} \setminus \mathbf{M}}^{(1)}$  contains a subset of the unordered monotonicity constrains in matrix  $\Psi_{\mathbf{U} \setminus \mathbf{M}}$  of Theorem 4. Matrix  $\Psi_{\mathbf{O} \setminus \mathbf{M}}^{(2)}$  contains constrains that are not in  $\Psi_{\mathbf{U} \setminus \mathbf{M}}$ . This is as expected since OM does not imply nor is implied by UM.

**Theorem 5** (Decomposing Ordered Verification). *The following relation holds for any response matrix  $\mathbf{R} \in \mathcal{T}^{N_Z \times N_S}$  :*

$$\|\Psi_{\mathbf{O}}\| = 0 \quad \Leftrightarrow \quad \|\Psi_{\mathbf{M}}\| + \|\Psi_{\mathbf{O} \setminus \mathbf{M}}^{(1)}\| + \|\Psi_{\mathbf{O} \setminus \mathbf{M}}^{(2)}\| = 0, \quad (33)$$

where

$$\Psi_{\mathbf{O} \setminus \mathbf{M}}^{(1)} \equiv \sum_{t_1 < \min(t_2, t_3)} \Psi(t_1, t_2, t_3, t_1),$$

$$\Psi_{O \setminus M}^{(2)} \equiv \sum_{t_1 < t_2 \leq t_3} \Psi(t_1, t_3, t_2, t_2) + \sum_{t_1 < t_2 < t_3} \Psi(t_1, t_3, t_3, t_2) + \sum_{t_4 < t_2, t_1 < t_3} \Psi(t_1, t_2, t_3, t_4).$$

*Proof.* See Appendix A □

## 6 An Economic Interpretation for Monotonicity Conditions

This section explores the economic content of the monotonicity criteria. We show that the minimal monotonicity condition (24) can be linked to a broad notion of rationality regarding treatment choices. This contrasts to ordered and unordered monotonicity, which as seen in Theorems 1 and 2 are equivalent to assuming particular ordered and unordered choice models.

Our analysis is based on the method of Pinto (2021); Buchinsky and Pinto (2021) who use revealed preference analysis to ascribe economic interpretation to response matrices.<sup>22</sup> The method uses the concept of an incentive matrix  $\mathbf{L}$  that characterizes the choice incentives (columns) generated by the IV-values (rows). Each column  $\mathbf{L}[\cdot, t]$  displays the relative ranking of incentives towards choice  $t \in \text{supp}(t)$  across the IV-values  $z \in \mathcal{Z}$ .  $\mathbf{L}[z', t] < \mathbf{L}[z, t]$  means that the IV-value  $z$  yields strictly greater incentives towards  $t$  than IV-value  $z'$ . The matrix is ordinal, monotonic transformations characterize equivalent choice incentives.

To fix ideas, consider the binary LATE model of Sections 4 where  $T$  denotes college enrollment;  $T = t_1$  for college enrollment and  $T_i = t_0$  otherwise. The instrument  $Z$  denotes a randomly assigned tuition discount such that  $Z = z_1$  if the discount is granted and  $Z = z_0$  otherwise. Incentive matrix (34) characterizes the choice incentives of the LATE model.  $\mathbf{L}[z_0, t_0] = \mathbf{L}[z_1, t_0] = 0$  means that the voucher offers no incentives for choice  $t_0$  (no college).  $\mathbf{L}[z_0, t_1] < \mathbf{L}[z_1, t_1]$  indicates that the tuition discount  $z_1$  incentives college enrollment  $t_1$ .

$$\text{LATE Incentive Matrix } \mathbf{L} = \begin{matrix} & \begin{matrix} t_0 & t_1 \end{matrix} \\ \begin{matrix} z_0 \\ z_1 \end{matrix} & \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \end{matrix} \quad (34)$$

Pinto (2021); Buchinsky and Pinto (2021) use revealed preference analysis to translate the incentive matrix into choice restrictions. They invoke the Weak Axiom of Revealed Preferences (WARP) and Choice Normality to generate the following choice rule:<sup>23</sup>

$$\text{Choice Rule: } T_i(z) = t \text{ and } \underbrace{\mathbf{L}[z', t'] - \mathbf{L}[z, t'] \leq \mathbf{L}[z', t] - \mathbf{L}[z, t]}_{\substack{\text{Switch from } z \text{ to } z' \text{ provides greater} \\ \text{incentives for } t \text{ than } t'}} \implies T_i(z') \neq t' \quad (35)$$

<sup>22</sup>This analysis is also similar to the encouragement designs proposed by Zelen (1979, 1990).

<sup>23</sup>[Define both] The choice rule would have a strict inequality if we were to assume WARP only.

Choice Rule (35) formalizes an intuitive behavioral restriction. If an agent  $i$  chooses  $t$  when exposed to instrument  $z$ , and the IV-shift from  $z$  to  $z'$  yields greater incentives towards  $t$  than  $t'$ , then agent  $i$  does not choose  $t'$  under  $z'$ . Otherwise stated, each instrumental value  $z$  is associated with an incentive gap between  $t$  and  $t'$ . If an agent decides for choice  $t$  given  $z$ , then  $t$  is revealed preferred to  $t'$ . The agent should  $t'$  only if the incentive gap between  $t$  and  $t'$  increases.

Applying choice rule (35) to the LATE incentive matrix (34) generates the following *choice restriction*:

$$T_i(z_0) = t_1 \quad \text{and} \quad \underbrace{\mathbf{L}[z_1, t_0] - \mathbf{L}[z_0, t_0]}_{=0} \leq \underbrace{\mathbf{L}[z_1, t_1] - \mathbf{L}[z_0, t_1]}_{=1} \implies T_i(z_1) \neq t_0. \quad (36)$$

Choice restriction (36) is summarized by  $T_i(z_0) = t_1 \implies T_i(z_1) \neq t_0$ . It states that if an agent chooses to attend college when not offered any incentives ( $T_i(z_0) = t_1$ ) they must also choose to attend college when offered the tuition discount ( $T_i(z_1) = t_1$ ). The restriction is equivalent to the monotonicity condition of Imbens and Angrist (1994), discussed in Section 3, which eliminates the defiers and enables the identification of LATE.

Buchinsky and Pinto (2021) characterize the class of incentive matrices that produce OM and UM conditions. For instance, they demonstrate that choice incentives characterized by lonesum matrices produce unordered choice models while increasing incentives such as those described by a Vandermonde matrix produce ordered choice models.

#### *Incentives and Minimal Monotonicity*

It is natural to inquire about which types of incentive designs ensure the MM condition. It turns out that the Choice Rule (35) itself ensures the MM condition. Theorem 6 asserts that the MM condition always arises whenever we apply the revealed preference analysis encoded by the choice rule to any choice incentives.

**Theorem 6.** *The minimal monotonicity condition (24) holds for all choice models generated by applying Choice Rule (35) to an arbitrary Incentive Matrix  $\mathbf{L}$ .*

Theorem 6 draws a sharp distinction in the interpretation of the monotonicity conditions. In essence, OM and UM can be interpreted as monotonicity conditions that arise when agents that display a rational behavior face a particular a class of choice incentives. This paradigm does not apply to MM, since MM is not a property ascribed to any particular pattern of incentives. Instead, MM is a supra-condition that can be justified by a weak notion of rationality itself.

With this in mind, the MM condition can be seen not as a final goal, but rather a starting point for generating and interpreting monotonicity criteria. To be more precise, minimal monotonicity ensures that a range of monotonicity conditions obtained by combining a broad notion of choice rationality with specific choice incentives satisfy particular basic properties, including interpretability of 2SLS-type estimands. Section 7 uses this insight to illustrate the flexibility of the MM condition

in empirical analysis.

## 7 Economic Examples of Monotonicity Conditions

We present several examples of response matrices generated by combining specific incentive structures with the choice rule (35) defined in Section 6. The first two examples demonstrate specific incentive designs that generate response matrices satisfying unordered and ordered monotonicity, respectively. Afterwards, we present some natural research designs that generate response matrices that do not comply with either ordered or unordered monotonicity. However, by Theorem 6, these response matrices still satisfy the minimal monotonicity condition. We discuss how minimal monotonicity may still enable the researcher to gain insight into causal effects under these research designs. Our minimal monotonicity examples include the popular Extensive Margin Compliers (EMCO) of Angrist and Imbens (1995) and a double RCT design. In all examples we consider incentive designs for a three-valued treatment choice  $\mathcal{T} = \{t_1, t_2, t_3\}$  and four instrumental values  $\mathcal{Z} = \{z_1, z_2, z_3, z_4\}$ .

### 7.1 A Case of Choice Incentives that Justify Unordered Monotonicity

In this example, let  $T$  denote the student's decision among college majors:  $t_1$  for humanities,  $t_2$  for social sciences, and  $t_3$  for the STEM fields of science, technology, engineering, and math. The instrumental variable  $Z$  represents a randomly assigned vouchers that offers a tuition discount that may apply to one, several or none of the majors. For example, consider the social experiment that randomly assigns one of the four vouchers  $z_1, z_2, z_3, z_4$  to college students:

1. Voucher  $z_1$  offers no tuition discount.
2. Voucher  $z_2$  applies only to STEM ( $t_3$ ).
3. Voucher  $z_3$  applies to either STEM ( $t_3$ ) or social sciences ( $t_2$ ).
4. Voucher  $z_4$  applies to all majors.

Assuming that the voucher amount is always the same, this design is characterized by incentive matrix  $\mathbf{L}$  in (37).<sup>24</sup>

$$\mathbf{L} = \begin{array}{ccc} & t_1 & t_2 & t_3 \\ \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} & z_1 & z_2 & z_3 & z_4 \end{array} \quad (37)$$

We use this first example to describe the machinery that translates choice incentives into monotonicity conditions and identification results. We adopt a more parsimonious approach in the

---

<sup>24</sup>Elements one indicate the presence of incentive (the tuition discount) while elements zero indicate the lack of it.

subsequent examples. We place detailed derivations in Appendix D.

Choice rule (35) converts the Incentive Matrix (37) into choice restrictions that determine the model response matrix  $\mathbf{R}$ . The choice rule applies to any two instrument-treatment pairs;  $((z, t), (z', t')) \in (\mathcal{Z} \times \mathcal{T})^2$ . To exemplify how this is done, Table 2 displays all the restrictions generated by applying Choice Rule (35) to an agent who chooses a humanities major ( $t_1$ ) when offered no tuition discount ( $z_1$ ). The first row of Table 2 applies the choice rule to the two instrument-treatment pairs,  $(z_1, t_1)$  and  $(z_2, t_2)$ . By applying the choice rule, we see that an agent  $i$  who chooses a humanities major ( $t_1$ ) when offered no tuition discount ( $z_1$ ) would not chose treatment a social sciences major ( $t_2$ ) when offered instrument a tuition discount only for STEM majors ( $z_2$ ). The incentives for choosing either  $t_1$  or  $t_2$  remain the same when the IV switches from  $z_1$  to  $z_2$ . The incentive inequality in (35) is satisfied and the choice restriction  $T_i(z_1) \neq t_2$  holds.

The other lines of Table 2 apply this same logic to all other instrument-treatment pairs. We can see that, in total, under the incentive structure summarized in (37) the Choice Rule (35) places the following restrictions on an agent who chooses a humanities major when offered no tuition discount

$$T_i(z_1) = t_1 \implies T_i(z_2) \neq t_2 \text{ and } T_i(z_4) \notin \{t_2, t_3\}$$

This analysis can be repeated for all types of agents, the sum total of all choice restrictions generated by applying the choice rule in this research design are presented in Table 3. All choice restrictions of Table 3 eliminate a total of 74 out of 81 possible response-types. The seven response-types that survive this elimination procedure are presented as columns of the following response matrix,  $\mathbf{R}$ , in (38):

Table 2: Applying Choice Rule (35) to  $T_i(z_1) = t_1$  and Incentive Matrix (37)

Counterfactual Choice	Incentive Condition			Choice Restriction
$T(z_1) = t_1$	$\mathbf{L}[z_2, t_2] - \mathbf{L}[z_1, t_2]$	$= 0 \leq 0 =$	$\mathbf{L}[z_2, t_1] - \mathbf{L}[z_1, t_1]$	$\Rightarrow T(z_2) \neq t_2$
$T(z_1) = t_1$	$\mathbf{L}[z_2, t_3] - \mathbf{L}[z_1, t_3]$	$= 1 \not\leq 0 =$	$\mathbf{L}[z_2, t_1] - \mathbf{L}[z_1, t_1]$	$\Rightarrow$ No Restriction
$T(z_1) = t_1$	$\mathbf{L}[z_3, t_2] - \mathbf{L}[z_1, t_2]$	$= 1 \not\leq 0 =$	$\mathbf{L}[z_3, t_1] - \mathbf{L}[z_1, t_1]$	$\Rightarrow$ No Restriction
$T(z_1) = t_1$	$\mathbf{L}[z_3, t_3] - \mathbf{L}[z_1, t_3]$	$= 1 \not\leq 0 =$	$\mathbf{L}[z_3, t_1] - \mathbf{L}[z_1, t_1]$	$\Rightarrow$ No Restriction
$T(z_1) = t_1$	$\mathbf{L}[z_4, t_2] - \mathbf{L}[z_1, t_2]$	$= 1 \leq 1 =$	$\mathbf{L}[z_4, t_1] - \mathbf{L}[z_1, t_1]$	$\Rightarrow T(z_4) \neq t_2$
$T(z_1) = t_1$	$\mathbf{L}[z_4, t_3] - \mathbf{L}[z_1, t_3]$	$= 1 \leq 1 =$	$\mathbf{L}[z_4, t_1] - \mathbf{L}[z_1, t_1]$	$\Rightarrow T(z_4) \neq t_3$

This table presents all the choice restrictions generated by applying the choice rule (35) to each of the tuples  $((z_1, t_1), (z', t'))$  where  $z' \in \{z_2, z_3, z_4\}$  and  $t' \in \{t_2, t_3, t_4\}$  according to the choice incentives displayed in the incentive matrix (37).

$$\mathbf{R} = \begin{matrix} & \mathbf{s}_1 & \mathbf{s}_2 & \mathbf{s}_3 & \mathbf{s}_4 & \mathbf{s}_5 & \mathbf{s}_6 & \mathbf{s}_7 & \\ \begin{bmatrix} t_1 & t_1 & t_1 & t_1 & t_2 & t_2 & t_3 \\ t_1 & t_1 & t_3 & t_3 & t_2 & t_3 & t_3 \\ t_1 & t_2 & t_2 & t_3 & t_2 & t_2 & t_3 \\ t_1 & t_1 & t_1 & t_1 & t_2 & t_2 & t_3 \end{bmatrix} & z_1 & z_2 & z_3 & z_4 & & & & \end{matrix} \quad (38)$$

Table 3: Choice Restrictions generated by Incentive Matrix (37)

1	$T_i(z_1) = t_1 \Rightarrow T_i(z_2) \neq t_2 \text{ and } T_i(z_4) \notin \{t_2, t_3\}$
2	$T_i(z_2) = t_1 \Rightarrow T_i(z_1) \notin \{t_2, t_3\} \text{ and } T_i(z_3) \neq t_3 \text{ and } T_i(z_4) \notin \{t_2, t_3\}$
3	$T_i(z_3) = t_1 \Rightarrow T_i(z_1) \notin \{t_2, t_3\} \text{ and } T_i(z_2) \notin \{t_2, t_3\} \text{ and } T_i(z_4) \notin \{t_2, t_3\}$
4	$T_i(z_4) = t_1 \Rightarrow T_i(z_1) \notin \{t_2, t_3\} \text{ and } T_i(z_2) \neq t_2$
5	$T_i(z_1) = t_2 \Rightarrow T_i(z_2) \neq t_1 \text{ and } T_i(z_3) \notin \{t_1, t_3\} \text{ and } T_i(z_4) \notin \{t_1, t_3\}$
6	$T_i(z_2) = t_2 \Rightarrow T_i(z_1) \notin \{t_1, t_3\} \text{ and } T_i(z_3) \notin \{t_1, t_3\} \text{ and } T_i(z_4) \notin \{t_1, t_3\}$
7	$T_i(z_3) = t_2 \Rightarrow T_i(z_1) \neq t_3 \text{ and } T_i(z_4) \neq t_3$
8	$T_i(z_4) = t_2 \Rightarrow T_i(z_1) \notin \{t_1, t_3\} \text{ and } T_i(z_2) \neq t_1 \text{ and } T_i(z_3) \notin \{t_1, t_3\}$
9	$T_i(z_1) = t_3 \Rightarrow T_i(z_2) \notin \{t_1, t_2\} \text{ and } T_i(z_3) \notin \{t_1, t_2\} \text{ and } T_i(z_4) \notin \{t_1, t_2\}$
10	$T_i(z_2) = t_3 \Rightarrow T_i(z_3) \neq t_1$
11	$T_i(z_3) = t_3 \Rightarrow T_i(z_1) \neq t_2 \text{ and } T_i(z_2) \notin \{t_1, t_2\} \text{ and } T_i(z_4) \neq t_2$
12	$T_i(z_4) = t_3 \Rightarrow T_i(z_1) \notin \{t_1, t_2\} \text{ and } T_i(z_2) \notin \{t_1, t_2\} \text{ and } T_i(z_3) \notin \{t_1, t_2\}$

This table presents all the choice restrictions generated by applying the choice rule (35) to each of the tuples  $((z, t), (z', t')) \in (\{t_1, t_2, t_3\} \times \{z_1, z_2, z_3, z_4\})^2$ . according to the incentive matrix (37).

Using the characterizations in Theorem 1, we can verify that unordered monotonicity holds for the response matrix  $\mathbf{R}$  presented in (38). Notice that there is no  $2 \times 2$  matrix of the matrix is of the type of the form  $\begin{pmatrix} t & t' \\ t'' & t \end{pmatrix}$  where  $t' \neq t$ , or  $t'' \neq t$ . Appendix D.1 corroborates the UM property using the verification matrix from item (iv) of Theorem 1.

While the incentive design described in (37) is plausible and the application of the choice rule a minimal behavioral requirement on the agents, it is helpful to remark on how this specific incentive structure generates unordered monotonicity. Note that the incentive structure increases in the sense that each successive instrument provides weakly more incentives for all treatment choices. No change in the instrument from  $z$  to  $z'$  would strictly decrease incentives for one treatment while strictly increasing incentives for another.

This property is crucial for generating unordered monotonicity. Under WARP, if an agent would choose treatment  $t$  under instrument value  $z$ , they must also choose treatment  $t$  under instrument value  $z'$  whenever the switch from  $z$  to  $z'$  weakly increases incentives for  $t$  relative to all other incentives. Because of the increasing nature of the incentive structure, each switch in the instrument value either increases the incentives for a choice  $t$  relative to *all* other treatments or decreases the incentives for the choice  $t$  relative to all other treatments. This, along with the choice rule, prevents an instrumental switch from moving one agent strictly towards choosing  $t$  while moving another agent strictly away from choice  $t$  and towards another treatment  $t'$ .

## 7.2 A Case of Choice Incentives that Justify Ordered Monotonicity

In this example, suppose the CEO of a company is interested in whether higher health insurance premiums lead to moral hazard in employees' safety behavior. Employees decide among three health insurance policies  $t_1, t_2, t_3$  that have increasing premiums. The co-pay of each policy off-sets the increasing premium such that all policies cost the same.

To study this, the CEO randomly assigns agents to one of four groups,  $z_1, z_2, z_3, z_4$ , that incentivize (say by offering an additional week of vacation) various insurance plan options. We consider the following scheme of choice incentives:

1. Group  $z_1$  is incentivized to choose treatment  $t_1$ .
2. Group  $z_2$  is offered no incentives.
3. Group  $z_3$  is offered incentives for all choices.
4. Group  $z_4$  is incentivized to choose treatment  $t_3$ .

Equation (39) presents the incentive matrix  $\mathbf{L}$  that characterizes the design of choice incentives.<sup>25</sup> This incentive design is rather peculiar because it is tailored to generate the OM criteria. Equation (39) also presents the corresponding response matrix  $\mathbf{R}$  generated by the method of revealed preference analysis described in Sections 6 and 7.1. Detailed derivations are presented in Appendix D.2.

$$\mathbf{L} = \begin{array}{ccc|c} t_1 & t_2 & t_3 & \\ \hline 1 & 0 & 0 & z_1 \\ 0 & 0 & 0 & z_2 \\ 1 & 1 & 1 & z_3 \\ 0 & 0 & 1 & z_4 \end{array} \Rightarrow \mathbf{R} = \begin{array}{cccccccc|c} \mathbf{s}_1 & \mathbf{s}_2 & \mathbf{s}_3 & \mathbf{s}_4 & \mathbf{s}_5 & \mathbf{s}_6 & \mathbf{s}_7 & \mathbf{s}_8 & \\ \hline t_1 & t_1 & t_1 & t_1 & t_1 & t_2 & t_2 & t_3 & z_1 \\ t_1 & t_1 & t_2 & t_2 & t_3 & t_2 & t_2 & t_3 & z_2 \\ t_1 & t_1 & t_2 & t_2 & t_3 & t_2 & t_2 & t_3 & z_3 \\ t_1 & t_3 & t_2 & t_3 & t_3 & t_2 & t_3 & t_3 & z_4 \end{array} \quad (39)$$

Using Theorem 2 it is easy to check that OM holds for response matrix (39). The indices of the treatment choices weakly increase as the instrument ranges along  $z_1 \rightarrow z_2 \rightarrow z_3 \rightarrow z_4$ . Ordered monotonicity is satisfied by assigning treatment values that satisfy  $t_1 \leq t_2 \leq t_3$ . By applying the identification results of Angrist and Imbens (1995), one can verify that the 2SLS has the causal interpretation of a weighted average of LATEs of the type  $E(Y(t_{k+1}) - Y(t_k)|\mathbf{S})$ ;  $k \in \{1, 2\}$ . Similarly via Theorem 1 it is also easy to verify that UM does not hold for response matrix (39). The  $2 \times 2$  submatrix of rows  $(z_1, z_4)$  and columns  $(\mathbf{s}_3, \mathbf{s}_7)$  displays the values  $\begin{pmatrix} t_1 & t_2 \\ t_2 & t_3 \end{pmatrix}$  which violates item (iii) of Theorem 1; the shift of IV-values from  $z_1$  to  $z_4$  induces some agents towards choice  $t_2$  while inducing others away from  $t_2$ .

Interestingly, the natural ranking on the treatment space is not important for generating ordered

<sup>25</sup>Elements one indicate the presence of incentive (an additional one week vacation) while elements zero indicate the lack of it.

monotonicity in this way. We could just as easily have considered a treatment space with no natural ranking, such as the choice of neighborhood to live in. The incentive design summarized by  $\mathbf{L}$  in (39) would have still generated the response matrix  $\mathbf{R}$  that satisfies ordered monotonicity. This demonstrates the usefulness of considering the slightly broader characterization of ordered monotonicity presented in OM Sequence (16). Had ordered monotonicity been ruled out a priori, the researcher may not have been able to take advantage of the Angrist and Imbens (1995) identification results.

### 7.3 Beyond Ordered or Unordered Monotonicity

The MM condition provides a theoretical foundation for a wide range choice behaviors that do not exactly conform to the paradigm imposed by ordered or unordered choices. It offers the necessary flexibility to examine economic settings that are not neatly described by ordered or unordered monotonicity. We illustrate this fact in the following examples.

#### *The Double Randomization Design*

A basic inquiry in social science is to evaluate the causal effect of a treatment  $t_1$  versus its absence. The standard IV experiment that would allow us to assess this effect is to randomly offer a voucher that incentivizes a set of agents to choose a treatment choice  $t_1$ . This experiment can be described by the binary LATE model discussed in Section 3.

A straightforward extension of this setup is to insert a second treatment  $t_2$  and randomize a second voucher that incentivizes  $t_2$  for the same set of agents. The combination of the two randomization runs generate four groups according to the voucher assignments. Notationally, our experiment consists of three choices  $T \in \{t_0, t_1, t_2\}$ , where  $t_0$  represents not choosing either treatment  $t_1$  or  $t_2$ , and four instrumental values  $\{z_1, z_2, z_3, z_4\}$  that classify the voucher recipients into:

1. Group  $z_1$  comprise agents that do not receive any voucher.
2. Group  $z_2$  comprise agents that receive a voucher that incentivizes choice ( $t_2$ ).
3. Group  $z_3$  comprise agents that receive a voucher that incentivizes choice ( $t_1$ ).
4. Group  $z_4$  are those that receive two vouchers, one for  $t_1$  and another for  $t_2$ .

Equation (40) presents an incentive matrix corresponding to this research design and the corresponding response matrix generated by the revealed preference analysis described in Section 7.1. See Appendix D.3 for detailed derivations.

$$\mathbf{L} = \begin{matrix} & t_0 & t_1 & t_2 \\ \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix} & z_1 \\ & z_2 \\ & z_3 \\ & z_4 \end{matrix} \Rightarrow \mathbf{R} = \begin{matrix} \mathbf{s}_1 & \mathbf{s}_2 & \mathbf{s}_3 & \mathbf{s}_4 & \mathbf{s}_5 & \mathbf{s}_6 & \mathbf{s}_7 & \mathbf{s}_8 & \mathbf{s}_9 \\ \begin{bmatrix} t_0 & t_0 & t_0 & t_0 & t_0 & t_1 & t_1 & t_2 & t_2 \\ t_0 & t_0 & t_2 & t_2 & t_2 & t_1 & t_2 & t_2 & t_2 \\ t_0 & t_1 & t_0 & t_1 & t_1 & t_1 & t_1 & t_1 & t_2 \\ t_0 & t_1 & t_2 & t_1 & t_2 & t_1 & t_1 & t_2 & t_2 \end{bmatrix} & z_1 \\ & z_2 \\ & z_3 \\ & z_4 \end{matrix} \quad (40)$$



Response matrix (40) satisfies neither UM nor OM. UM does not hold because the  $2 \times 2$  submatrix generated by rows  $(z_2, z_3)$  and columns  $(\mathbf{s}_2, \mathbf{s}_3)$  displays matrix  $\begin{pmatrix} t_0 & t_2 \\ t_1 & t_0 \end{pmatrix}$ , which violates item (iii) of Theorem 1. For the ordering  $t_0 \leq t_1 \leq t_2$  on  $\mathcal{T}$ , columns  $(\mathbf{s}_2, \mathbf{s}_3)$  also preclude OM. Under this ordering, no matter how we order the IV values, we cannot generate that the sequence of treatments is increasing for both  $\mathbf{s}_2$  and  $\mathbf{s}_3$ . In fact, no matter which ordering we take on  $\mathcal{T}$ , we can always find a pair of response types whose treatment uptake patterns violate OM Sequence.<sup>26</sup>

However, as guaranteed by Theorem 6, the response matrix  $\mathbf{R}$  in (40) satisfies minimal monotonicity (MM). This can be further verified via the minimal monotonicity characterization in Theorem 3 by checking that there is no  $2 \times 2$  submatrix in (40) of the form  $\begin{pmatrix} t & t' \\ t' & t \end{pmatrix}$ . Appendix D.3 additionally corroborates this fact by evaluating the verification matrix  $\Psi_M$  from Theorem 3(iii).

Despite the fact that neither UM nor OM holds, we can still explore causal relationships in this choice model. In particular, the response matrix still enables us to identify causal parameters using 2SLS type estimands. For example, if the researcher was interested in the effect of treatment  $t_2$  against alternate treatments, the 2SLS-type estimand

$$\beta_{z_4, z_3} \equiv E[Y | Z = z_4] - E[Y | Z = z_3].$$

recovers a weighted average of  $E[Y(t_2) - Y(t_1) | \mathbf{S} \in \mathcal{S}_{2,1}]$  and  $E[Y(t_2) - Y(t_0) | \mathbf{S} \in \mathcal{S}_{2,0}]$ , with positive weights, for some sets of response types  $\mathcal{S}_{2,1}$  and  $\mathcal{S}_{2,0}$  that can be found using the decomposition in Appendix B. If one was alternatively interested in the effect of  $t_1$  against alternate treatments, the 2SLS-type estimand

$$\beta_{z_3, z_1} \equiv E[Y | Z = z_3] - E[Y | Z = z_2]$$

recovers a weighted average of  $E[Y(t_1) - Y(t_0) | \mathbf{S} \in \mathcal{S}_{1,0}]$  and  $E[Y(t_1) - Y(t_2) | \mathbf{S} \in \mathcal{S}_{1,2}]$  for two alternate sets of response types  $\mathcal{S}_{1,0}$  and  $\mathcal{S}_{1,2}$ .

#### *Incentives that Justify Extensive Margin Compliers Only*

We next exemplify how the choice rationale can be used to justify monotonicity criteria that are more restrictive than UM and OM. Consider a group of students of a technical college that decide among three possible majors: computer science ( $t_1$ ), electrical engineering ( $t_2$ ), or mechanical engineering ( $t_3$ ).

College administration perform a double randomization of two types of tuition vouchers. The first voucher offers a tuition discount for computer science ( $t_1$ ) while the other for the engineering courses ( $t_2$  or  $t_3$ ). Students can be divided into four groups according to the voucher assignment:

1. Group  $z_1$  receives no voucher;

---

<sup>26</sup>Using the violation of unordered monotonicity with  $t_0$  we have that  $t_0$  cannot be ranked highest or lowest in any ordering that satisfies OM; this would mean that some agents increase their treatment as the instrument ranges from  $z_2$  to  $z_3$  while other agents decrease their treatment. We thus only have to consider the orderings  $t_1 \leq t_0 \leq t_2$  and  $t_2 \leq t_0 \leq t_1$ , which can both be eliminated by considering the alternating patterns displayed by response types  $\mathbf{s}_4$  and  $\mathbf{s}_7$ .

2. Group  $z_2$  receives the voucher for computer science ( $t_1$ ) only;
3. Group  $z_3$  receives the voucher that incentivizes electrical ( $t_2$ ) or mechanical ( $t_3$ ) engineering.
4. Group  $z_4$  receives both vouchers which offers incentives to all three choices.

Equation (41) presents the incentive matrix associated with this experimental design and its corresponding response matrix. See Appendix D.4 for derivation details.

$$\mathbf{L} = \begin{array}{ccc|c} t_1 & t_2 & t_3 & \\ \hline 0 & 0 & 0 & z_1 \\ 1 & 0 & 0 & z_2 \\ 0 & 1 & 1 & z_3 \\ 1 & 1 & 1 & z_4 \end{array} \Rightarrow \mathbf{R} = \begin{array}{cccccc|c} \mathbf{s}_1 & \mathbf{s}_2 & \mathbf{s}_3 & \mathbf{s}_4 & \mathbf{s}_5 & \mathbf{s}_6 & \mathbf{s}_7 & \\ \hline t_1 & t_1 & t_1 & t_2 & t_2 & t_3 & t_3 & z_1 \\ t_1 & t_1 & t_1 & t_1 & t_2 & t_1 & t_3 & z_2 \\ t_1 & t_2 & t_3 & t_2 & t_2 & t_3 & t_3 & z_3 \\ t_1 & t_1 & t_1 & t_2 & t_2 & t_3 & t_3 & z_4 \end{array} \quad (41)$$

Response matrix (41) is an example where both OM and UM are satisfied. We can check that OM holds by assigning values (1, 2, 3) to ( $t_1, t_2, t_3$ ) and reordering the IV-values from  $z_1, z_2, z_3, z_4$  to  $z_2, z_1, z_4, z_3$ . The resulting response matrix is presented in (42) which shows that treatment values weakly increase as  $Z$  ranges along its values. It is easy to check that each of the binary matrices  $\mathbf{B}_t = \mathbf{1}[\mathbf{R} = t]$ ;  $t \in \{1, 2, 3\}$  that indicate the treatment choices of response matrix (42) is triangular (i.e. lonesum). This implies that UM holds by item (iv) of Theorem 1.

$$\text{Reordered } \mathbf{R} = \begin{array}{cccccc|c} \mathbf{s}_1 & \mathbf{s}_2 & \mathbf{s}_3 & \mathbf{s}_4 & \mathbf{s}_5 & \mathbf{s}_6 & \mathbf{s}_7 & \\ \hline 1 & 1 & 1 & 1 & 2 & 1 & 3 & z_2 \\ 1 & 1 & 1 & 2 & 2 & 3 & 3 & z_1 \\ 1 & 1 & 1 & 2 & 2 & 3 & 3 & z_4 \\ 1 & 2 & 3 & 2 & 2 & 3 & 3 & z_3 \end{array} \quad (42)$$

Response matrix (41) has a special property beyond UM and OM: each of its compliers takes only two treatment values, one of them being  $t_1$ . Specifically, the matrix has four response-types that display a variation of treatment choice, these are the compliers ( $\mathbf{s}_2, \mathbf{s}_3, \mathbf{s}_4, \mathbf{s}_6$ ). The choice values of response-types  $\mathbf{s}_2, \mathbf{s}_4$  are  $t_1$  or  $t_2$  and the choice values of  $\mathbf{s}_3, \mathbf{s}_6$  are  $t_1$  or  $t_3$ . This special property is called Extensive Margin Compliers Only (EMCO) which is formalized in (43).<sup>27</sup>

**EMCO:** There exists a treatment choice  $t_1 \in \mathcal{T}$  such that for any  $z, z' \in \text{supp}(Z)$  we have that

$$T_i(z) \neq T_i(z') \Rightarrow T_i(z) = t_1 \text{ or } T_i(z') = t_1 \text{ for all } i \in \mathcal{I} \quad (43)$$

EMCO (43) simplifies the multiple-choice decision of compliers into a binary decision that debates between choosing  $t_1$  or not. In our example, compliers  $\mathbf{s}_2, \mathbf{s}_4$  debate between choosing computer science  $t_1$  or electrical engineering  $t_2$  while compliers  $\mathbf{s}_3, \mathbf{s}_6$  debate between computer science  $t_1$  or

<sup>27</sup>See Rose and Shem-Tov (2021); Angrist and Imbens (1995).

mechanical engineering  $t_3$ . None of the compliers debate between electrical or mechanical engineering. Instead, they decide between choosing computer science or not.

EMCO (43) enables us to recode the multiple choice  $T_i \in \{t_1, t_2, t_3\}$  into a binary choice  $D_i = \mathbf{1}[T_i = t_1]$  that indicates if the agent  $i$  chooses  $t_1$ . The 2SLS regression that uses the binary indicator as the endogenous treatment evaluates a weighted average of LATE-type effects between choosing  $t_1$  and not across compliers.

In particular, the comparison between two IV-values identifies the causal effect for of choosing  $t_1$  versus not choosing  $t_1$  for a sub-set of compliers. For instance, consider the IV-values  $z_1$  and  $z_2$ . We can use equation (9) to obtain the following identification result:

$$\frac{E(Y|Z = z_2) - E(Y|Z = z_1)}{P(T = t_2|Z = z_2) - P(T = t_2|Z = z_1)} = \quad (44)$$

$$\frac{E(Y(t_1) - Y(t_2)|\mathbf{S} = \mathbf{s}_4)P(\mathbf{S} = \mathbf{s}_4) + E(Y(t_1) - Y(t_3)|\mathbf{S} = \mathbf{s}_6)P(\mathbf{S} = \mathbf{s}_6)}{P(\mathbf{S} = \mathbf{s}_4) + P(\mathbf{S} = \mathbf{s}_6)}. \quad (45)$$

Equations (44)–(45) show that the comparison between IV-values  $z_1$  and  $z_2$  identifies the causal effect of choosing  $t_1$  versus not choosing  $t_1$  conditional on response-types  $\mathbf{s}_4, \mathbf{s}_6$ . The equations are similar to the LATE identification equation of Imbens and Angrist (1994). They imply that we can evaluate the causal effect via the 2SLS regression that uses the sub-sample of agents assigned to  $z_1$  and  $z_2$ .

### *Orthogonal Array Design*

We additionally examine an IV choice model based on the orthogonal array experimental design. Orthogonal arrays are a widely popular experimental design developed by C.D. Rao (Rao, 1946a,b, 1947, 1949). Orthogonal arrays are widely used in Agricultural and Industrial sciences to determine the optimum mix of treatments that maximize production yield. The method is based on the random assignment of a combinatorial arrangements of treatments for each randomization arm. We adapt this setup to an instrumental variable setting by exogenously providing incentives for one or more treatments instead of directly assigning agents to treatment arms. Below, we will see that this incentive structure allows for a broad range of identification results.

Formally, a binary orthogonal array is a matrix of zeros and ones such that any two-column sub-matrix displays all possible combinations of zeros and ones. In other words, the tuples

$$\{(0, 0), (0, 1), (1, 0), (1, 1)\}$$

are all rows in any two-column submatrix. We say an incentive design is an *orthogonal array incentive design* if its associated incentive matrix is a binary orthogonal array. The incentive matrix in (46) displays an example of an orthogonal array incentive design and the corresponding response matrix generated by applying Choice Rule (35).

$$\mathbf{L} = \begin{matrix} & t_1 & t_2 & t_3 \\ \begin{bmatrix} 0 & 1 & 1 \\ 0 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} & z_1 \\ & z_2 \\ & z_3 \\ & z_4 \end{matrix} \Rightarrow \mathbf{R} = \begin{matrix} & \mathbf{s}_1 & \mathbf{s}_2 & \mathbf{s}_3 & \mathbf{s}_4 & \mathbf{s}_5 & \mathbf{s}_6 & \mathbf{s}_7 & \mathbf{s}_8 & \mathbf{s}_9 \\ \begin{bmatrix} t_1 & t_2 & t_2 & t_2 & t_2 & t_3 & t_3 & t_3 & t_3 \\ t_1 & t_1 & t_2 & t_2 & t_2 & t_1 & t_3 & t_3 & t_3 \\ t_1 & t_1 & t_2 & t_2 & t_2 & t_1 & t_1 & t_2 & t_3 \\ t_1 & t_1 & t_1 & t_2 & t_3 & t_1 & t_3 & t_3 & t_3 \end{bmatrix} & z_1 \\ & z_2 \\ & z_3 \\ & z_4 \end{matrix} \quad (46)$$

This response matrix satisfies neither unordered nor ordered monotonicity. When the instrument switches from  $z_1$  to  $z_4$ , agents in response type  $\mathbf{s}_5$  move from treatment  $t_2$  to treatment  $t_3$  while agents in response type  $\mathbf{s}_6$  move away from  $t_3$  and towards  $t_1$ . This represents a violation of ordered monotonicity and also prevents  $t_3$  from being ordered the highest or lowest in any ordering on  $\mathcal{T}$  that would satisfy ordered monotonicity.<sup>28</sup> Similarly we can see a switch from  $z_3$  to  $z_4$  induces agents in response type  $\mathbf{s}_3$  to move from treatment  $t_2$  to treatment  $t_1$  while inducing agents in response type  $\mathbf{s}_7$  to move away from treatment  $t_1$  and towards treatment  $t_3$ . This again represents a violation of unordered monotonicity and prevents  $t_1$  from being ordered either the highest or the lowest in any ordering  $\mathcal{T}$  that would satisfy ordered monotonicity. Since all orderings on  $\mathcal{T} = \{t_1, t_2, t_3\}$  must have either  $t_1$  or  $t_3$  as the largest or smallest element, this means there is no ordering on  $\mathcal{T}$  that satisfies ordered monotonicity.

Again, however, Theorem 6 guarantees that the response matrix satisfies minimal monotonicity (MM). Any simple 2SLS estimands exploiting the variation between any two instruments (as in Theorem 3) identifies an interpretable causal parameter. For example, if the researcher was interested in the effect of treatment  $t_1$  against alternative treatments, the 2SLS estimand  $\beta_{z_1, z_2} = E[Y | Z = z_1] - E[Y | Z = z_2]$  recovers a weighted average of  $E[Y(t_1) - Y(t_2) | \mathbf{S} = \mathbf{s}_2]$  and  $E[Y(t_1) - Y(t_3) | \mathbf{S} = \mathbf{s}_6]$ .

Interestingly, the response matrix  $\mathbf{R}$  in (46) also satisfies the extensive margin compliers only (EMCO) condition of (41) when considering any three instruments at a time. As in the last example, this allows for 2SLS estimands to additionally be interpreted as recovering the causal effect of one treatment against another.

## 8 Conclusion

Analysis of ordered and unordered IV choice models has largely been conducted in parallel, with little overlap between the two strands of the literature. Ordered choice models are commonly analyzed using ordered monotonicity (OM), introduced by Angrist and Imbens (1995), while unordered choice models are commonly analyzed using the unordered monotonicity (UM) of Heckman and Pinto (2018).

<sup>28</sup>If  $t_3$  is ranked highest a movement away from  $t_3$  represents moving towards a lower treatment while a towards  $t_3$  represents moving towards a higher treatment. Vice versa, if  $t_3$  is ranked lowest a movement towards  $t_3$  represents moving towards a lower treatment while a movement away from  $t_3$  represents moving towards a higher treatment.

This paper bridges the gap between analysis of ordered and unordered IV choice models. We note symmetric features of ordered and unordered monotonicity and use them to derive symmetric characterizations of the two. The symmetric characterizations offer deep insights into the relationship between the two monotonicity criterion. Moreover they provide computationally tractable ways to verify the two monotonicity criterion, which may be useful to researchers who wish to utilize both sets of identification results.

The symmetric characterizations illuminate a shared monotonicity property, which we term the minimal monotonicity (MM) condition. We characterize this novel criterion and show it is the minimal requirement needed to identify interpretable causal parameters using 2SLS type estimands. Moreover, minimal monotonicity can be associated with a notion of rationality that enables the investigation of a range of economic choice models that do not comply with ordered or unordered monotonicity.

## References

- Aliprantis, D. (2012). Compulsory schooling laws and educational attainment. *Journal of Educational and Behavioral Statistics* 37, 316–338.
- Angrist, J. D., K. Graddy, and G. W. Imbens (2000, July). The interpretation of instrumental variables estimators in simultaneous equations models with an application to the demand for fish. *Review of Economic Studies* 67(3), 499–527.
- Angrist, J. D. and G. W. Imbens (1995, June). Two-stage least squares estimation of average causal effects in models with variable treatment intensity. *Journal of the American Statistical Association* 90(430), 431–442.
- Angrist, J. D., G. W. Imbens, and D. Rubin (1996). Identification of causal effects using instrumental variables. *Journal of the American Statistical Association* 91(434), 444–455.
- Barua, R. and K. Lang (2016). School entry, educational attainment, and quarter of birth: A cautionary tale of a local average treatment effect. *Journal of Human Capital* 10(3), 347–376.
- Brinch, C. N., M. Mogstad, and M. Wiswall (2017). Beyond late with a discrete instrument. *Journal of Political Economy* 125(4), 985–1039.
- Buchinsky, M. and R. Pinto (2021). Economics of monotonicity conditions. *NBER Working Paper*.
- Cameron, S. V. and J. J. Heckman (1998, April). Life cycle schooling and dynamic selection bias: Models and evidence for five cohorts of American males. *Journal of Political Economy* 106(2), 262–333.
- Carneiro, P., K. Hansen, and J. J. Heckman (2003, May). Estimating distributions of treatment effects with an application to the returns to schooling and measurement of the effects of uncertainty on college choice. *International Economic Review* 44(2), 361–422.
- Cunha, F., J. J. Heckman, and S. Navarro (2007, November). The identification and economic content of ordered choice models with stochastic cutoffs. *International Economic Review* 48(4), 1273–1309.
- Dahl, C. M., M. Huber, and G. Mellace (2017). It’s never too late: A new look at local average treatment effects with or without defiers. *Discussion Papers on Business and Economics*.
- de Chaisemartin, C. (2017). Tolerating defiance? local average treatment effects without monotonicity. *Quantitative Economics* 8(2), 367–396.
- Feller, A., T. Grindal, L. Miratrix, and L. C. Page (2016). Compared to what? variation in the impacts of early childhood education by alternative care type. *The Annals of Applied Statistics* 10(3), 1245–1285.

- Frölich, M. (2007). Nonparametric iv estimation of local average treatment effects with covariates. *Journal of Econometrics* 139(1), 35–75. Endogeneity, instruments and identification.
- Goldberger, A. S. (1972, November). Structural equation methods in the social sciences. *Econometrica* 40(6), 979–1001.
- Heckman, J. and R. Pinto (2018). Unordered monotonicity. *Econometrica* 86, 1–35.
- Heckman, J. J. (1976, December). The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models. *Annals of Economic and Social Measurement* 5(4), 475–492.
- Heckman, J. J. (1979, January). Sample selection bias as a specification error. *Econometrica* 47(1), 153–162.
- Heckman, J. J. and R. Pinto (2014). Causal analysis after haavelmo. *Econometric Theory*, 1–37.
- Heckman, J. J. and R. Robb (1985, October–November). Alternative methods for evaluating the impact of interventions: An overview. *Journal of Econometrics* 30(1–2), 239–267.
- Heckman, J. J. and S. Urzúa (2010, May). Comparing IV with structural models: What simple IV can and cannot identify. *Journal of Econometrics* 156(1), 27–37.
- Heckman, J. J., S. Urzua, and E. Vytlacil (2006). Understanding instrumental variables in models with essential heterogeneity. *The Review of Economics and Statistics* 88(3), 389–432.
- Heckman, J. J., S. Urzúa, and E. J. Vytlacil (2006). Understanding instrumental variables in models with essential heterogeneity. *Review of Economics and Statistics* 88(3), 389–432.
- Heckman, J. J., S. Urzúa, and E. J. Vytlacil (2008). Instrumental variables in models with multiple outcomes: The general unordered case. *Annales d’Economie et de Statistique* 91–92, 151–174.
- Heckman, J. J. and E. J. Vytlacil (1999, April). Local instrumental variables and latent variable models for identifying and bounding treatment effects. *Proceedings of the National Academy of Sciences* 96(8), 4730–4734.
- Heckman, J. J. and E. J. Vytlacil (2005, May). Structural equations, treatment effects and econometric policy evaluation. *Econometrica* 73(3), 669–738.
- Heckman, J. J. and E. J. Vytlacil (2007a). Econometric evaluation of social programs, part I: Causal models, structural models and econometric policy evaluation. In J. J. Heckman and E. E. Leamer (Eds.), *Handbook of Econometrics*, Volume 6B, Chapter 70, pp. 4779–4874. Amsterdam: Elsevier B. V.
- Heckman, J. J. and E. J. Vytlacil (2007b). Econometric evaluation of social programs, part II: Using the marginal treatment effect to organize alternative economic estimators to evaluate

- social programs, and to forecast their effects in new environments. In J. J. Heckman and E. E. Leamer (Eds.), *Handbook of Econometrics*, Volume 6B, Chapter 71, pp. 4875–5143. Amsterdam: Elsevier B. V.
- Holland, P. W. (1986, December). Statistics and causal inference. *Journal of the American Statistical Association* 81(396), 945–960.
- Huber, M., L. Laffers, and G. Mellace (2017). Sharp iv bounds on average treatment effects on the treated and other populations under endogeneity and noncompliance. *Journal of Applied Econometrics* 32(1), 56–79.
- Huber, M. and G. Mellace (2012). Relaxing monotonicity in the identification of local average treatment effects. Unpublished manuscript, Department of Economics, University of St. Gallen.
- Huber, M. and G. Mellace (2015). Testing instrument validity for late identification based on inequality moment constraints. *Review of Economics and Statistics* 97, 398–411.
- Hull, P. (2018). Isolateing: Identifying counterfactual-specific treatment effects with cross-stratum comparisons. *Unpublished Manuscript*.
- Imbens, G. W. and J. D. Angrist (1994, March). Identification and estimation of local average treatment effects. *Econometrica* 62(2), 467–475.
- Imbens, G. W. and D. B. Rubin (1997, October). Estimating outcome distributions for compliers in instrumental variables models. *Review of Economic Studies* 64(4), 555–574.
- Kamat, V. (2021). Identifying the effects of a program offer with an application to head start. *Unpublished Manuscript*.
- Kirkeboen, L. J., E. Leuven, and M. Mogstad (2016). Field of study, earnings, and self-selection. *The Quarterly Journal of Economics* 131(3), 1057–1111.
- Klein, T. J. (2010). Heterogeneous treatment effects: Instrumental variables without monotonicity? *Journal of Econometrics* 155, 99–116.
- Kline, P. and C. R. Walters (2016, 07). Evaluating public programs with close substitutes: The case of head start. *The Quarterly Journal of Economics* 131(4), 1795–1848.
- Lee, S. and B. Salanié (2018). Identifying effects of multivalued treatments. *Econometrica* 86, 1939–1963.
- Mogstad, M., A. Santos, and A. Torgovitsky (2018). Using instrumental variables for inference about policy relevant treatment parameters. *Econometrica* 86(5), 1589–1619.
- Mogstad, M. and A. Torgovitsky (2018). Identification and extrapolation of causal effects with instrumental variables. *Annual Review of Economics* 10(1), 577–613.



- Mogstad, M., A. Torgovitsky, and C. R. Walters (2021). The causal interpretation of two-stage least squares with multiple instrumental variables. *American Economic Review* 111(11), 3663–3698.
- Mountjoy, J. (2021). Community colleges and upward mobility. *Unpublished Manuscript*.
- Pinto, R. (2021). Beyond intention to treat: Using the incentives in moving to opportunity to identify neighborhood effects. *NBER Working Paper*.
- Pinto, R. and J. Heckman (2021). The econometric model for causal policy analysis. *Manuscript Prepared for Annual Review of Economics (2021) edition*.
- Rao, C. R. (1946a). Difference sets and combinatorial arrangements derivable from finite geometries. *Proceedings of the Indian National Science Academy* 12(3), 123–135.
- Rao, C. R. (1946b). Hypercubes of strength ‘d’ leading to confounded designs in factorial experiments. *Bulletin of the Calcutta Mathematical Society* 38, 67–78.
- Rao, C. R. (1947). Factorial experiments derivable from combinatorial arrangements of arrays. *Supplement to the Journal of the Royal Statistical Society* 9(1), 128–139.
- Rao, C. R. (1949). On a class of arrangements. *Proceedings of the Edinburgh Mathematical Society* 8(3), 119–125.
- Rose, E. K. and Y. Shem-Tov (2021). On recoding ordered treatments as binary indicators. *Unpublished Manuscript*.
- Rubin, D. B. (1974, October). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology* 66(5), 688–701.
- Rubin, D. B. (1978a). Bayesian inference for causal effects. *The Annals of Statistics* 6(1), 34–58.
- Rubin, D. B. (1978b, January). Bayesian inference for causal effects: The role of randomization. *Annals of Statistics* 6(1), 34–58.
- Ryser, H. (1957). Combinatorial properties of matrices of zeros and ones. *Canadian Journal of Mathematics* 9, 371–377.
- Small, D. S. and Z. Tan (2007). A stochastic monotonicity assumption for the instrumental variables method. *Technical report, Department of Statistics, Wharton School, University of Pennsylvania*.
- Theil, H. (1953). *Estimation and Simultaneous Correlation in Complete Equation Systems*. The Hague: Central Planning Bureau. Mimeographed memorandum.
- Vytlacil, E. J. (2002, January). Independence, monotonicity, and latent index models: An equivalence result. *Econometrica* 70(1), 331–341.

- Vytlacil, E. J. (2006, August). Ordered discrete-choice selection models and local average treatment effect assumptions: Equivalence, nonequivalence, and representation results. *Review of Economics and Statistics* 88(3), 578–581.
- Zelen, M. (1979). A new design for randomized clinical trials. *New England Journal of Medicine* 300(22), 1242–1245. PMID: 431682.
- Zelen, M. (1990). Randomized consent designs for clinical trials: An update. *Statistics in Medicine* 9(6), 645–656.

## A Proofs of Main Results

### A.1 Lonesum Matrix Characterizations

Using the characterization of unordered monotonicity in UM-Sequence (17), unordered monotonicity is equivalent to there being a permutation of the rows of  $\mathbf{B}_t$  such that each column of  $\mathbf{B}_t$  is weakly increasing.<sup>1</sup> Existence of such a reordering characterizes a class of binary matrices known as lonesum matrices, which are a generalization of lower triangular binary matrices. The lonesum property, and various characterizations of it, end up forming the basis of much of our proof strategy. We discuss the property briefly below and note a useful characterization of the property.

#### Lonesum Matrices

Following Ryser (1957), a binary matrix  $\mathbf{A}$  is lonesum if each of its entries is uniquely determined by its column and row sums. Matrix  $\mathbf{A}$  below is an example of such a lonesum matrix:

$$\begin{array}{rcc}
 \mathbf{A} & = & \begin{array}{ccccc}
 & & \text{row} & & \text{row-sum} \\
 \begin{bmatrix} 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 & 0 \end{bmatrix} & \begin{array}{l} r_1 \\ r_2 \\ r_3 \end{array} & \begin{array}{l} 2 \\ 4 \\ 1 \end{array} \\
 \text{column} & & c_1 & c_2 & c_3 & c_4 & c_5 \\
 \text{column-sum} & & 0 & 3 & 1 & 2 & 1
 \end{array} \Rightarrow \begin{array}{ccccc}
 & & \text{row} & & \text{row-sum} \\
 \begin{bmatrix} 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 1 & 1 & 1 & 1 \end{bmatrix} & \begin{array}{l} r_3 \\ r_1 \\ r_2 \end{array} & \begin{array}{l} 1 \\ 2 \\ 4 \end{array} \\
 c_1 & c_3 & c_5 & c_4 & c_2 \\
 0 & 1 & 1 & 2 & 3
 \end{array}
 \end{array}
 \end{array}
 \underbrace{\hspace{10em}}_{\text{Original Matrix}} \quad \underbrace{\hspace{10em}}_{\text{Reordered Matrix}}$$

We can reorder the rows of the matrix  $\mathbf{A}$  such that the elements of each column are weakly increasing. We can also reorder the columns so that the matrix  $\mathbf{A}$  is lower triangular, which is why the lonesum property is considered a generalization of binary lower triangular matrices. The lonesum matrix property can also be productively characterized in the following ways:

**Lemma 3** (Lonesum Matrices). *A binary matrix  $\mathbf{A} \in \{0, 1\}^{m \times n}$  is lonesum if and only if:*

- (i). *Matrix  $\mathbf{A}$  is lower-triangular under column and row permutations.*
- (ii). *There are no  $2 \times 2$  submatrix in  $\mathbf{A}$  of the form either:*

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \text{ or } \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}. \tag{A.1}$$

- (iii).  *$\mathbf{1}^\top((\mathbf{1} - \mathbf{A})^\top \mathbf{A}) \odot ((\mathbf{1} - \mathbf{A})^\top \mathbf{A})^\top \mathbf{1} = 0$ , where  $\mathbf{1}$  is a  $n$ -dimensional vector of elements ones and  $\mathbf{1}$  is a  $m \times n$  matrix of element ones.*

- (iv). *Let  $r_i(\mathbf{A})$  and  $c_j(\mathbf{A})$  represent the row sum of row  $i$  and the column sum of column  $j$ ,*

<sup>1</sup>This permutation can differ for each  $\mathbf{B}_t$ , but there must be such a permutation for each  $t \in \mathcal{T}$ .

respectively. Each entry  $\mathbf{A}[i, j]$ , for  $1 \leq i \leq m$  and  $1 \leq j \leq n$ , can be expressed as:

$$\mathbf{A}[i, j] = \mathbf{1} \left[ r_i(\mathbf{A}) \geq \sum_{j'=1}^n \mathbf{1} [c_{j'}(\mathbf{A}) \geq c_j(\mathbf{A})] \right] \quad (\text{A.2})$$

The sum  $\sum_{j'=1}^n \mathbf{1} [c_{j'}(\mathbf{A}) \geq c_j(\mathbf{A})]$  represents the number of columns of  $\mathbf{A}$  with a weakly larger column sum than column  $j$

*Proof.* The second item comes from [Ryser \(1957\)](#). We next show a series of equivalences.

(iii)  $\iff$  (ii). Notice that (iii) is equivalent to the matrix

$$(\mathbf{1} - \mathbf{A})^\top \mathbf{A} \odot \mathbf{A}^\top (\mathbf{1} - \mathbf{A})$$

being a matrix of all zeros. By direct calculation, the  $ij^{\text{th}}$  element of  $(\mathbf{1} - \mathbf{A})^\top \mathbf{A}$  is given

$$\sum_{k=1}^m \mathbf{A}[k, i] - \mathbf{A}[k, j] \mathbf{A}[k, i] = \sum_{k=1}^m \mathbf{A}[k, i] (1 - \mathbf{A}[k, j]).$$

This is non-zero if and only if  $\mathbf{A}[k, i] = 1$  and  $\mathbf{A}[k, j] = 0$  for some  $k$ . Conversely, the  $ij^{\text{th}}$  element of  $\mathbf{A}^\top (\mathbf{1} - \mathbf{A})$  can be expressed

$$\sum_{k=1}^m \mathbf{A}[k, j] - \mathbf{A}[k, i] \mathbf{A}[k, j] = \sum_{k=1}^m \mathbf{A}[k, j] (1 - \mathbf{A}[k, i]).$$

This is non-zero if and only if  $\mathbf{A}[k, i] = 0$  and  $\mathbf{A}[k, j] = 1$  for some  $k$ . The  $ij^{\text{th}}$  element of  $(\mathbf{1} - \mathbf{A})^\top \mathbf{A} \odot \mathbf{A}^\top (\mathbf{1} - \mathbf{A})$  is non-zero if and only if each of the terms above are nonzero, which in turn is equivalent to their existing two rows  $k, k'$  such that

$$\begin{pmatrix} \mathbf{A}[k, i] & \mathbf{A}[k, j] \\ \mathbf{A}[k', i] & \mathbf{A}[k', j] \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

which is equivalent to one of the restricted submatrices in item (ii) up to a relabeling of  $k$  and  $k'$ .

(iv)  $\iff$  (ii). For the forward direction, notice that if (iv) holds, then the matrix  $\mathbf{A}$  can clearly be reproduced uniquely from its row and column sums, so it is lonesum and thus (ii) holds. We then want to show that (ii)  $\implies$  (iv). Consider the contrapositive. First, suppose that there is some element  $\mathbf{A}[i, j]$  such that  $\mathbf{A}[i, j] = 1$  but

$$r_i(\mathbf{A}) < \sum_{j'=1}^n \mathbf{1} [c_{j'}(\mathbf{A}) \geq c_j(\mathbf{A})]. \quad (\text{A.3})$$

That is, the row sum of row  $i$  is less than the number of columns  $j'$  with column sum larger

than column  $j$ . For this to be the case, there must be some column  $j'$  such that  $\mathbf{A}[i, j'] = 0$  but  $c_{j'}(\mathbf{A}) \geq c_j(\mathbf{A})$ .<sup>2</sup> Because  $\mathbf{A}[i, j'] = 0$  we know that  $j \neq j'$ .

Because  $\mathbf{A}[i, j] = 1$  but  $\mathbf{A}[i, j'] = 0$ , for the column sum of  $j'$  to be weakly larger than the column sum of  $j$ , there must be some other row  $i' \neq i$  such that  $\mathbf{A}[i', j] = 0$  but  $\mathbf{A}[i', j'] = 1$ . This generates the restricted pattern which violates (ii) (up to a relabeling of columns  $j$  and  $j'$ ).

$$\begin{pmatrix} \mathbf{A}[i, j] & \mathbf{A}[i, j'] \\ \mathbf{A}[i', j] & \mathbf{A}[i', j'] \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Conversely, suppose that  $\mathbf{A}[i, j] = 0$  but

$$r_i(\mathbf{A}) \geq \sum_{j'=1}^n \mathbf{1}[c_{j'}(\mathbf{A}) \geq c_j(\mathbf{A})] \quad (\text{A.4})$$

By similar logic as the previous case, this means that there must be some column  $j'$  such that  $\mathbf{A}[i, j'] = 1$  but  $c_{j'}(\mathbf{A}) < c_j(\mathbf{A})$ . Since column  $j$  does not contribute to the row sum on the right hand side of (A.4) but does contribute to the column sum on the right hand side, there must be some other column that does the opposite in order for (A.4) to hold.

Because  $\mathbf{A}[i, j'] = 1$  and  $\mathbf{A}[i, j] = 0$  but  $c_{j'}(\mathbf{A}) < c_j(\mathbf{A})$ , there must be some other row  $i' \neq i$  such that  $\mathbf{A}[i', j] = 1$  but  $\mathbf{A}[i', j'] = 0$ . As before, we can show that this generates the restricted pattern which violates (ii) (up to a relabeling of columns  $j$  and  $j'$ ).

(i)  $\iff$  (ii). That (i)  $\implies$  (ii) is clear, since the existence of the restricted pattern is stable under row and column permutations and the existence of such a restricted pattern prevents a matrix from being lower triangular. To see that (ii)  $\implies$  (i), suppose that  $\mathbf{A}$  is lonesum and let  $\tilde{\mathbf{A}}$  be the matrix generated by ordering the rows of  $\mathbf{A}$  in increasing row-sum and the columns of  $\mathbf{A}$  in decreasing column sum. Now, notice that  $\tilde{\mathbf{A}}$  must also be lonesum, since if the restricted pattern appears in  $\tilde{\mathbf{A}}$  it must also appear in  $\mathbf{A}$ ; we cannot generate the restricted pattern using row and column permutations of a lonesum matrix. If  $\tilde{\mathbf{A}}$  is lonesum, then (iv) must hold, that is

$$\tilde{\mathbf{A}}[i, j] = \mathbf{1} \left[ r_i(\tilde{\mathbf{A}}) \geq \sum_{j'=1}^n \mathbf{1} [c_{j'}(\tilde{\mathbf{A}}) \geq c_j(\tilde{\mathbf{A}})] \right]. \quad (\text{A.5})$$

Since the columns of  $\tilde{\mathbf{A}}$  are ordered in increasing column sum, we must have by (A.5) that  $\tilde{\mathbf{A}}[i, j] \leq \tilde{\mathbf{A}}[i, j']$  for  $j' > j$ . Similarly, since the rows are ordered in decreasing row sum, (A.5) implies that  $\tilde{\mathbf{A}}[i, j] \leq \tilde{\mathbf{A}}[i', j]$  for  $i' < i$ . Together, these imply that  $\tilde{\mathbf{A}}$  is lower triangular. □

---

<sup>2</sup>There must be some column  $j'$  that contributes to the right hand side of (A.3) but not to the row sum on the left hand side.

## A.2 Proofs of Results in Section 4

### A.2.1 Proof of Theorem 1

We prove Theorem 1 via a series of implications:

(i)  $\implies$  (ii). If there is a response type  $\mathbf{s}$  and a treatment  $t$  for which (ii) is not true, then the sequence

$$\left( \mathbf{1}[\mathbf{s}[z_1^{(t)}] = t], \dots, \mathbf{1}[\mathbf{s}[z_{N_Z}^{(t)}] = t] \right)$$

is not increasing.

(ii)  $\implies$  (iii). If there is a  $2 \times 2$  matrix of  $\mathbf{R}$  of the form

$$\begin{array}{cc} \mathbf{s} & \mathbf{s}' \\ \left( \begin{array}{cc} t & t' \\ t'' & t \end{array} \right) & \begin{array}{c} z \\ z' \end{array} \end{array}$$

then we have  $\mathbf{1}[\mathbf{s}[z] = t] > \mathbf{1}[\mathbf{s}[z'] = t]$  while  $\mathbf{1}[\mathbf{s}'[z] = t] < \mathbf{1}[\mathbf{s}'[z'] = t]$ , a violation of (ii). A symmetric argument holds for the other restricted submatrix of  $\mathbf{R}$ .

(iii)  $\implies$  (iv). First notice that (iv) is equivalent to the matrix  $\mathbf{U}$  being lonesum by part three of Lemma 3. Further notice that the matrix  $\mathbf{U}$  is lonesum if and only if each  $\mathbf{B}_t$  is lonesum. Because there are no restricted submatrices of  $\mathbf{R}$  of the form in (iii) there are no submatrices of any  $\mathbf{B}_t$  of the form either

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad \text{or} \quad \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

By part (ii) of Lemma 3, this is equivalent to each  $\mathbf{B}_t$  being lonesum.

(iv)  $\implies$  (v). Item (iv) is equivalent to each  $\mathbf{B}_t$  being lonesum. We seek to use the fourth item of Lemma 3. With this in mind, define the functions

$\zeta(z, t) \equiv$  row sum of the  $z^{\text{th}}$  row of the matrix  $\mathbf{B}_t$

$\varphi(\mathbf{s}, t) \equiv$  # of columns of  $\mathbf{B}_t$  with a larger column sum than that of its  $\mathbf{s}^{\text{th}}$  column.

Because  $\mathbf{S}$  is implicitly a function of unobserved confounders  $\mathbf{V}$ , we can also write  $\varphi(\cdot, t)$  as a function of  $\mathbf{V}$ . By the definition of  $\mathbf{B}_t = \mathbf{1}[\mathbf{R} = t]$  and the fourth item of Lemma 3, we can then write

$$\mathbf{1}[T = t] = \mathbf{1}[\zeta(Z, t) \geq \varphi(\mathbf{V}, t)].$$

By the first item of Lemma 3, there is a reordering of the rows and columns of  $\mathbf{B}_t$  such that  $\mathbf{B}_t$  is lower triangular. Let  $z_1^{(t)}$  denote the instrument associated with the “top” row of this matrix,  $z_2^{(t)}$  denote the second row of this matrix, and so on till  $z_{N_Z}^{(t)}$ .

Then, the function  $\zeta$  satisfies  $\zeta(z_{k+1}^{(t)}, t) \geq \zeta(z_k^{(t)}, t)$  for  $k = 1, \dots, N_Z - 1$ , by definition of the sequence  $(z_1^{(t)}, \dots, z_{N_Z}^{(t)})$ ; weakly more response types must be taking up treatment  $t$  for each successive value of this sequence.

$(v) \implies (i)$ . Since  $\zeta(z_{k+1}^{(t)}, t) > \zeta(z_k^{(t)}, t)$  for  $k = 1, \dots, N_Z - 1$ , if  $(v)$  holds we must have that

$$(\mathbf{1}[T_i(z_1^{(t)}) = t], \dots, \mathbf{1}[T_i(z_{N_Z}^{(t)}) = t])$$

is a weakly increasing sequence in  $\{0, 1\}$  for all  $i \in \mathcal{I}$ .

### A.2.2 Proof of Theorem 2

We prove Theorem 2 via a series of implications:

$(i) \implies (ii)$ . Take any strict ordering on  $\mathcal{T}$ . Suppose there is a violation of  $(ii)$  for some  $i \in \mathcal{I}$ . Then for that particular  $i \in \mathcal{I}$  the sequence

$$(T_i(z_1), \dots, T_i(z_{N_Z}))$$

is not increasing with respect to the ordering on  $\mathcal{T}$ . If there is no such ordering satisfying  $(ii)$ , then  $(i)$  cannot be satisfied.

$(ii) \implies (iii)$ . Suppose there is a  $2 \times 2$  submatrix of  $\mathbf{R}$  of the form:

$$\begin{pmatrix} \mathbf{s} & \mathbf{s}' \\ t & t' \\ t'' & t''' \end{pmatrix} \begin{matrix} z \\ z' \end{matrix}$$

for some  $t'' > t$  and  $t''' < t'$ . This means that  $\mathbf{s}[z'] > \mathbf{s}[z]$  while  $\mathbf{s}'[z] > \mathbf{s}'[z']$ . These two statements cannot both be true under  $(ii)$ . A symmetric argument applies for the other  $2 \times 2$  restricted submatrix.

$(iii) \implies (iv)$ . Notice that by part 3 of Lemma 3 that  $(iv)$  is equivalent to  $\mathbf{O}$  being lonesum. Suppose that  $\mathbf{O} = [\mathbf{B}_{t_1}^*, \dots, \mathbf{B}_{t_{N_T}}^*]$  is not lonesum. That is, by the lonesum characterization (A.1) there is a  $2 \times 2$  submatrix of  $\mathbf{O}$  of the form

$$\begin{pmatrix} \mathbf{s} & \tilde{\mathbf{s}} \\ 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{matrix} z \\ z' \end{matrix} \quad \text{or} \quad \begin{pmatrix} \mathbf{s} & \tilde{\mathbf{s}} \\ 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{matrix} z \\ z' \end{matrix} .$$

WLOG suppose there is a  $2 \times 2$  submatrix of  $\mathbf{O}$  of the first form. This indicates that for some  $t \in \mathcal{T}$ , the instrumental switch from  $z$  to  $z'$  induces agents of response type  $\mathbf{s}$  to switch from a treatment weakly below  $t$  to a treatment strictly greater than  $t$ . That is  $\mathbf{s}[z] < \mathbf{s}[z']$ . Conversely, for some

treatment  $t' \in \mathcal{T}$  the instrumental switch from  $z$  to  $z'$  induces agents of response  $\tilde{\mathbf{s}}$  to switch from a treatment strictly greater than  $t'$  to a treatment weakly lower than  $t'$ . That is  $\tilde{\mathbf{s}}[z] < \tilde{\mathbf{s}}[z']$ .

These two statements are incompatible with each other, so we cannot have that  $\mathbf{s} = \tilde{\mathbf{s}}$ . Leting  $t = \mathbf{s}[z], t' = \tilde{\mathbf{s}}[z], t'' = \mathbf{s}[z']$ , and  $t''' = \tilde{\mathbf{s}}[z']$ , this implies a  $2 \times 2$  submatrix of  $\mathbf{R}$  of the form

$$\begin{array}{cc} \mathbf{s} & \tilde{\mathbf{s}} \\ \left( \begin{array}{cc} t & t' \\ t'' & t''' \end{array} \right) & \begin{array}{c} z \\ z' \end{array} \end{array}$$

with  $t < t''$  and  $t' > t'''$ . This is a violation of (iii). Similarly, considering  $2 \times 2$  submatrices of  $\mathbf{O}$  of the second form, we can find a violation of the other pattern restricted by (iii).

(iv)  $\implies$  (v). Item (iv) is equivalent to  $\mathbf{O}$  being lonesum. This in turn implies that  $\mathbf{B}_t^* = \mathbf{1}[\mathbf{R} \geq t]$  is lonesum for each  $t \in \mathcal{T}$ . With this in mind, define the functions

$$\zeta(z, t) \equiv \text{row sum of the } z^{\text{th}} \text{ row of the matrix } \mathbf{B}_t^*$$

$$\varphi(\mathbf{s}, t) \equiv \# \text{ of columns of } \mathbf{B}_t^* \text{ with a larger column sum than that of its } \mathbf{s}^{\text{th}} \text{ column.}$$

By the first item of Lemma 3, there is a reordering of the rows and columns of  $\mathbf{O}$  such that  $\mathbf{O}$  is lower triangular. Let  $z_1^{(t)}$  denote the instrument associated with the “top” row of this matrix,  $z_2^{(t)}$  denote the second row of this matrix, and so on till  $z_{N_Z}^{(t)}$ .

Because  $\mathbf{S}$  is implicitly a function of unobserved confounders  $\mathbf{V}$ , we can also write  $\varphi(\cdot, t)$  as a function of  $\mathbf{V}$ . By the definition of  $\mathbf{B}_t^* = \mathbf{1}[\mathbf{R} \geq t]$  and the fourth item of Lemma 3, we can then write

$$\mathbf{1}[T \geq t] = \mathbf{1}[\zeta(Z, t) \geq \varphi(\mathbf{V}, t)]$$

for each  $t$ . Moreover, by definition of the sequence  $(z_1, \dots, z_{N_Z})$  we know that, for each treatment  $t \in \mathcal{T}$ , weakly more response types take up treatments larger than  $t$  as the instrument cycles through the sequence  $(z_1, \dots, z_{N_Z})$ . By definition of the  $\zeta(\cdot, \cdot)$  function then,  $\zeta(z_{k+1}, t) > \zeta(z_k, t)$  for  $k = 1, \dots, N_Z - 1$  and all  $t$ .

(v)  $\implies$  (i). Since  $\zeta(z_{k+1}, t) > \zeta(z_k, t)$  for  $k = 1, \dots, N_Z - 1$  and all  $t$ , if holds (v) we must have that, for all  $t$

$$(\mathbf{1}[T_i(z_1) \geq t], \dots, \mathbf{1}[T_i(z_{N_Z}) \geq t])$$

is a weakly increasing sequence in  $\{0, 1\}$  for all  $i \in \mathcal{I}$ . This implies that

$$(T_i(z_1), \dots, T_i(z_{N_Z}))$$

must be a weakly increasing sequence with respect to the ordering on  $\mathcal{T}$  for all  $i \in \mathcal{I}$ .



### A.3 Proofs of Results in Section 5

#### A.3.1 Proof of Theorem 3

We show a system of implications.

(i)  $\iff$  (ii). This is provided by Lemma 1.

(ii)  $\iff$  (iv). This follows from the discussion in Appendix B, namely the decomposition of  $\beta_{z,z'}$  in equation (B.6), and the definition of an interpretable causal parameter in (26). The decomposition of  $\beta_{z,z'}$  gives the forward direction. The definition of an interpretable causal parameter gives the backwards direction: if there is a negative weight there must be a pair of treatments  $t, t'$  such that some agents that are switching from  $t$  to  $t'$  as the instrument ranges from  $z$  to  $z'$  whereas that same instrumental switch moves other agents from  $t'$  to  $t$ .

(iii)  $\implies$  (ii). We consider the contrapositive. Suppose there is no binary matrix  $\mathbf{B}$  element-wise less than or equal to  $\sum_{t'' \neq t, t'} \mathbf{B}_{t''}$  such that  $\mathbf{B}_t + \mathbf{B}$  is lonesum. This means that  $\mathbf{B}_t$  is lonesum, so that by Theorem 1 there is a  $2 \times 2$  submatrix  $\mathbf{R}$  of either the form

$$\begin{pmatrix} t & t''' \\ t'' & t \end{pmatrix} \text{ or } \begin{pmatrix} t'' & t \\ t & t''' \end{pmatrix},$$

for some  $t'', t''' \neq t$ . If either  $t'' \neq t'$  or  $t''' \neq t'$ , then we can find a binary matrix that is element wise less than  $\sum_{t'' \neq t, t'} \mathbf{B}_{t''}$  to “fill in the gap” and get rid of the restricted pattern. In particular we can take  $\tilde{\mathbf{B}}$  to be the matrix that is equal to one at the position of either  $t''$  or  $t'''$  and zero everywhere else. If there is no such matrix that eliminates the restricted pattern then both  $t'' = t'$  and  $t''' = t'$ . So we have the restricted pattern (25) in  $\mathbf{R}$ .

(ii)  $\iff$  (iii). The  $ij^{\text{th}}$  element of  $\mathbf{B}_t^\top \mathbf{B}_{t'}$  is given

$$\sum_{z=z_1}^{N_Z} \mathbf{1}[s_i[z] = t] \mathbf{1}[s_j[z] = t'],$$

this is nonzero if and only if we have  $s_i[z] = t$  and  $s_j[z] = t'$  for some instrument value  $z$ . Similarly, the  $ij^{\text{th}}$  element of  $(\mathbf{B}_t^\top \mathbf{B}_{t'})^\top = \mathbf{B}_{t'}^\top \mathbf{B}_t$  is given

$$\sum_{z=z_1}^{N_Z} \mathbf{1}[s_i[z] = t'] \mathbf{1}[s_j[z] = t].$$

This is non-zero if and only if we have  $s_i[z'] = t'$  and  $s_j[z'] = t$  for some instrument value  $z'$ .

Then, the  $ij^{\text{th}}$  element of the Hadamard product  $(\mathbf{B}_t^\top \mathbf{B}_{t'}) \odot (\mathbf{B}_t^\top \mathbf{B}_{t'})^\top$  is non-zero if and only if the  $ij^{\text{th}}$  elements of both  $(\mathbf{B}_t^\top \mathbf{B}_{t'})$  and  $(\mathbf{B}_t^\top \mathbf{B}_{t'})^\top$  are non-zero. By the characterizations above and because each response type is a well defined function, this is equivalent to  $s_i[z] = t, s_i[z'] = t'$  but

$\mathbf{s}_j[z] = t', \mathbf{s}_j[z'] = t$  for some  $z' \neq z$ . This is equivalent to the restricted pattern (25) existing between  $\mathbf{s}_i$  and  $\mathbf{s}_j$  for instrument values  $z, z'$  and the specific treatment values  $t$  and  $t'$ .

All elements of  $(\mathbf{B}_t^\top \mathbf{B}_{t'}) \odot (\mathbf{B}_t^\top \mathbf{B}_{t'})^\top$  being equal to zero is then equivalent to their being no restricted patterns (25) between the specific treatment values  $t$  and  $t'$  in the response matrix  $\mathbf{R}$ .

Because each  $(\mathbf{B}_t^\top \mathbf{B}_{t'}) \odot (\mathbf{B}_t^\top \mathbf{B}_{t'})^\top$  has weakly positive entries, checking whether

$$\iota^\top \left( \sum_{(t,t') \in \mathcal{C}_2(\mathcal{T})} (\mathbf{B}_t^\top \mathbf{B}_{t'}) \odot (\mathbf{B}_t^\top \mathbf{B}_{t'})^\top \right) \iota = 0$$

is equivalent to checking whether each  $(\mathbf{B}_t^\top \mathbf{B}_{t'}) \odot (\mathbf{B}_t^\top \mathbf{B}_{t'})^\top$  is equal to the zero matrix. By the discussion above, this is equivalent to checking whether there are no matrices of the form (25) for any  $t, t' \in \mathcal{T}$ .

### A.3.2 Proof of Theorem 4

Let  $\Psi_U(t) = ((\tilde{\mathbf{1}} - \mathbf{B}_t)^\top \mathbf{B}_t) \odot ((\mathbf{1} - \mathbf{B}_t)^\top \mathbf{B}_t)^\top$ , where  $\tilde{\mathbf{1}}$  is a  $N_Z \times N_S$  matrix of element ones. Using this notation, we can rewrite the UM verification in Item (iv) of Theorem 1 as the following sum:

$$\begin{aligned} \|\Psi_U\| &= \|((\mathbf{1} - \mathbf{U})^\top \mathbf{U}) \odot ((\mathbf{1} - \mathbf{U})^\top \mathbf{U})^\top\| \\ &= \left\| \sum_{t \in \mathcal{T}} ((\tilde{\mathbf{1}} - \mathbf{B}_t)^\top \mathbf{B}_t) \odot ((\tilde{\mathbf{1}} - \mathbf{B}_t)^\top \mathbf{B}_t)^\top \right\| \\ &= \sum_{t \in \mathcal{T}} \|((\tilde{\mathbf{1}} - \mathbf{B}_t)^\top \mathbf{B}_t) \odot ((\tilde{\mathbf{1}} - \mathbf{B}_t)^\top \mathbf{B}_t)^\top\| \\ &= \sum_{t \in \mathcal{T}} \|\Psi_U(t)\| \end{aligned}$$

The first equality is from the definition of the verification in Theorem 1. The second equality arises from the construction of matrix  $\mathbf{U}$ . The third equality is due to the fact that all elements of  $\Psi_U(t)$  are either zero or a natural number.

We can use the fact that  $\sum_{t \in \mathcal{T}} \mathbf{B}_t = \tilde{\mathbf{1}}$  to express matrix  $\Psi_U(t)$  as:

$$\begin{aligned} \Psi_U(t) &= \left( (\tilde{\mathbf{1}} - \mathbf{B}_t)^\top \mathbf{B}_t \right) \odot \left( (\tilde{\mathbf{1}} - \mathbf{B}_t)^\top \mathbf{B}_t \right)^\top \\ &= \left( \left( \left( \sum_{t' \in \mathcal{T}} \mathbf{B}_{t'} \right) - \mathbf{B}_t \right)^\top \mathbf{B}_t \right) \odot \left( \left( \left( \sum_{t' \in \mathcal{T}} \mathbf{B}_{t'} \right) - \mathbf{B}_t \right)^\top \mathbf{B}_t \right)^\top \\ &= \left( \left( \sum_{t' \in \mathcal{T} \setminus \{t\}} \mathbf{B}_{t'} \right)^\top \mathbf{B}_t \right) \odot \left( \left( \sum_{t' \in \mathcal{T} \setminus \{t\}} \mathbf{B}_{t'} \right)^\top \mathbf{B}_t \right)^\top \end{aligned}$$

$$\begin{aligned}
&= \left( \sum_{t' \in \mathcal{T} \setminus \{t\}} \mathbf{B}_{t'}^\top \mathbf{B}_t \right) \odot \left( \sum_{t' \in \mathcal{T} \setminus \{t\}} \mathbf{B}_{t'}^\top \mathbf{B}_t \right)^\top \\
&= \left( \sum_{t' \in \mathcal{T} \setminus \{t\}} \mathbf{B}_{t'}^\top \mathbf{B}_t \right) \odot \left( \sum_{t' \in \mathcal{T} \setminus \{t\}} \mathbf{B}_t^\top \mathbf{B}_{t'} \right) \\
&= \sum_{t' \in \mathcal{T} \setminus \{t\}} (\mathbf{B}_{t'}^\top \mathbf{B}_t) \odot (\mathbf{B}_{t'}^\top \mathbf{B}_t)^\top + 2 \sum_{t', t'' \in \mathcal{T} \setminus \{t\}} (\mathbf{B}_{t'}^\top \mathbf{B}_t) \odot (\mathbf{B}_t^\top \mathbf{B}_{t''})^\top \\
&= \sum_{t' \in \mathcal{T} \setminus \{t\}} \Psi(t', t, t, t') + 2 \sum_{t', t'' \in \mathcal{T} \setminus \{t\}} \Psi(t', t, t, t'')
\end{aligned}$$

The derivation above use simple rules of matrix algebra and the formula for the product of sums. We use the fact that  $\Psi(t', t, t, t'')^\top = \Psi(t, t', t'', t)$  and express the transpose of  $\Psi_U(t)$  as:

$$\Psi_U(t)^\top = \sum_{t' \in \mathcal{T} \setminus \{t\}} \Psi(t, t', t', t) + 2 \sum_{(t', t'') \in \mathcal{C}_2(\mathcal{T} \setminus \{t\})} \Psi(t, t', t'', t) \quad (\text{A.6})$$

Recall that the elements of matrix  $\Psi(t, t', t'', t''')$  are either zero or natural numbers. Thus, equation (A.6) implies that  $\|\Psi_U(t)\| = 0$  (or equivalently  $\|\Psi_U(t)^\top\| = 0$ ) if and only if:

$$\|\Psi(t, t', t', t)\| = 0 \text{ for all } t' \in \mathcal{T} \setminus \{t\} \quad (\text{A.7})$$

$$\text{and } \|\Psi(t, t', t'', t)\| = 0 \text{ for all combinations of } t', t'' \in \mathcal{T} \setminus \{t\}. \quad (\text{A.8})$$

Now  $\|\Psi_U\| = 0$  only and only if  $\|\Psi_U(t)\| = 0$  for all  $t \in \mathcal{T}$ , which completes the proof.

## A.4 Proof of Theorem 5

### A.4.1 Lemmas

**Proof of Lemma 1.** Suppose there is a violation of MM (24). This is equivalent to there being pair of response types  $\mathbf{s}, \mathbf{s}'$ , a pair of treatments  $z, z'$ , and a pair of treatments  $t, t'$  such that

$$\begin{aligned}
&\mathbf{1}[\mathbf{s}[z] = t] \mathbf{1}[\mathbf{s}'[z'] = t'] > \mathbf{1}[\mathbf{s}[z] = t'] \mathbf{1}[\mathbf{s}'[z'] = t] \\
&\text{and } \mathbf{1}[\mathbf{s}'[z] = t] \mathbf{1}[\mathbf{s}[z'] = t'] < \mathbf{1}[\mathbf{s}'[z] = t'] \mathbf{1}[\mathbf{s}[z'] = t].
\end{aligned}$$

This is in turn equivalent to  $\mathbf{s}[z] = t, \mathbf{s}[z'] = t'$  and  $\mathbf{s}'[z] = t', \mathbf{s}'[z'] = t$ , which is equivalent (up to a relabeling of  $\mathbf{s}$  and  $\mathbf{s}'$ ) to the restricted pattern (25) appearing in the response matrix  $\mathbf{R}$ .

**Proof of Lemma 2.** Follows from Theorems 4 and 5. That MM holds when OM and UM fail can be seen via examples in Section 7.

## A.5 Proof of Results in Section 6

### A.5.1 Proof of Theorem 6

Consider any pair of instrument values  $z, z' \in \mathcal{Z}$  and any pair of treatments  $t, t' \in \mathcal{T}$ . Without loss of generality, it is enough show there are no  $2 \times 2$  submatrices of the form

$$\begin{array}{cc} \mathbf{s} & \mathbf{s}' \\ \begin{pmatrix} t & t' \\ t' & t \end{pmatrix} & \begin{matrix} z \\ z' \end{matrix} \end{array}.$$

Define  $\Delta_t \equiv \mathbf{L}[z', t] - \mathbf{L}[z, t]$  and  $\Delta_{t'} \equiv \mathbf{L}[z', t'] - \mathbf{L}[z, t']$ . There are two scenarios. Either  $\Delta_t \leq \Delta_{t'}$  or  $\Delta_t \geq \Delta_{t'}$ . In each case, we have the following behavioral restrictions from the Choice Rule (35).

$$\begin{aligned} \text{If } \Delta_t \leq \Delta_{t'} \text{ then } T_i(z) = t' &\implies T_i(z') \neq t \\ \text{If } \Delta_{t'} \leq \Delta_t \text{ then } T_i(z) = t &\implies T_i(z') \neq t' \end{aligned}$$

The first restriction would eliminate the response type  $\mathbf{s}'$  from the matrix  $\mathbf{R}$  while the second restriction would eliminate the response type  $\mathbf{s}$  from the response matrix  $\mathbf{R}$ . In either case, we cannot have the restricted  $2 \times 2$  submatrix displayed at the top of the proof.

## B 2SLS Analysis

### B.1 Interpretation of 2SLS under Ordered and Unordered Monotonicity

Under ordered monotonicity and a binary instruments, Angrist and Imbens (1995) show that the 2SLS estimand identifies the following:

$$\begin{aligned} \beta_{2\text{SLS}} &= \frac{E[Y | Z = z_1] - E[Y | Z = z_0]}{E[T | Z = z_1] - E[T | Z = z_0]} \\ &= \sum_{j=1}^{N_T} \omega_{t_j, t_{j-1}} E[Y(t_j) - Y(t_{j-1}) | \mathbf{S} \in \mathcal{S}_{t_j, t_{j-1}}] \end{aligned} \tag{B.1}$$

where  $\mathcal{S}_{t_j, t_{j-1}} \equiv \{\mathbf{s} \in \mathcal{S}; \mathbf{s}[z_1] \geq t_j > \mathbf{s}[z_0]\}$ , that is the sets of response types for whom a change in instrument receipt from  $z_0$  to  $z_1$  induces a change in treatment from strictly “below”  $t_j$  to weakly “above”  $t_j$ . The weights  $\omega_{t_j, t_{j-1}}$  are positive and given:

$$\omega_{t_j, t_{j-1}} = \frac{\Pr(\mathbf{S} \in \mathcal{S}_{t_j, t_{j-1}})}{\sum_{j=1}^{N_T} \Pr(\mathbf{S} \in \mathcal{S}_{t_j, t_{j-1}})}. \tag{B.2}$$

Unordered monotonicity also allows 2SLS type estimands to be expressed in terms of a weighted average of LATE parameters with positive weights.<sup>3</sup> In this setting the 2SLS numerator can be decomposed

$$\begin{aligned}\beta_{2\text{SLS}}^u(z, z') &= E[Y | Z = z] - E[Y | Z = z'] \\ &= \sum_{\{t, t'\}, t \neq t'} \omega_{t, t'}^u E[Y(t) - Y(t') | \mathbf{S} \in \mathcal{S}_{t, t'}(z, z')]\end{aligned}\quad (\text{B.3})$$

where  $\mathcal{S}_{t, t'}(z, z') \equiv \{\mathbf{s} : \mathbf{s}[z] = t, \mathbf{s}[z'] = t'\}$  is the set of response types that switch treatments from  $t$  to  $t'$  as the instrument varies from  $z$  to  $z'$ . Under unordered monotonicity, the same instrument switch cannot induce some agents to switch towards choice  $t$  while inducing others to switch away from choice  $t$ . So, we must have either  $\mathcal{S}_{t, t'}(z, z') = \emptyset$  or  $\mathcal{S}_{t', t}(z, z') = \emptyset$  (or both). The weights  $\omega_{t, t'}^u$  are weakly positive and given  $\omega_{t, t'}^u = \Pr(\mathbf{S} \in \mathcal{S}_{t, t'})$ .

## B.2 General Unique Decomposition

Using the identification equality in (9) we can rewrite

$$\begin{aligned}E[Y | Z = z] &= \sum_{\mathbf{s} \in \text{supp}(\mathbf{S})} \sum_{t \in \mathcal{T}} \mathbf{1}[\mathbf{s}[z] = t] E[Y(t) | \mathbf{S} = \mathbf{s}] \Pr(\mathbf{S} = \mathbf{s}) \\ E[Y | Z = z'] &= \sum_{\mathbf{s} \in \text{supp}(\mathbf{S})} \sum_{t \in \mathcal{T}} \mathbf{1}[\mathbf{s}[z'] = t] E[Y(t) | \mathbf{S} = \mathbf{s}] \Pr(\mathbf{S} = \mathbf{s})\end{aligned}$$

Using these, we can express the quasi-2SLS estimand as the following

$$\beta_{z, z'} \equiv E[Y | Z = z] - E[Y | Z = z'] \quad (\text{B.4})$$

$$= \sum_{\mathbf{s} \in \text{supp}(\mathbf{S})} \sum_{t \in \mathcal{T}} (\mathbf{1}[\mathbf{s}[z] = t] - \mathbf{1}[\mathbf{s}[z'] = t]) E[Y(t) | \mathbf{S} = \mathbf{s}] \Pr(\mathbf{S} = \mathbf{s}) \quad (\text{B.5})$$

$$= \sum_{\{t, t'\}, t \neq t'} E[Y(t) - Y(t') | \mathbf{S} \in \mathcal{S}_{t, t'}(z, z')] \Pr(\mathbf{S} \in \mathcal{S}_{t, t'}(z, z')), \quad (\text{B.6})$$

where the last equality is due to the fact that sets  $\mathcal{S}_{t, t'}(z, z')$ , defined below (B.3), form a partition of  $\text{supp}(\mathbf{S})$  as  $t, t'$  ranges in  $\mathcal{T}$ . Equation (B.6) holds regardless of any monotonicity assumption. That is, no matter what the restriction is on the support of  $\mathbf{S}$ , we will always be able to rewrite the 2SLS numerator as in (B.6).

In view of Lemma 1, a violation of MM is equivalent to there being a pair of treatments  $t, t'$  such

<sup>3</sup>In addition, [Buchinsky and Pinto \(2021\)](#) show that any variation in the instrumental variable can be used to identify a meaningful counterfactual outcome mean. For instance, the 2SLS estimate that uses a choice indicator for  $t$  and any IV-values  $z, z' \in \mathcal{Z}$ , such that  $P(T = t | Z = z) > P(T = t' | Z = z)$ , identifies the following parameter:

$$\beta_{2\text{SLS}}^t(z, z') = \frac{E[Y \mathbf{1}[T = t] | Z = z] - E[Y \mathbf{1}[T = t] | Z = z']}{\Pr(T = t | Z = z) - \Pr(T = t' | Z = z')} = E(Y(t) | \mathbf{S} \in \mathcal{S}_{z, z'}^t),$$

where  $\mathcal{S}_{z, z'}^t = \{\mathbf{s} : \mathbf{s}[z] = t, \mathbf{s}[z'] \neq t\}$  is the set of response types that switch from treatment choice  $t$  to any other treatment choice as the instrument varies from  $z$  to  $z'$ .

that the sets  $\mathcal{S}_{t,t'}(z, z')$  and  $\mathcal{S}_{t',t}(z, z')$  are *both* nonempty. This induces negative weights in the 2SLS estimand; both  $E[Y(t) - Y(t') \mid \mathcal{S}_{t,t'}(z, z')]$  and  $E[Y(t') - Y(t) \mid \mathcal{S}_{t',t}(z, z')]$  are represented in the decomposition (B.6). This in turn limits our ability to use the 2SLS estimand to gain useful insight into the direction of causal effects. The partial minimal monotonicity criterion is then crucial for interpreting  $\beta_{z,z'}$  as a type of interpretable causal parameter defined in (26).

### C Ordered vs. Unordered Example

We consider a setup with three treatments,  $\mathcal{T} = \{t_1, t_2, t_3\}$ , and three instruments,  $\mathcal{Z} = \{z_1, z_2, z_3\}$ . Response matrices (C.1)–(C.2) below are useful to understand the difference between ordered and unordered monotonicity conditions:

$$\mathbf{R}_1 = \begin{array}{c} \overbrace{\begin{array}{cccccc} \mathbf{s}_1 & \mathbf{s}_2 & \mathbf{s}_3 & \mathbf{s}_4 & \mathbf{s}_5 & \mathbf{s}_6 & \mathbf{s}_7 \end{array}}^{\text{Ordered but NOT Unordered}} \\ \left[ \begin{array}{cccccc|c} t_1 & t_2 & t_3 & t_1 & t_1 & t_2 & t_1 \\ t_1 & t_2 & t_3 & t_2 & t_3 & t_2 & t_1 \\ t_1 & t_2 & t_3 & t_3 & t_3 & t_3 & t_2 \end{array} \right] \begin{array}{l} z_1 \\ z_2 \\ z_3 \end{array} \end{array} \quad (\text{C.1})$$

$$\mathbf{R}_2 = \begin{array}{c} \overbrace{\begin{array}{cccccc} \mathbf{s}_1 & \mathbf{s}_2 & \mathbf{s}_3 & \mathbf{s}_4 & \mathbf{s}_5 & \mathbf{s}_6 \end{array}}^{\text{Ordered AND Unordered}} \\ \left[ \begin{array}{cccccc|c} t_1 & t_2 & t_3 & t_1 & t_1 & t_2 & t_1^* \\ t_1 & t_2 & t_3 & t_2 & t_3 & t_2 & t_2 \\ t_1 & t_2 & t_3 & t_3 & t_3 & t_3 & t_1 \end{array} \right] \begin{array}{l} z_1 \\ z_2 \\ z_3 \end{array} \end{array} \quad (\text{C.2})$$

\underbrace{\hspace{15em}}\_{\text{Unordered but NOT Ordered}}

Columns  $\mathbf{s}_1, \dots, \mathbf{s}_7$  of response matrix  $\mathbf{R}_1$  (C.1) denote response-types. Each column describes the sequence of counterfactual choices,  $(T_i(z_1), T_i(z_2), T_i(z_3))$ , for an agent in that column's response type. The counterfactual treatment in each of these sequences is weakly increasing with respect to the ordering  $t_1 \leq t_2 \leq t_3$ ; for any agent  $i \in \mathcal{I}$ ,  $T_i(z_1) \leq T_i(z_2) \leq T_i(z_3)$ . Thus OM-Sequence (16) holds.

However, response types  $\mathbf{s}_6, \mathbf{s}_7$  in  $\mathbf{R}_1$  violate the sequential representation of unordered monotonicity in (17) for choice  $t_2$ . Consider two agents  $i, i' \in \mathcal{I}$  such that  $\mathbf{S}_i = \mathbf{s}_6$  and  $\mathbf{S}_{i'} = \mathbf{s}_7$ . The sequence of  $t_2$ -indicators for agent  $i$  is weakly decreasing while the sequence for agent  $i'$  is weakly increasing

$$\begin{aligned} (\mathbf{1}[T_i(z_1) = t_2], \mathbf{1}[T_i(z_2) = t_2], \mathbf{1}[T_i(z_3) = t_2]) &= (1, 1, 0) \\ (\mathbf{1}[T_{i'}(z_1) = t_2], \mathbf{1}[T_{i'}(z_2) = t_2], \mathbf{1}[T_{i'}(z_3) = t_2]) &= (0, 0, 1). \end{aligned}$$

This represents a violation of UM-Sequence (17) for the sequencing of  $\mathcal{Z}$ ,  $(z_1, z_2, z_3)$ . Moreover, because the switch from  $z_2$  to  $z_3$  induces agent  $i$  to move strictly away from treatment choice  $t_2$  while moving agent  $i'$  strictly towards treatment choice  $t_2$ , there is no other sequencing of  $\mathcal{Z}$  that

would satisfy the requirement of UM Sequence (17). We can conclude that the response matrix  $\mathbf{R}_1$  does not satisfy unordered monotonicity.

Response matrix  $\mathbf{R}_2$  (C.2) replaces  $\mathbf{s}_7$  in  $\mathbf{R}_1$  with  $\mathbf{s}_7^*$ . The treatment indexes in  $\mathbf{s}_7^*$  are not weakly increasing with respect to the ordering  $t_1 \leq t_2 \leq t_3$ . Thus the ordered monotonicity that held for  $\mathbf{R}_1$  does not hold for  $\mathbf{R}_2$ . Indeed, the reader can confirm that there is no ordering  $\mathcal{T}$  that satisfies OM-Sequence (16). In this case, though, the response matrix  $\mathbf{R}_2$  satisfies unordered monotonicity. Equations (C.3)–(C.5) are instructive in establishing this fact.

$$\text{Reordered rows of } \mathbf{R}_2 \text{ for } t_1 = \begin{bmatrix} \mathbf{s}_1 & \mathbf{s}_2 & \mathbf{s}_3 & \mathbf{s}_4 & \mathbf{s}_5 & \mathbf{s}_6 & \mathbf{s}_7^* \\ t_1 & t_2 & t_3 & t_2 & t_3 & t_2 & t_2 \\ t_1 & t_2 & t_3 & t_3 & t_3 & t_3 & t_1 \\ t_1 & t_2 & t_3 & t_1 & t_1 & t_2 & t_1 \end{bmatrix} \begin{matrix} z_2 \\ z_3 \\ z_1 \end{matrix}, \quad (\text{C.3})$$

$$\text{Reordered rows of } \mathbf{R}_2 \text{ for } t_2 = \begin{bmatrix} \mathbf{s}_1 & \mathbf{s}_2 & \mathbf{s}_3 & \mathbf{s}_4 & \mathbf{s}_5 & \mathbf{s}_6 & \mathbf{s}_7^* \\ t_1 & t_2 & t_3 & t_3 & t_3 & t_3 & t_1 \\ t_1 & t_2 & t_3 & t_1 & t_1 & t_2 & t_1 \\ t_1 & t_2 & t_3 & t_2 & t_3 & t_2 & t_2 \end{bmatrix} \begin{matrix} z_3 \\ z_1 \\ z_2 \end{matrix}, \quad (\text{C.4})$$

$$\text{Reordered rows of } \mathbf{R}_2 \text{ for } t_3 = \begin{bmatrix} \mathbf{s}_1 & \mathbf{s}_2 & \mathbf{s}_3 & \mathbf{s}_4 & \mathbf{s}_5 & \mathbf{s}_6 & \mathbf{s}_7^* \\ t_1 & t_2 & t_3 & t_1 & t_1 & t_2 & t_1 \\ t_1 & t_2 & t_3 & t_2 & t_3 & t_2 & t_2 \\ t_1 & t_2 & t_3 & t_3 & t_3 & t_3 & t_1 \end{bmatrix} \begin{matrix} z_1 \\ z_2 \\ z_3 \end{matrix}. \quad (\text{C.5})$$

Equation (C.3) reorders the rows of  $\mathbf{R}_2$  in (C.2) from  $(z_1, z_2, z_3)$  to  $(z_2, z_3, z_1)$ . At each move along the sequence  $(z_2, z_3, z_1)$ , additional response types switch to treatment  $t_1$  and no response types switch away from  $t_1$ . This means that

$$\mathbf{1}[T_i(z_2) = t_1] \leq \mathbf{1}[T_i(z_3) = t_1] \leq \mathbf{1}[T_i(z_1) = t_1]$$

holds for all agents  $i \in \mathcal{I}$ , regardless of response type. Thus UM Sequence (17) holds for  $t_1$ . By symmetric logic, equation (C.4) demonstrates that UM Sequence (17) holds for  $t_2$  using the IV sequence  $(z_3, z_1, z_2)$  and equation (C.5) shows that UM Sequence holds for  $t_3$  using sequence  $(z_1, z_2, z_3)$ . We conclude that unordered monotonicity holds.

Response matrices  $\mathbf{R}_1$  and  $\mathbf{R}_2$  in (C.1)–(C.2) show that ordered monotonicity does not imply unordered monotonicity nor vice-versa. Ordered monotonicity holds for  $\mathbf{R}_1$  but not for  $\mathbf{R}_2$  while unordered monotonicity holds for  $\mathbf{R}_2$  but not for  $\mathbf{R}_1$ . The two monotonicity conditions can intersect, both ordered and unordered monotonicity hold for the submatrix generated by response-types  $\mathbf{s}_1$  to  $\mathbf{s}_6$ .

## D Additional Information Regarding the Examples of Section 7

### D.1 Verifying Unordered Monotonicity

We seek to show that the response matrix (38) is a case of UM (17) using the verification matrix of item (iv) of Theorem 1. The matrix is presented below for convenience.

$$\mathbf{R} = \begin{array}{cccccccc} & \mathbf{s}_1 & \mathbf{s}_2 & \mathbf{s}_3 & \mathbf{s}_4 & \mathbf{s}_5 & \mathbf{s}_6 & \mathbf{s}_7 & \\ \begin{bmatrix} t_1 & t_1 & t_1 & t_1 & t_2 & t_2 & t_3 \\ t_1 & t_1 & t_3 & t_3 & t_2 & t_3 & t_3 \\ t_1 & t_2 & t_2 & t_3 & t_2 & t_2 & t_3 \\ t_1 & t_1 & t_1 & t_1 & t_2 & t_2 & t_3 \end{bmatrix} & z_1 \\ & & & & & & & & z_2 \\ & & & & & & & & z_3 \\ & & & & & & & & z_4 \end{array}$$

Let  $\mathbf{B}_t = \mathbf{1}[\mathbf{R} = t]; t \in \{t_1, t_2, t_3\}$  denote the binary matrices corresponding to response matrix (38). Those are displayed below:

$$\mathbf{B}_{t_1} = \begin{array}{cccccccc} & \mathbf{s}_1 & \mathbf{s}_2 & \mathbf{s}_3 & \mathbf{s}_4 & \mathbf{s}_5 & \mathbf{s}_6 & \mathbf{s}_7 & \\ \begin{bmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 \end{bmatrix} & z_1 \\ & & & & & & & & z_2 \\ & & & & & & & & z_3 \\ & & & & & & & & z_4 \end{array}$$

$$\mathbf{B}_{t_2} = \begin{array}{cccccccc} & \mathbf{s}_1 & \mathbf{s}_2 & \mathbf{s}_3 & \mathbf{s}_4 & \mathbf{s}_5 & \mathbf{s}_6 & \mathbf{s}_7 & \\ \begin{bmatrix} 0 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 \end{bmatrix} & z_1 \\ & & & & & & & & z_2 \\ & & & & & & & & z_3 \\ & & & & & & & & z_4 \end{array}$$

$$\mathbf{B}_{t_3} = \begin{array}{cccccccc} & \mathbf{s}_1 & \mathbf{s}_2 & \mathbf{s}_3 & \mathbf{s}_4 & \mathbf{s}_5 & \mathbf{s}_6 & \mathbf{s}_7 & \\ \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} & z_1 \\ & & & & & & & & z_2 \\ & & & & & & & & z_3 \\ & & & & & & & & z_4 \end{array}$$

Unordered monotonicity holds if and only if the binary matrices  $\mathbf{B}_{t_1}, \mathbf{B}_{t_2}, \mathbf{B}_{t_3}$  are lonesum. For item (iv) of Theorem 1 to hold, it suffices to show that  $\|\Psi_U(t)\| = 0$  for all  $t \in \{t_1, t_2, t_3\}$  where  $\Psi_U(t)$  is given by  $\Psi_U(t) \equiv ((\mathbf{1} - \mathbf{B}_t)^\top \mathbf{B}_t) \odot ((\mathbf{1} - \mathbf{B}_t)^\top \mathbf{B}_t)^\top$ . It is useful to express  $\Psi_U(t_1)$  as  $\Psi_U(t) = \tilde{\Psi}_U(t) \odot \tilde{\Psi}_U(t)^\top$  where  $\tilde{\Psi}_U(t) = ((\mathbf{1} - \mathbf{B}_t)^\top \mathbf{B}_t)$ .

The matrices  $\tilde{\Psi}_U(t_1), \tilde{\Psi}_U(t_2), \tilde{\Psi}_U(t_3)$  are computed below. Note that  $\|\Psi_U(t)\| = 0$  if  $\tilde{\Psi}_U(t)$  is a triangular matrix with a zero diagonal. Thus it suffices to evaluate matrix  $\tilde{\Psi}_U(t)$  for  $t \in \{t_1, t_2, t_3\}$ .

$$\tilde{\Psi}_U(t_1) = \underbrace{\begin{bmatrix} 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 \end{bmatrix}^\top}_{(\mathbf{1} - \mathbf{B}_{t_1})^\top} \underbrace{\begin{bmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 \end{bmatrix}}_{\mathbf{B}_{t_1}} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 2 & 1 & 0 & 0 & 0 & 0 & 0 \\ 2 & 1 & 0 & 0 & 0 & 0 & 0 \\ 4 & 3 & 2 & 2 & 0 & 0 & 0 \\ 4 & 3 & 2 & 2 & 0 & 0 & 0 \\ 4 & 3 & 2 & 2 & 0 & 0 & 0 \end{bmatrix}$$



$$\begin{aligned}
\tilde{\Psi}_{\mathcal{U}}(t_2) &= \underbrace{\begin{bmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 1 \\ 1 & 1 & 1 & 1 & 0 & 1 & 1 \\ 1 & 0 & 0 & 1 & 0 & 0 & 1 \\ 1 & 1 & 1 & 1 & 0 & 0 & 1 \end{bmatrix}^\top}_{(1-\mathcal{B}_{t_2})^\top} \underbrace{\begin{bmatrix} 0 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 \end{bmatrix}}_{\mathcal{B}_{t_2}} = \begin{bmatrix} 0 & 1 & 1 & 0 & 4 & 3 & 0 \\ 0 & 0 & 0 & 0 & 3 & 2 & 0 \\ 0 & 0 & 0 & 0 & 3 & 2 & 0 \\ 0 & 1 & 1 & 0 & 4 & 3 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 4 & 3 & 0 \end{bmatrix} \\
\tilde{\Psi}_{\mathcal{U}}(t_3) &= \underbrace{\begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 & 0 \end{bmatrix}^\top}_{(1-\mathcal{B}_{t_3})^\top} \underbrace{\begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}}_{\mathcal{B}_{t_3}} = \begin{bmatrix} 0 & 0 & 1 & 2 & 0 & 1 & 4 \\ 0 & 0 & 1 & 2 & 0 & 1 & 4 \\ 0 & 0 & 0 & 1 & 0 & 0 & 3 \\ 0 & 0 & 0 & 0 & 0 & 0 & 2 \\ 0 & 0 & 1 & 2 & 0 & 1 & 4 \\ 0 & 0 & 0 & 1 & 0 & 0 & 3 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}
\end{aligned}$$

It is easy to see that in  $\tilde{\Psi}_{\mathcal{U}}(t_1) \odot \tilde{\Psi}_{\mathcal{U}}(t_1)^\top$  is equal to a matrix of zeros. Indeed, the matrix  $\tilde{\Psi}_{\mathcal{U}}(t_1)$  is triangular with a zero diagonal. Thus, when we perform the element wise multiplication of  $\tilde{\Psi}_{\mathcal{U}}(t_1)$  and its transpose, at least one of the elements of the multiplication will be zero. The same occurs for matrices  $\tilde{\Psi}_{\mathcal{U}}(t_2)$  and  $\tilde{\Psi}_{\mathcal{U}}(t_3)$ .

## D.2 A Case of Choice Incentives for Ordered Monotonicity

The can summarize the above incentive structure the binary incentive matrix given below:

$$\mathbf{L} = \begin{array}{ccc} & t_1 & t_2 & t_3 \\ \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix} & z_1 & z_2 & z_3 & z_4 \end{array} \quad (\text{D.1})$$

We use this first example to describe the machinery that translates choice incentives into monotonicity conditions and identification results. We adopt a more parsimonious approach in the subsequent examples.

Choice rule (35) converts the Incentive Matrix (D.1) into choice restrictions that determine the model response matrix  $\mathbf{R}$ . These choice restrictions are displayed in Table 4. Choice restrictions in Table 4 are in turn used to eliminate the response-types that are not economically justifiable.

Each counterfactual choice  $T(z)$  of the response vector  $\mathbf{S} = [T(z_1), T(z_2), T(z_3), T(z_4)]'$  takes up to three values in  $\{t_1, t_2, t_3\}$ . Thus, there are  $3^4 = 81$  potential response types. The combination of all choice restrictions of Table 4 eliminate a total of 74 out of the 81 potential response-types. Response matrix  $\mathbf{R}$  in (D.2) displays the resulting seven response-types that survive the elimination process.

Table 4: Choice Restrictions generated by Incentive Matrix (D.1)

1	$T_i(z_1) = t_1 \Rightarrow \emptyset$
2	$T_i(z_2) = t_1 \Rightarrow T_i(z_1) \notin \{t_2, t_3\}$ and $T_i(z_3) \notin \{t_2, t_3\}$ and $T_i(z_4) \neq t_2$
3	$T_i(z_3) = t_1 \Rightarrow T_i(z_1) \notin \{t_2, t_3\}$ and $T_i(z_2) \notin \{t_2, t_3\}$ and $T_i(z_4) \neq t_2$
4	$T_i(z_4) = t_1 \Rightarrow T_i(z_1) \notin \{t_2, t_3\}$ and $T_i(z_2) \notin \{t_2, t_3\}$ and $T_i(z_3) \notin \{t_2, t_3\}$
5	$T_i(z_1) = t_2 \Rightarrow T_i(z_2) \notin \{t_1, t_3\}$ and $T_i(z_3) \notin \{t_1, t_3\}$ and $T_i(z_4) \neq t_1$
6	$T_i(z_2) = t_2 \Rightarrow T_i(z_1) \neq t_3$ and $T_i(z_3) \notin \{t_1, t_3\}$ and $T_i(z_4) \neq t_1$
7	$T_i(z_3) = t_2 \Rightarrow T_i(z_1) \neq t_3$ and $T_i(z_2) \notin \{t_1, t_3\}$ and $T_i(z_4) \neq t_1$
8	$T_i(z_4) = t_2 \Rightarrow T_i(z_1) \neq t_3$ and $T_i(z_2) \notin \{t_1, t_3\}$ and $T_i(z_3) \notin \{t_1, t_3\}$
9	$T_i(z_1) = t_3 \Rightarrow T_i(z_2) \notin \{t_1, t_2\}$ and $T_i(z_3) \notin \{t_1, t_2\}$ and $T_i(z_4) \notin \{t_1, t_2\}$
10	$T_i(z_2) = t_3 \Rightarrow T_i(z_1) \neq t_2$ and $T_i(z_3) \notin \{t_1, t_2\}$ and $T_i(z_4) \notin \{t_1, t_2\}$
11	$T_i(z_3) = t_3 \Rightarrow T_i(z_1) \neq t_2$ and $T_i(z_2) \notin \{t_1, t_2\}$ and $T_i(z_4) \notin \{t_1, t_2\}$
12	$T_i(z_4) = t_3 \Rightarrow \emptyset$

This table presents all the choice restrictions generated by applying the choice rule (35) to each of the combination of choices  $(t, t') \in \{t_1, t_2, t_3\}$  and instrumental values  $(z, z') \in \{z_1, z_2, z_3, z_4\}$  of the incentive matrix (D.1).

$$\mathbf{R} = \begin{bmatrix} \mathbf{s}_1 & \mathbf{s}_2 & \mathbf{s}_3 & \mathbf{s}_4 & \mathbf{s}_5 & \mathbf{s}_6 & \mathbf{s}_7 & \mathbf{s}_8 \\ t_1 & t_1 & t_1 & t_1 & t_1 & t_2 & t_2 & t_3 \\ t_1 & t_1 & t_2 & t_2 & t_3 & t_2 & t_2 & t_3 \\ t_1 & t_1 & t_2 & t_2 & t_3 & t_2 & t_2 & t_3 \\ t_1 & t_3 & t_2 & t_3 & t_3 & t_2 & t_1 & t_3 \end{bmatrix} \begin{matrix} z_1 \\ z_2 \\ z_3 \\ z_4 \end{matrix} \quad (\text{D.2})$$

We use equations (12)–(13) to evaluate the causal parameters identified by response matrix (39). The response-matrix (39) enables the identification of eight response-type probabilities:

Point Identified	$P(\mathbf{S} = \mathbf{s}_1), P(\mathbf{S} = \mathbf{s}_2), P(\mathbf{S} = \mathbf{s}_5), P(\mathbf{S} = \mathbf{s}_8).$
Partially Identified	$P(\mathbf{S} \in \{\mathbf{s}_3, \mathbf{s}_4\}), P(\mathbf{S} \in \{\mathbf{s}_3, \mathbf{s}_6\}), P(\mathbf{S} \in \{\mathbf{s}_4, \mathbf{s}_7\}), P(\mathbf{S} \in \{\mathbf{s}_6, \mathbf{s}_7\}).$

as well as the following counterfactual outcomes.

Always-takers	$E(Y(t_1) \mathbf{S} = \mathbf{s}_1)$	-	$E(Y(t_3) \mathbf{S} = \mathbf{s}_8)$
Switchers	$E(Y(t_1) \mathbf{S} = \mathbf{s}_2)$	-	$E(Y(t_3) \mathbf{S} = \mathbf{s}_5)$
Partially Identified	$E(Y(t_1) \mathbf{S} \in \{\mathbf{s}_3, \mathbf{s}_4, \mathbf{s}_5\})$	$E(Y(t_2) \mathbf{S} \in \{\mathbf{s}_4, \mathbf{s}_7\})$	$E(Y(t_3) \mathbf{S} \in \{\mathbf{s}_2, \mathbf{s}_4, \mathbf{s}_7\})$
		$E(Y(t_2) \mathbf{S} \in \{\mathbf{s}_6, \mathbf{s}_7\})$	
		$E(Y(t_2) \mathbf{S} \in \{\mathbf{s}_3, \mathbf{s}_6\})$	
		$E(Y(t_2) \mathbf{S} \in \{\mathbf{s}_3, \mathbf{s}_4\})$	

The identification results above state that only four out of nine response-type probabilities are point-identified. Most of the counterfactual outcomes are partially identified. Only four counterfactual outcome means are point-identified, none of these for choice  $t_2$ . In contrast, the unordered response matrix (38) in Lemma ?? secures the point-identification of all response-type probabilities and most of the counterfactual outcome means.

### D.3 MM under the Double Randomization Design

We consider the emergence of MM in a “Double Randomization” design in which two vouchers are randomly assigned to the same sample of prospective students. The first voucher offers a tuition discount that applies to a natural science major. The second one applies to social science majors. We can divide the students into four groups:

1. Group  $z_1$  does not receive any voucher.
2. Group  $z_2$  receives only the social sciences voucher ( $t_3$ ).
3. Group  $z_3$  receives only the natural sciences voucher ( $t_2$ ).
4. Group  $z_4$  receives both the social sciences and natural sciences voucher.

Assuming the social sciences and natural sciences vouchers are of the same amount and that students cannot double major (so that they can only apply one voucher at a time), the IV design described above can be summarized by the incentive matrix in (D.3).

$$\mathbf{L} = \begin{array}{ccc|c} & t_1 & t_2 & t_3 \\ \hline & 0 & 0 & 0 \\ & 0 & 0 & 1 \\ & 0 & 1 & 0 \\ & 0 & 1 & 1 \end{array} \begin{array}{l} z_1 \\ z_2 \\ z_3 \\ z_4 \end{array} \quad (\text{D.3})$$

Table 5: Choice Restrictions generated by Incentive Matrix (40)

1	$T_i(z_1) = t_1 \Rightarrow T_i(z_2) \neq t_2 \text{ and } T_i(z_3) \neq t_3$
2	$T_i(z_2) = t_1 \Rightarrow T_i(z_1) \notin \{t_2, t_3\} \text{ and } T_i(z_3) \neq t_3 \text{ and } T_i(z_4) \neq t_3$
3	$T_i(z_3) = t_1 \Rightarrow T_i(z_1) \notin \{t_2, t_3\} \text{ and } T_i(z_2) \neq t_2 \text{ and } T_i(z_4) \neq t_2$
4	$T_i(z_4) = t_1 \Rightarrow T_i(z_1) \notin \{t_2, t_3\} \text{ and } T_i(z_2) \notin \{t_2, t_3\} \text{ and } T_i(z_3) \notin \{t_2, t_3\}$
5	$T_i(z_1) = t_2 \Rightarrow T_i(z_2) \neq t_1 \text{ and } T_i(z_3) \notin \{t_1, t_3\} \text{ and } T_i(z_4) \notin \{t_1, t_3\}$
6	$T_i(z_2) = t_2 \Rightarrow T_i(z_1) \notin \{t_1, t_3\} \text{ and } T_i(z_3) \notin \{t_1, t_3\} \text{ and } T_i(z_4) \notin \{t_1, t_3\}$
7	$T_i(z_3) = t_2 \Rightarrow T_i(z_4) \neq t_1$
8	$T_i(z_4) = t_2 \Rightarrow T_i(z_1) \neq t_3 \text{ and } T_i(z_3) \notin \{t_1, t_3\}$
9	$T_i(z_1) = t_3 \Rightarrow T_i(z_2) \notin \{t_1, t_2\} \text{ and } T_i(z_3) \neq t_1 \text{ and } T_i(z_4) \notin \{t_1, t_2\}$
10	$T_i(z_2) = t_3 \Rightarrow T_i(z_4) \neq t_1$
11	$T_i(z_3) = t_3 \Rightarrow T_i(z_1) \notin \{t_1, t_2\} \text{ and } T_i(z_2) \notin \{t_1, t_2\} \text{ and } T_i(z_4) \notin \{t_1, t_2\}$
12	$T_i(z_4) = t_3 \Rightarrow T_i(z_1) \neq t_2 \text{ and } T_i(z_2) \notin \{t_1, t_2\}$

This table presents all the choice restrictions generated by applying the choice rule (35) to each of the combination of choices  $(t, t') \in \{t_1, t_2, t_3\}$  and instrumental values  $(z, z') \in \{z_1, z_2, z_3, z_4\}$  of the incentive matrix (40).

Applying the Choice Rule (35) from above generates the choice restrictions of Table 5. These in turn generate the response matrix  $\mathbf{R}$  in (D.4).

$$\mathbf{R} = \begin{matrix} & \mathbf{s}_1 & \mathbf{s}_2 & \mathbf{s}_3 & \mathbf{s}_4 & \mathbf{s}_5 & \mathbf{s}_6 & \mathbf{s}_7 & \mathbf{s}_8 & \mathbf{s}_9 \\ \begin{bmatrix} t_1 & t_1 & t_1 & t_1 & t_1 & t_2 & t_2 & t_3 & t_3 \\ t_1 & t_1 & t_3 & t_3 & t_3 & t_2 & t_3 & t_3 & t_3 \\ t_1 & t_2 & t_1 & t_2 & t_2 & t_2 & t_2 & t_2 & t_3 \\ t_1 & t_2 & t_3 & t_2 & t_3 & t_2 & t_2 & t_3 & t_3 \end{bmatrix} & z_1 \\ & & & & & & & & & z_2 \\ & & & & & & & & & z_3 \\ & & & & & & & & & z_4 \end{matrix} \quad (\text{D.4})$$

Applying equations (12)–(13) to response-matrix (D.4) gives that all response-type probabilities are identified,  $P(\mathbf{S} = \mathbf{s}_j)$ ;  $j = 1, \dots, 9$ , as well as the following counterfactual outcomes:

Always-takers	$E(Y(t_0) \mathbf{S} = \mathbf{s}_1)$	$E(Y(t_1) \mathbf{S} = \mathbf{s}_6)$	$E(Y(t_2) \mathbf{S} = \mathbf{s}_9)$
Switchers	$E(Y(t_0) \mathbf{S} = \mathbf{s}_2)$ $E(Y(t_0) \mathbf{S} = \mathbf{s}_3)$	$E(Y(t_1) \mathbf{S} = \mathbf{s}_7)$	$E(Y(t_2) \mathbf{S} = \mathbf{s}_8)$
Partially Identified	$E(Y(t_0) \mathbf{S} \in \{\mathbf{s}_4, \mathbf{s}_5\})$	$E(Y(t_1) \mathbf{S} \in \{\mathbf{s}_2, \mathbf{s}_4\})$ $E(Y(t_1) \mathbf{S} \in \{\mathbf{s}_5, \mathbf{s}_8\})$	$E(Y(t_2) \mathbf{S} \in \{\mathbf{s}_3, \mathbf{s}_5\})$ $E(Y(t_2) \mathbf{S} \in \{\mathbf{s}_4, \mathbf{s}_7\})$

#### D.4 MM under the Extensive Margin Compliers Only (EMCO) Design

We revisit the incentive design described in (41), presented again in  $\mathbf{L}$  (D.5) below

$$\mathbf{L} = \begin{matrix} & t_0 & t_1 & t_2 \\ \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} & z_1 \\ & & & z_2 \\ & & & z_3 \\ & & & z_4 \end{matrix} \quad (\text{D.5})$$

Applying the Choice Rule (35) to the incentive design summarized in (D.5) we generate the following choice restrictions. These choice restrictions will in turn be used to eliminate response types, i.e restrict  $\text{supp}(\mathbf{S})$ .

After exhausting the choice restrictions in Table 6 we are left with 7 out of a possible 81 response types. These response types are consolidated and displayed in the response matrix  $\mathbf{R}$  (D.6) below.

$$\mathbf{R} = \begin{matrix} & \mathbf{s}_1 & \mathbf{s}_2 & \mathbf{s}_3 & \mathbf{s}_4 & \mathbf{s}_5 & \mathbf{s}_6 & \mathbf{s}_7 \\ \begin{bmatrix} t_1 & t_1 & t_1 & t_2 & t_2 & t_3 & t_3 \\ t_1 & t_1 & t_1 & t_1 & t_2 & t_1 & t_3 \\ t_1 & t_2 & t_3 & t_2 & t_2 & t_3 & t_3 \\ t_1 & t_1 & t_1 & t_2 & t_2 & t_3 & t_3 \end{bmatrix} & z_1 \\ & & & & & & & z_2 \\ & & & & & & & z_3 \\ & & & & & & & z_4 \end{matrix} \quad (\text{D.6})$$

We can apply the identification results of Heckman and Pinto (2018) to response-matrix (D.6) in order to identify all the response-type probabilities  $P(\mathbf{S} = \mathbf{s}_j)$ ;  $j = 1, \dots, 7$  as well as the following counterfactual outcomes:

Table 6: Choice Restrictions generated by Incentive Matrix (D.5)

1	$T_i(z_1) = t_1 \Rightarrow T_i(z_2) \notin \{t_2, t_3\}$ and $T_i(z_4) \notin \{t_2, t_3\}$
2	$T_i(z_2) = t_1 \Rightarrow \emptyset$
3	$T_i(z_3) = t_1 \Rightarrow T_i(z_1) \notin \{t_2, t_3\}$ and $T_i(z_2) \notin \{t_2, t_3\}$ and $T_i(z_4) \notin \{t_2, t_3\}$
4	$T_i(z_4) = t_1 \Rightarrow T_i(z_1) \notin \{t_2, t_3\}$ and $T_i(z_2) \notin \{t_2, t_3\}$
5	$T_i(z_1) = t_2 \Rightarrow T_i(z_2) \neq t_3$ and $T_i(z_3) \notin \{t_1, t_3\}$ and $T_i(z_4) \notin \{t_1, t_3\}$
6	$T_i(z_2) = t_2 \Rightarrow T_i(z_1) \notin \{t_1, t_3\}$ and $T_i(z_3) \notin \{t_1, t_3\}$ and $T_i(z_4) \notin \{t_1, t_3\}$
7	$T_i(z_3) = t_2 \Rightarrow T_i(z_1) \neq t_3$ and $T_i(z_2) \neq t_3$ and $T_i(z_4) \neq t_3$
8	$T_i(z_4) = t_2 \Rightarrow T_i(z_1) \notin \{t_1, t_3\}$ and $T_i(z_2) \neq t_3$ and $T_i(z_3) \notin \{t_1, t_3\}$
9	$T_i(z_1) = t_3 \Rightarrow T_i(z_2) \neq t_2$ and $T_i(z_3) \notin \{t_1, t_2\}$ and $T_i(z_4) \notin \{t_1, t_2\}$
10	$T_i(z_2) = t_3 \Rightarrow T_i(z_1) \notin \{t_1, t_2\}$ and $T_i(z_3) \notin \{t_1, t_2\}$ and $T_i(z_4) \notin \{t_1, t_2\}$
11	$T_i(z_3) = t_3 \Rightarrow T_i(z_1) \neq t_2$ and $T_i(z_2) \neq t_2$ and $T_i(z_4) \neq t_2$
12	$T_i(z_4) = t_3 \Rightarrow T_i(z_1) \notin \{t_1, t_2\}$ and $T_i(z_2) \neq t_2$ and $T_i(z_3) \notin \{t_1, t_2\}$

This table presents all the choice restrictions generated by applying the choice rule (35) to each of the combination of choices  $(t, t') \in \{t_1, t_2, t_3\}$  and instrumental values  $(z, z') \in \{z_1, z_2, z_3, z_4\}$  of the incentive matrix (D.5).

Always-takers	$E(Y(t_1) S = s_1)$	$E(Y(t_2) S = s_5)$	$E(Y(t_3) S = s_7)$
Switchers		$E(Y(t_2) S = s_2)$	$E(Y(t_3) S = s_3)$
		$E(Y(t_2) S = s_4)$	$E(Y(t_3) S = s_6)$
Partially Identified	$E(Y(t_1) S \in \{s_2, s_3\})$		
	$E(Y(t_1) S \in \{s_4, s_6\})$		

## D.5 MM under Orthogonal Array Design

We additionally examine an IV choice model based on the popular orthogonal array experimental design. Orthogonal arrays are a widely popular experimental design developed by CD Rao (Rao, 1946a,b, 1947, 1949). Orthogonal arrays are widely used in Agricultural and Industrial sciences to determine the optimum mix of treatments that maximize production yield. The method is based on the random assignment of a combinatorial arrangements of treatments for each randomization arm. We adapt this setup to an instrumental variable setting by exogenously providing incentives for one or more treatments instead of directly assigning agents to treatment arms. Below, we will see that this incentive structure allows for a broad range of identification results.

Formally, a binary orthogonal array is a matrix of zeros and ones such that any two-column submatrix displays all possible combinations of zeros and ones. In other words, the tuples

$$\{(0, 0), (0, 1), (1, 0), (1, 1)\}$$

are all rows in any two-column submatrix. An orthogonal array incentive design if its associated incentive matrix is a binary orthogonal array. The incentive matrix in (46) displays an example of an orthogonal array incentive design. In context of the college choice example, we can rationalize

the orthogonal array incentive design (46) with the following research design:

1. Group  $z_1$  receives a cash voucher if they choose to major in the natural sciences ( $t_2$ ) or the social sciences ( $t_3$ ).
2. Group  $z_2$  receives no cash voucher.
3. Group  $z_3$  receives a cash voucher if they do not go to college ( $t_1$ ) or if they major in the natural sciences ( $t_2$ ).
4. Group  $z_3$  receives a cash voucher if they do not go to college ( $t_1$ ) or if they major in the social sciences ( $t_3$ ).

Table 7 displays the choice restrictions generated by applying the Choice Rule (35) to the orthogonal array incentive design (46). After using these choice restrictions to eliminate response types, we are left with nine total response types summarized in the response matrix (D.7).

Table 7: Choice Restrictions generated by Incentive Matrix (46)

1	$T_i(z_1) = t_1 \Rightarrow T_i(z_2) \notin \{t_2, t_3\}$ and $T_i(z_3) \notin \{t_2, t_3\}$ and $T_i(z_4) \notin \{t_2, t_3\}$
2	$T_i(z_2) = t_1 \Rightarrow T_i(z_3) \notin \{t_2, t_3\}$ and $T_i(z_4) \notin \{t_2, t_3\}$
3	$T_i(z_3) = t_1 \Rightarrow T_i(z_2) \neq t_2$ and $T_i(z_4) \neq t_2$
4	$T_i(z_4) = t_1 \Rightarrow T_i(z_2) \neq t_3$ and $T_i(z_3) \neq t_3$
5	$T_i(z_1) = t_2 \Rightarrow T_i(z_2) \neq t_3$ and $T_i(z_3) \neq t_3$
6	$T_i(z_2) = t_2 \Rightarrow T_i(z_1) \notin \{t_1, t_3\}$ and $T_i(z_3) \notin \{t_1, t_3\}$
7	$T_i(z_3) = t_2 \Rightarrow T_i(z_1) \neq t_1$ and $T_i(z_2) \neq t_1$
8	$T_i(z_4) = t_2 \Rightarrow T_i(z_1) \notin \{t_1, t_3\}$ and $T_i(z_2) \notin \{t_1, t_3\}$ and $T_i(z_3) \notin \{t_1, t_3\}$
9	$T_i(z_1) = t_3 \Rightarrow T_i(z_2) \neq t_2$ and $T_i(z_4) \neq t_2$
10	$T_i(z_2) = t_3 \Rightarrow T_i(z_1) \notin \{t_1, t_2\}$ and $T_i(z_4) \notin \{t_1, t_2\}$
11	$T_i(z_3) = t_3 \Rightarrow T_i(z_1) \notin \{t_1, t_2\}$ and $T_i(z_2) \notin \{t_1, t_2\}$ and $T_i(z_4) \notin \{t_1, t_2\}$
12	$T_i(z_4) = t_3 \Rightarrow T_i(z_1) \neq t_1$ and $T_i(z_2) \neq t_1$

This table presents all the choice restrictions generated by applying the choice rule (35) to each of the combination of choices  $(t, t') \in \{t_1, t_2, t_3\}$  and instrumental values  $(z, z') \in \{z_1, z_2, z_3, z_4\}$  of the incentive matrix (46).

$$\mathbf{R} = \begin{matrix} & \mathbf{s}_1 & \mathbf{s}_2 & \mathbf{s}_3 & \mathbf{s}_4 & \mathbf{s}_5 & \mathbf{s}_6 & \mathbf{s}_7 & \mathbf{s}_8 & \mathbf{s}_9 & \\ \begin{bmatrix} t_1 & t_2 & t_2 & t_2 & t_2 & t_3 & t_3 & t_3 & t_3 \\ t_1 & t_1 & t_2 & t_2 & t_2 & t_1 & t_3 & t_3 & t_3 \\ t_1 & t_1 & t_2 & t_2 & t_2 & t_1 & t_1 & t_2 & t_3 \\ t_1 & t_1 & t_1 & t_2 & t_3 & t_1 & t_3 & t_3 & t_3 \end{bmatrix} & z_1 \\ & & & & & & & & & & z_2 \\ & & & & & & & & & & z_3 \\ & & & & & & & & & & z_4 \end{matrix} \tag{D.7}$$

This response matrix satisfies neither unordered nor ordered monotonicity. When the instrument switches from  $z_1$  to  $z_4$ , agents in response type  $\mathbf{s}_3$  move from treatment  $t_2$  to treatment  $t_3$  while agents in response type  $\mathbf{s}_6$  move away from  $t_3$  and towards  $t_1$ . This represents a violation of ordered

monotonicity and also prevents  $t_3$  from being ordered the highest or lowest in any ordering on  $\mathcal{T}$  that would satisfy ordered monotonicity.<sup>4</sup> Similarly we can see a switch from  $z_3$  to  $z_4$  induces agents in response type  $\mathbf{s}_3$  to move from treatment  $t_2$  to treatment  $t_1$  while inducing agents in response type  $\mathbf{s}_7$  to move away from treatment  $t_1$  and towards treatment  $t_3$ . This again represents a violation of unordered monotonicity and prevents  $t_1$  from being ordered either the highest or the lowest in any ordering  $\mathcal{T}$  that would satisfy ordered monotonicity. Since all orderings on  $\mathcal{T} = \{t_1, t_2, t_3\}$  must have either  $t_1$  or  $t_3$  as the largest or smallest element, this means there is no ordering on  $\mathcal{T}$  that satisfies ordered monotonicity.

Despite this, we can once again use Theorem 3 to verify that this matrix does indeed satisfy MM. Thus we can still use 2SLS type estimands to recover interpretable causal parameters as defined in (26). Moreover, by applying (12)–(13) we can see that all response types probabilities  $P(\mathbf{S} = \mathbf{s}_j)$ ,  $j = 1, \dots, 9$  are identified. Additionally, using (12)–(13) we obtain that the following counterfactual outcomes are identified

Always-takers	$E(Y(t_1) \mathbf{S} = \mathbf{s}_1)$	$E(Y(t_2) \mathbf{S} = \mathbf{s}_4)$	$E(Y(t_3) \mathbf{S} = \mathbf{s}_9)$
Switchers	$E(Y(t_1) \mathbf{S} = \mathbf{s}_3)$	$E(Y(t_2) \mathbf{S} = \mathbf{s}_2)$	$E(Y(t_3) \mathbf{S} = \mathbf{s}_5)$
	$E(Y(t_1) \mathbf{S} = \mathbf{s}_7)$	$E(Y(t_2) \mathbf{S} = \mathbf{s}_8)$	$E(Y(t_3) \mathbf{S} = \mathbf{s}_6)$
Partially Identified	$E(Y(t_1) \mathbf{S} \in \{\mathbf{s}_2, \mathbf{s}_6\})$	$E(Y(t_2) \mathbf{S} \in \{\mathbf{s}_3, \mathbf{s}_5\})$	$E(Y(t_3) \mathbf{S} \in \{\mathbf{s}_7, \mathbf{s}_8\})$

<sup>4</sup>If  $t_3$  is ranked highest a movement away from  $t_3$  represents moving towards a lower treatment while a towards  $t_3$  represents moving towards a higher treatment. Vice versa, if  $t_3$  is ranked lowest a movement towards  $t_3$  represents moving towards a lower treatment while a movement away from  $t_3$  represents moving towards a higher treatment.