Journal of Quantitative Economics (2022) 20 (Suppl 1):S7–S30
https://doi.org/10.1007/s40953-022-00307-w

ORIGINAL ARTICLE

Causal Inference of Social Experiments Using Orthogonal Designs

James J. Heckman[1] · **Rodrigo Pinto**[2]

Accepted: 27 April 2022 / Published online: 12 September 2022

Abstract
Orthogonal arrays are a powerful class of experimental designs that has been widely used to determine efficient arrangements of treatment factors in randomized controlled trials. Despite its popularity, the method is seldom used in social sciences. Social experiments must cope with randomization compromises such as noncompliance that often prevent the use of elaborate designs. We present a novel application of orthogonal designs that addresses the particular challenges arising in social experiments. We characterize the identification of counterfactual variables as a finite mixture problem in which choice incentives, rather than treatment factors, are randomly assigned. We show that the causal inference generated by an orthogonal array of incentives greatly outperforms a traditional design.

Keywords Strata · Discrete mixtures · Causal models · Experiments

Introduction

This paper investigates the problem of making causal inferences in social experiments under noncompliance. We develop two themes motivated by C.R. Rao's fundamental contributions to the characterization of distributions and the study of experiments. We use instrumental variables to characterize the identification of causal parameters as the solution to a mixing distribution problem. We then

James J. Heckman and Rodrigo Pinto contributed equally to this work.

✉ Rodrigo Pinto
rodrig@econ.ucla.edu

James J. Heckman
jjh@uchicago.edu

[1] Department of Economics and the Center for the Economics of Human Development, University of Chicago, 1126 East 59th Street, Chicago 60637, IL, USA

[2] Department of Economics, University of California, Los Angeles, 8385 Bunche Hall, Los Angeles 90095, CA, USA

explore orthogonal array designs to correct for the selection bias generated by noncompliance.

Statisticians widely use Rao's research on orthogonal arrays to design efficient arrangements of treatment factors in randomized controlled trials (RCTs). See, e.g., Stinson (2004). Despite its popularity, Rao's research has not been broadly applied to evaluate treatment effects in social sciences. Social experiments are commonly plagued by randomization compromises, such as noncompliance, that often prevent the use of elaborate designs. This paper uses recently developed econometric tools to repurpose Rao's original ideas into a novel framework where orthogonal arrays of *incentives* play a central role in solving compliance problems in social experiments.

In his M.A. thesis at Calcutta University, C. R. Rao (1943) introduced a powerful class of experimental designs called orthogonal arrays. This design employs combinatorial arrangements of factors (or treatments) for each randomization arm. Rao developed the theory of orthogonal arrays in a series of seminal papers (C. R. Rao 1946a, b, 1947, 1949).

The following matrix is an example of an orthogonal array:

$$\boldsymbol{A} = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix} \tag{1}$$

Matrix $\boldsymbol{A}$ is a 2-1evel orthogonal array because it only uses two elements, 0 and 1. Any two columns of the matrix display all the possible combinations of zeros and ones, that is, (0, 0), (0, 1), (1, 0), and (1, 1). The matrix has four "runs" (rows) corresponding to treatment conditions and three "factors" (columns) corresponding to treatments. The matrix is classified as *OA* (4, 3, 2, 2), where the first number 2 is the *level* and the second number 2 is the *strength*, which is the number of columns where we are guaranteed to see all the possible combinations of zeros and ones.

Orthogonal arrays such as *OA* (4, 3, 2, 2) are widely used to design experiments that determine the optimum mix of factors (or treatments) that maximize production yield. In these experiments, the researcher can choose the combination of inputs in each randomization arm.

A fundamental difference between RCTs in the natural and social sciences is that social scientists often cannot force compliance with intended treatments. In a natural experiment, the experimenter can determine the treatment of each randomization unit. In a social experiment, the randomization units consist of economic agents. The experimenter can attempt to persuade agents but can seldom impose an intended treatment status on them. The final treatment status depends on the agent's decision to comply or not comply with the initial treatment assignment.

Noncompliance violates the principle of randomization that secures the identification of causal effects in perfectly implemented RCTs. Agents that choose to deviate from their assigned treatment may differ from those who do not. The compliance decision introduces the danger of an unobserved confounding

variable that may cause both the treatment choice and the outcomes of interest. Noncompliance prevents the use of sophisticated designs, making it especially difficult to reap the benefits of Rao's orthogonal array design.

We present a novel approach to Rao's orthogonal array design to aid the nonparametric identification of causal effects in RCTs with noncompliance. We draw on research by Heckman and Pinto (2018) and Pinto, R. (2021a)[1] and use a choice-theoretic instrumental variable (IV) model. The identification of causal parameters hinges on methods that control for unobserved characteristics of agents. We use discrete instruments to generate a finite partition of unobserved variables. This partition enables us to characterize the identification of causal parameters as a problem of identifying a mixture of unobserved distributions. The partition induced by the instruments enables us to determine the necessary and sufficient conditions for identifying counterfactual outcomes. We use this framework to investigate how the orthogonal design of choice incentives outperforms the traditional approach to social experiments.

Section Causal Model with Choice and Compliance presents a choice-theoretic causal model using instrumental variables. Section Using IV to Control for Unobserved Variables explains how to nonparametrically control for an agent's unobservable characteristics using discrete instruments. Section Identification as a Mixture Problem describes the identification of causal effects as a problem of identifying a finite mixture of unobserved distributions. Section Using Rao's Orthogonal Design to Address Identification Problems Arising from Noncompliance in Social Experiments explains how to use Rao's orthogonal design to identify and estimate causal parameters. Section Conclusion concludes.

Causal Model with Choice and Compliance

In social experiments, the treatment status is typically determined by agents' decisions to comply with the treatment choice. This generates the problem of selection bias, which makes it difficult to identify causal effects. Economists have long used instrumental variables to solve the problem of selection bias and to identify causal effects in choice models. This paper examines the case of multivalued-choice models with categorical instrumental variables and heterogeneous agents.

Decision-Theoretic Foundation

The economic literature offers several theoretical foundations to model an economic agent ω's treatment choice t among the available treatments in a choice set $\mathcal{T}$.

The classical microeconomic theory assumes a rational agent that maximizes the utility among available choices. Agents, however, do not need to be rational to generate predictable choice behavior (Thaler 2016). As noted by Becker (1962), the key

[1] Pinto, R. (2021a). Beyond intention to treat: Using the incentives in moving to opportunity to identify neighborhood effects (unpublished manuscript). Department of Economics, University of California, Los Angeles. https://www.rodrigopinto.net/_files/ugd/95d94d_90f491ec1afa45cf8ef1e9a77346c9a8.pdf.

features of choice theory are a notion of preferences based on the agent's information set and some choice constraints, such as a budget set, that shape the agent's behavior—however rational or not.

We do not assume the full rationality of agents, but we allow for purposive actions under different information and constraint sets. We adopt a flexible choice equation consistent with a broad array of decision mechanisms. We denote the preferences of an agent ω over the choice set $\mathcal{T}$ by an unobserved random vector $\boldsymbol{V}_\omega$ of arbitrary but finite dimension. Choice constraints are indexed by the elements z in a finite set $\mathcal{Z}$. We keep the information sets of agents implicitly so that the treatment choice of agent ω given a restriction $z \in \mathcal{Z}$ is expressed as $T_\omega(z) = f_T(z, \boldsymbol{V}_\omega)$.

We map the choice behavior onto a standard IV model where treatment values $t \in \mathcal{T}$ and restriction indexes $z \in \mathcal{Z}$ become potential values in the support of the random variables T and Z, respectively. We use $\boldsymbol{X}$ for the random vector of baseline variables that occur prior to treatment choice. All variables are defined on the probability space $(\Omega, \mathcal{F}, P)$, and $Z_\omega, T_\omega, \boldsymbol{V}_\omega, \boldsymbol{X}_\omega$ denote the realized values of random variables $Z, T, \boldsymbol{V}, \boldsymbol{X}$ for an agent $\omega \in \Omega$.

The Instrumental Variable Model

The IV model has been a standard analytical framework in economics since Reiersöl (1945). In the economic context, the IV model consists of four observed variables: (1) an instrument Z taking N_Z discrete values in the support $\text{supp}\,(Z) = \{z_1, \ldots, z_{N_Z}\}$; (2) a treatment choice T taking N_T discrete values in $\text{supp}\,(T) = \{t_1, \ldots, t_{N_T}\}$; (3) a real-valued outcome[2] Y in $\mathbb{R}$; and (4) a pre-treatment random vector $\boldsymbol{X}$ of finite dimension taking values in $\mathbb{R}^{|X|}$. Notationally, we use $D_t = \mathbf{1}[T = t], t \in \text{supp}\,(T)$, and $D_z = \mathbf{1}[Z = z], z \in \text{supp}\,(Z)$, as indicators of treatment and instrument values, respectively.[3]

Observed variables are related according to two policy-invariant equations that determine causal relationships among the variables:[4]

$$\text{Choice Equation:} \qquad T = f_T(Z, \boldsymbol{V}, \boldsymbol{X}), \tag{2}$$

$$\text{Outcome Equation:} \qquad Y = f_Y(T, \boldsymbol{V}, \boldsymbol{X}, \epsilon_Y), \tag{3}$$

where ϵ_Y is an unobserved error term[5] in $\mathbb{R}$. As mentioned, the choice Eq. (2) is general and might be motivated by several choice mechanisms, including utility maximization (see, e.g., McFadden 1981). The unobserved random vector $\boldsymbol{V}$ subsumes

[2] Our analysis holds if outcome Y represents a vector-valued variable denoting multiple outcomes.

[3] The indicator function $\mathbf{1}[A]$ equals one if event A occurs and zero otherwise.

[4] By policy-invariant, we mean functions whose maps remain invariant under manipulation of the arguments. This is the notation of autonomy developed by Frisch (1938) and Haavelmo (1944). For a recent discussion of these conditions, see Heckman and Pinto (2015) and Pinto and Heckman (2021).

[5] Such error terms are often called "shocks" in structural equation models. f_T is a deterministic function that can be interpreted as a random function if we introduced shock ϵ_T of arbitrary dimension as one of its arguments.

not only the agent's preferences but all the unobserved (by the analyst) variables that affect both the choice T and outcome Y. Vector $\boldsymbol{V}$ is a *confounder*, and it is the source of selection bias. Choice probability $P(T = t \mid Z = z, \boldsymbol{X})$ is the *propensity score* of choosing t given z and $\boldsymbol{X}$.

The two main assumptions of the IV model are:

$$\text{Independence:} \quad Z \perp\!\!\!\perp (\boldsymbol{V}, \epsilon_Y) \mid \boldsymbol{X} \text{ are statistically independent,} \tag{4}$$

$$\text{IV Relevance:} \quad P(T = t \mid Z = z, \boldsymbol{X}) \text{ is a positive and non-degenerate function of z, for all } (t, z) \in \text{supp}\,(T) \times \text{supp}\,(Z). \tag{5}$$

Independence condition (4) states that the instrument Z is statistically independent of the confounder $\boldsymbol{V}$ and error term ϵ conditioned on baseline variables $\boldsymbol{X}$. Given that $\boldsymbol{V}$ is arbitrary, we can, without loss of generality, assume that $\boldsymbol{V}$ and ϵ are statistically independent; that is, $\boldsymbol{V} \perp\!\!\!\perp \epsilon \mid \boldsymbol{X}$. The independence condition implies that the instrument affects the outcome only through its impact on the treatment T.

IV relevance (5) guarantees that there exists agents who will choose t for any instrumental value z. The condition rules out the possibility that equivalent instrumental values have an identical impact on the treatment. We also assume as a regularity condition that the outcome expectation exists $E(Y^2) < \infty$. To simplify notation, we henceforth suppress the background variables $\boldsymbol{X}$. Our analysis can be interpreted as conditioned on such variables.

Counterfactuals

Counterfactual choice is defined by fixing Z in the choice Eq. (2) to a value $z \in \text{supp}\,(Z)$; that is, $T(z) = f_T(z, \boldsymbol{V})$. The counterfactual outcome is defined by fixing T in (3) to a value $t \in \text{supp}\,(T)$; that is, $Y(t) = f_Y(t, \boldsymbol{V}, \epsilon_Y)$.[6] The observed choice T and outcome Y can be described as switching regressions (Quandt 1958, 1972) by the following equations:

$$T = \sum_{z \in \text{supp}\,(Z)} T(z) \cdot D_z \equiv T(Z), \tag{6}$$

$$Y = \sum_{t \in \text{supp}\,(T)} Y(t) \cdot D_t \equiv Y(T). \tag{7}$$

Equation (6) describes choice T as the counterfactual choice $T(z)$ multiplied by the indicator D_z that takes value one if $Z = z$ and zero otherwise. Equation (7) describes the outcome Y in terms of the counterfactual outcomes $Y(t)$ multiplied by the choice indicator D_t.

[6] *Fixing* is a causal operation that captures the notion of external (*ceteris paribus*) manipulation. It is a central concept in the study of causality and dates back to Haavelmo (1943). See Heckman and Pinto (2015) for a recent discussion of fixing and causality.

The independence condition (4) generates two useful relations regarding counterfactuals:

$$\text{Exogeneity:} \quad Z \perp\!\!\!\perp (T(z), Y(t)) \text{ for all } (z,t) \in \text{supp}\,(Z) \times \text{supp}\,(T), \tag{8}$$

$$\text{Matching:} \quad Y(t) \perp\!\!\!\perp T \mid \boldsymbol{V} \text{ for all } t \in \text{supp}\,(T). \tag{9}$$

The exogeneity condition (8) is commonly used to describe IV models. It states that the instrument Z is independent of the counterfactuals. The matching property (9) states that controlling for the confounder $\boldsymbol{V}$ renders the outcome counterfactuals $Y(t)$ statistically independent of the treatment choice T.

Causal Inference

Causal analysis seeks to make inferences about counterfactual outcomes $Y(t)$. The causal effect of switching the treatment from t to t' for agent ω is given by $Y_\omega(t') - Y_\omega(t)$. A fundamental problem in causal inference is that, in any cross-section, we only observe a single outcome for each agent ω. Causal inference copes with this problem by focusing on the evaluation of average causal effects, specifically, the causal effect over a sub-population $\Omega' \subseteq \Omega$ of the agents:

$$E\big(Y(t') - Y(t) \mid \omega \in \Omega'\big) = \frac{\int_{\omega \in \Omega'} \big[Y_\omega(t') - Y_\omega(t)\big] dP}{P(\omega \in \Omega')}. \tag{10}$$

If $\Omega' = \Omega$ in (10), we obtain the average treatment effect of t' versus t on the outcome ATE $= E(Y(t') - Y(t))$.

Controlling for Unobservables

The identification of causal effects hinges on our ability to control for the confounder $\boldsymbol{V}$. By conditioning on $\boldsymbol{V}$, we are able to relate counterfactual outcome $E(Y(t) \mid \boldsymbol{V})$ and conditional outcome $E(Y \mid T = t, \boldsymbol{V})$:

$$\begin{aligned} E\big(Y(t) \mid \boldsymbol{V}\big) &= E\big(Y(t) \mid T = t, \boldsymbol{V}\big) \\ &= E\Bigg(\sum_{t \in \text{supp}\,(T)} Y(t) \cdot D_t \mid D_t = 1, \boldsymbol{V}\Bigg) = E\big(Y \mid T = t, \boldsymbol{V}\big), \end{aligned} \tag{11}$$

where the first equality is due to matching property (9) and the second equality is due to (7). If $\boldsymbol{V}$ were observed, we would be able to identify the counterfactual expectation $E(Y(t) \mid T = t, \boldsymbol{V})$ by the conditional expectation $E(Y \mid T = t, \boldsymbol{V})$. In addition, if $\boldsymbol{V}$ were observed, we would be able to identify its probability distribution. The counterfactual mean $E(Y(t))$ could be evaluated by integrating the conditional expectation $E(Y \mid T = t, \boldsymbol{V})$ over the unconditional distribution of $\boldsymbol{V}$:

$$
\begin{aligned}
E(Y(t)) &= \int_{v} E(Y(t) \mid \boldsymbol{V} = \boldsymbol{v}) dF_{\boldsymbol{V}}(\boldsymbol{v}) \\
&= \int_{v} E(Y \mid T = t, \boldsymbol{V} = \boldsymbol{v}) dF_{\boldsymbol{V}}(\boldsymbol{v}),
\end{aligned} \tag{12}
$$

where the second equality is due to (9), and $dF_{\boldsymbol{V}}(\boldsymbol{v})$ denotes the probability density of the confounder $\boldsymbol{V}$ at point $\boldsymbol{v}$.

The Identification Problem

Unfortunately, when $\boldsymbol{V}$ is not observed, the conditional expectation of the outcome $E(Y \mid T = t)$ does not identify the counterfactual mean $E(Y(t))$:

$$
\begin{aligned}
E\big(Y \mid T = t\big) &= \int_{v} E\big(Y \mid T = t, \boldsymbol{V} = \boldsymbol{v}\big) dF_{\boldsymbol{V}|T=t}(\boldsymbol{v}) \\
&= \int_{v} E\big(Y(t) \mid \boldsymbol{V} = \boldsymbol{v}\big) dF_{\boldsymbol{V}|T=t}(\boldsymbol{v}).
\end{aligned} \tag{13}
$$

Equation (13) clarifies how the outcome expectation $E(Y \mid T = t)$ differs from the counterfactual mean $E(Y(t))$. Outcome expectation $E(Y \mid T = t)$ is the weighted average of the counterfactual outcome $E(Y(t) \mid \boldsymbol{V} = \boldsymbol{v})$ over the *conditional* probability of $\boldsymbol{V}$ given $T = t$. On the other hand, the counterfactual mean $E(Y(t))$ is the weighted average of the counterfactual outcome $E(Y(t) \mid \boldsymbol{V} = \boldsymbol{v})$ over the *unconditional* probability distribution of $\boldsymbol{V}$. This mismatch prevents the identification of causal effects and can promote misleading conclusions. For instance, the difference-in-means estimator for the binary outcome $T \in \{0, 1\}$ evaluates the following parameter:

$$
\begin{aligned}
&E\big(Y \mid T = 1\big) - E\big(Y \mid T = 0\big) \\
&= \int_{v} E\big(Y(1) \mid \boldsymbol{V} = \boldsymbol{v}\big) dF_{\boldsymbol{V}|T=1}(\boldsymbol{v}) - \int_{v} E\big(Y(0) \mid \boldsymbol{V} = \boldsymbol{v}\big) dF_{\boldsymbol{V}|T=0}(\boldsymbol{v}).
\end{aligned} \tag{14}
$$

An identification problem arises because agent self-selection induces a correlation between choice T and the unobserved variables in $\boldsymbol{V}$. Large values of the difference in means in (14) could arise from the difference between the distribution of $\boldsymbol{V}$ conditioned on the treatment choices instead of the impact of the treatment on the outcome.

RCTs are supposed to solve the problem of selection bias by randomly assigning the treatments. The randomization secures statistical independence between the treatment T and the unobserved characteristics of the agents, namely, the confounder $\boldsymbol{V}$. The independence relationship $\boldsymbol{V} \perp\!\!\!\perp T$ implies that the distribution of $\boldsymbol{V}$ conditional on T is equal to the unconditional distribution of $\boldsymbol{V}$, and therefore, the outcome difference-in-means identified the average treatment effect.

Noncompliance in RCTs potentially compromises the independence relationship between agents' unobserved variables $\boldsymbol{V}$ and their final treatment assignment T. Effectively, noncompliance transforms the intended RCT experiment into an IV model where the randomization arms determine the instrumental variable.

Using IV to Control for Unobserved Variables

Identification strategies in IV models use instruments Z to control for the unobserved confounder $\boldsymbol{V}$ (Heckman and Pinto 2015). One approach assumes parametric models that impose functional restrictions on the choice Eq. (2) and the outcome Eq. (3). An example of this approach is Two-Stage Least Squares (Theil 1958, 1971).

Heckman and Pinto (2018) propose a nonparametric approach that explores the choice behavior induced by the instrument Z. They use counterfactual choices to determine a partition of the support of supp $(\boldsymbol{V})$ that renders T statistically independent of the counterfactual outcomes $Y(t)$. This independence property enables them to characterize the observed data as a mixture of unobserved counterfactuals over the partition set of supp $(\boldsymbol{V})$. We use this characterization to determine the necessary and sufficient conditions to point-identify counterfactual outcomes. Additional notation is necessary to introduce their results.

The Response Vector

We control for the unobservables $\boldsymbol{V}$ using a partition of it generated by the choice variation induced by the instrument. A central concept in our analysis is the *response vector*. This is the N_Z-dimensional random vector of counterfactual choices $T(z)$ across all the instrumental values $z_1, ..., z_{N_Z}$:

$$\boldsymbol{S} = \left[T(z_1), \ \dots \ , T\left(z_{N_Z}\right)\right]'. \tag{15}$$

The support of the response vector is given by supp $(S) = \{\boldsymbol{s}_1, \dots, \boldsymbol{s}_{N_S}\}$, and each element $\boldsymbol{s} \in$ supp (S) is called a *response-type*. The response vector for an agent ω is given by $\boldsymbol{S}_\omega = [T_\omega(z_1), \ \dots \ , T_\omega(z_{N_Z})]'$. It lists the treatment choices that agent ω would take if it were to face each instrumental value.[7]

Response vector $\boldsymbol{S}$ has been used by several authors in distinct fields, starting with Robins and Greenland (1992) and Balke and Pearl (1993), who studied bounds for causal effects for the binary choice model. Angrist et al. (1996) use response-types to study the identification of a binary choice model.

Response vectors are called "principal strata" by Frangakis and Rubin (2002) and can be understood as the control functions of Heckman and Robb (1985) and Powell (1994). Our approach differs from these interpretations. We use the response vector $\boldsymbol{S}$ as a criterion to control for the unobserved confounding variable $\boldsymbol{V}$.

Equation (16) expresses the response vector $\boldsymbol{S}$ as a function of $\boldsymbol{V}$, while Eq. (17) expresses choice T as a function of the response vector $\boldsymbol{S}$ and the instrument Z. Figure 1 displays these causal relationships graphically as directed acyclic graphs (DAGs).

[7] The response-types can be viewed as "types" in the sense of Keane and Wolpin (1997).

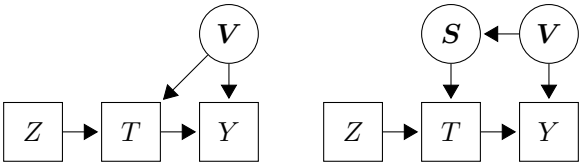

Fig. 1 IV Models with and without the Response Vector $\boldsymbol{S}$. These two diagrams depict equivalent IV models as DAGs. Arrows represent direct causal relations. Circles represent unobserved variables. Squares represent observed variables. The error term ϵ is kept implicit. The left-hand side diagram shows the standard IV model without the response vector $\boldsymbol{S}$, while the right-hand side diagram includes the response vector $\boldsymbol{S}$.

$$\boldsymbol{S} = \left[f_T(z_1, \boldsymbol{V}), \ldots, f_T(z_{N_Z}, \boldsymbol{V})\right]' = h(\boldsymbol{V}), \tag{16}$$

$$T = \left[D_{z_1}, \ldots, D_{z_{N_Z}}\right] \cdot \boldsymbol{S} = g(Z, \boldsymbol{S}). \tag{17}$$

Equation (18) lists three useful properties of the response vector $\boldsymbol{S}$:

$$\text{(i) } \boldsymbol{S} \perp\!\!\!\perp Z, \quad \text{(ii) } Y(t) \perp\!\!\!\perp T \mid \boldsymbol{S}, \quad \text{(iii) } Y \perp\!\!\!\perp T \mid (\boldsymbol{S}, Z). \tag{18}$$

Property (i) states that the response vector is independent of the IV. This independence relationship stems from $\boldsymbol{V} \perp\!\!\!\perp Z$ in (4) and from the fact that $\boldsymbol{S}$ is a function of $\boldsymbol{V}$. Property (ii) states a matching condition where $\boldsymbol{S}$ plays the role of a balancing score for $\boldsymbol{V}$.[8] The relationship stems from $(Y(t), \boldsymbol{V}) \perp\!\!\!\perp Z$ and from the fact that $\boldsymbol{S}$ is a function of $\boldsymbol{V}$, while T is a function of Z and $\boldsymbol{V}$.[9] Indeed, conditioned on $\boldsymbol{S}$, T depends only on Z, which is independent of $Y(t)$. The last property (iii) is due to the fact that T is deterministic given T and $\boldsymbol{S}$.

The properties of the response vector in (18) enable us to describe a coarse partition of $\text{supp}(\boldsymbol{V})$ that renders the treatment statistically independent of counterfactual outcomes. According to (16), each $\boldsymbol{v} \in \text{supp}(\boldsymbol{V})$ corresponds to one and only one response-type $\boldsymbol{s} \in \text{supp}(\boldsymbol{S})$ such that $h(\boldsymbol{v}) = \boldsymbol{s}$. Thus, for each response-type $\boldsymbol{s}_n \in \text{supp}(\boldsymbol{S})$, we can define a subset $\mathcal{V}_n \subset \text{supp}(\boldsymbol{V})$ as:

$$\mathcal{V}_n = \left\{\boldsymbol{v} \in \text{supp}(\boldsymbol{V}) \text{ such that } \left[f_T(z_1, \boldsymbol{v}), \ldots, f_T(z_{N_Z}, \boldsymbol{v})\right]' = \boldsymbol{s}_n\right\}. \tag{19}$$

Sets $\mathcal{V}_1, \ldots, \mathcal{V}_{N_S}$ constitute a disjoint partition of $\text{supp}(\boldsymbol{V})$ and their union spans the full set; that is,

[8] A balancing score for $\boldsymbol{V}$ constitutes a function of $\boldsymbol{V}$ that preserves the matching condition in $Y(t) \perp\!\!\!\perp T \mid \boldsymbol{V}$ (9). See Rosenbaum and Rubin (1983).

[9] Formally, $(Y(t), \boldsymbol{V}) \perp\!\!\!\perp Z \Rightarrow (Y(t), h(\boldsymbol{V})) \perp\!\!\!\perp Z \Rightarrow (Y(t), \boldsymbol{S}) \perp\!\!\!\perp Z \Rightarrow Y(t) \perp\!\!\!\perp Z \mid \boldsymbol{S} \Rightarrow Y(t) \perp\!\!\!\perp g(Z, \boldsymbol{S}) \mid \boldsymbol{S} \Rightarrow Y(t) \perp\!\!\!\perp T \mid \boldsymbol{S}$.

$$\text{supp}\,(S) = \bigcup_{n=1}^{N_S} \mathcal{V}_n \text{ such that } \mathcal{V}_n \bigcap \mathcal{V}_{n'} = \varnothing. \tag{20}$$

Note that the events $\boldsymbol{S} = \boldsymbol{s}_n$ and $\boldsymbol{V} \in \mathcal{V}_n$ are equivalent. The matching property (ii) in (18) states that $Y(t) \perp\!\!\!\perp T \mid (\boldsymbol{S} = \boldsymbol{s}_n)$, so

$$T \perp\!\!\!\perp Y(t) \mid (\boldsymbol{V} \in \mathcal{V}_n) \quad \text{for each } n \in \{1, \ldots, N_S\}. \tag{21}$$

Equations (19)–(21) imply that the treatment T can be understood as being randomly assigned when we condition on the subset of agents ω that share the same response-type $\boldsymbol{s}$. If response-types were observed, we could use (ii) in (18) to identify the expected value of counterfactual outcomes by taking the expected values of the observed outcome conditioned on the treatment choice and the response-types.

A significant challenge is that the response-types that determine the partition of the support of $\boldsymbol{V}$ are not observed. Nevertheless, the partition substantially simplifies the identification problem. It reframes the identification of counterfactuals as a problem of identifying a finite mixture of unobserved distributions.

Identification as a Mixture Problem

We gain a deeper understanding by reframing the identification problem as a particular case of the identification of unobserved mixture distributions (B. L. S. P. Rao 1992). The general mixture model is given by:

$$F(Y) = \int F_\theta(Y) dG(\theta), \tag{22}$$

where $F(Y)$ stands for the cumulative distribution function (cdf) of an observed outcome Y, $(F_\theta(Y))_{\theta \in \Theta}$ is a collection of cdf's indexed by a random variable $\theta \in \Theta$ that takes a value in the (possibly infinite) set Θ, and G denotes the cdf of θ. $F(Y)$ is a mixture distribution, the cdf's $(F_\theta(Y))_{\theta \in \Theta}$ are component distributions, G is the mixing distribution, and θ is the unobserved latent (or mixing) variable. B. L. S. P. Rao (1992) notes that if the mixing distribution G is finite, then a necessary and sufficient condition for its identification is that the family of cdf's $(F_\theta(Y))_{\theta \in \Theta}$ be linearly independent as functions on Y. We use the mixture model (22) as a starting point.

As mentioned, the identification of causal parameters hinges on controlling for unobserved variables $\boldsymbol{V}$. A natural candidate for the values of θ in (22) are the elements $\boldsymbol{v} \in \text{supp}\,(\boldsymbol{V})$. We replace the cdf's in (22) by the expectation of $\kappa(Y)$, where $\kappa : \text{supp}\,(Y) \to \mathbb{R}$ is an arbitrary real-valued function.

$$E(\kappa(Y)) = \int_{\boldsymbol{v}} E(\kappa(Y) \mid \boldsymbol{V} = \boldsymbol{v}) dF_{\boldsymbol{V}}(\boldsymbol{v}) \tag{23}$$

$$= \sum_{n=1}^{N_S} E(\kappa(Y) \mid \boldsymbol{V} \in \mathcal{V}_n) P(\boldsymbol{V} \in \mathcal{V}_n). \tag{24}$$

Equation (23) describes the expected outcome using the mixture model in (22), where θ stands for the elements $\boldsymbol{v} \in \text{supp}(\boldsymbol{V})$. Equation (24) uses the partition of $\text{supp}(\boldsymbol{V})$ in (19) to generate a discrete mixing distribution across the partition sets of the support of $\boldsymbol{V}$. Condition (21) in Section Using IV to Control for Unobserved Variables enables us to express the conditional expectation $E(\kappa(Y) \mid T = t)$ in terms of the conditional counterfactuals $E(\kappa(Y(t)) \mid \boldsymbol{V})$:

$$E\big(\kappa(Y) \mid T = t\big) = \sum_{n=1}^{N_S} E\Big(\kappa\big(Y(t)\big) \mid \boldsymbol{V} \in \mathcal{V}_n\Big) P\big(\boldsymbol{V} \in \mathcal{V}_n \mid T = t\big). \tag{25}$$

Equation (25) relates a single conditional outcome expectation with several outcome counterfactuals for each choice value $t \in \text{supp}(T)$. The equation does not assess sufficient information on observed data to secure the identification of the counterfactual outcomes. The instrumental variable Z generates additional variation of observed quantities (left-hand side of (25)) without increasing the number of unobserved counterfactuals (right-hand side of (25)):

$$\begin{aligned} &E\big(\kappa(Y) \mid T = t, Z = z\big) \\ &= \sum_{n=1}^{N_S} E\Big(\kappa\big(Y(t)\big) \mid Z = z, \boldsymbol{V} \in \mathcal{V}_n\Big) P\big(\boldsymbol{V} \in \mathcal{V}_n \mid T = t, Z = z\big) \end{aligned} \tag{26}$$

$$= \sum_{n=1}^{N_S} E(\kappa(Y(t)) \mid Z = z, \boldsymbol{S} = \boldsymbol{s}_n) P(\boldsymbol{S} = \boldsymbol{s}_n \mid T = t, Z = z) \tag{27}$$

$$= \sum_{n=1}^{N_S} E\Big(\kappa\big(Y(t)\big) \mid \boldsymbol{S} = \boldsymbol{s}_n\Big) \frac{P\big(T{=}t \mid Z{=}z, \boldsymbol{S}{=}\boldsymbol{s}_n\big) P\big(\boldsymbol{S}{=}\boldsymbol{s}_n \mid Z{=}z\big)}{P\big(T = t \mid Z = z\big)} \tag{28}$$

$$= \sum_{n=1}^{N_S} \mathbf{1}\big[T = t \mid Z = z, \boldsymbol{S} = \boldsymbol{s}_n\big] E\Big(\kappa\big(Y(t)\big) \mid \boldsymbol{S} = \boldsymbol{s}_n\Big) \frac{P(\boldsymbol{S} = \boldsymbol{s}_n)}{P\big(T = t \mid Z = z\big)}. \tag{29}$$

Equation (26) rewrites (25) conditioning on instrument Z. Equation (27) uses the fact that $Z \perp\!\!\!\perp \boldsymbol{S}$ and that $\boldsymbol{V} \in \mathcal{V}_n$ and $\boldsymbol{S} = \boldsymbol{s}_n$ are equivalent events. Equation (28) uses Bayes rule to rewrite the conditional expectation $P(\boldsymbol{S} = \boldsymbol{s}_n \mid T = t, Z = z)$. Equation (29) employs $Z \perp\!\!\!\perp \boldsymbol{S}$ again and invokes the fact that T is deterministic when conditioned on $\boldsymbol{S}$ and Z. The response vector $\boldsymbol{S}$ enables us to connect observed data with a mixture of counterfactual outcomes conditioned on response-types. This produces our main equation:

$$
\begin{aligned}
&E\big(\kappa(Y) \mid T = t, Z = z\big)P\big(T = t \mid Z = z\big) \\
&= \sum_{n=1}^{N_S} \mathbf{1}\big[T = t \mid Z = z, \boldsymbol{S} = \boldsymbol{s}_n\big]E\big(\kappa(Y(t)) \mid \boldsymbol{S} = \boldsymbol{s}_n\big)P(\boldsymbol{S} = \boldsymbol{s}_n).
\end{aligned} \tag{30}
$$

If $\kappa(Y) = Y$, (30) generates an equality relating the expected values of observed outcomes with expected counterfactual outcomes. Setting $\kappa(Y) = \mathbf{1}[Y \leq y]$ generates the cdf of the observed outcome with the unobserved cdf of counterfactual outcomes. Setting $\kappa(Y)$ to 1 in (30) generates the propensity score equality:

$$
P\big(T = t \mid Z = z\big) = \sum_{n=1}^{N_S} \mathbf{1}\big[T = t \mid \boldsymbol{S} = \boldsymbol{s}_n, Z = z\big]P(\boldsymbol{S} = \boldsymbol{s}_n). \tag{31}
$$

Replacing $\kappa(Y)$ by any variable X such that $X \perp\!\!\!\perp T \mid \boldsymbol{S}$ generates an equation that relates baseline variables with response-types:

$$
\begin{aligned}
&E\big(X \mid T = t, Z\big)P\big(T = t \mid Z\big) \\
&= \sum_{n=1}^{N_S} \mathbf{1}\big[T = t \mid \boldsymbol{S} = \boldsymbol{s}_n, Z\big]E\big(X \mid \boldsymbol{S} = \boldsymbol{s}_n\big)P(\boldsymbol{S} = \boldsymbol{s}_n).
\end{aligned} \tag{32}
$$

Identification Criteria

We now investigate the necessary and sufficient conditions for identifying counterfactual outcomes and response-type probabilities. To do so, we express our main Eq. (30) as a system of linear equations.

Observed parameters are stacked in vectors $\boldsymbol{P}_Z(t)$ and $\boldsymbol{Q}_Z(t)$ below:

$$
\boldsymbol{P}_Z(t) = \big[P(T = t \mid Z = z_1), \ldots, P(T = t \mid Z = z_{N_Z})\big]', \tag{33}
$$

$$
\boldsymbol{Q}_Z(t) = \Big[E\big(\kappa(Y) \mid T = t, Z = z_1\big), \ldots, E\big(\kappa(Y) \mid T = t, Z = z_{N_Z}\big)\Big]', \tag{34}
$$

where $\boldsymbol{P}_Z(t)$ is the vector of observed propensity scores, and $\boldsymbol{Q}_Z(t)$ is the vector of outcome expectations. The unobserved parameters are stacked in the vectors $\boldsymbol{P}_S$ and $\boldsymbol{Q}_S(t)$ below:

$$
\boldsymbol{P}_S = \big[P(\boldsymbol{S} = \boldsymbol{s}_1), \ldots, P(\boldsymbol{S} = \boldsymbol{s}_{N_S})\big]', \tag{35}
$$

$$
\boldsymbol{Q}_S(t) = \Big[E\Big(\kappa\big(Y(t)\big) \mid \boldsymbol{S} = \boldsymbol{s}_1\Big), \ldots, E\Big(\kappa\big(Y(t)\big) \mid \boldsymbol{S} = \boldsymbol{s}_{N_S}\Big)\Big]', \tag{36}
$$

where $\boldsymbol{P}_S$ is the vector of response-type probabilities, and $\boldsymbol{Q}_S(t)$ is the vector of counterfactual outcomes conditioned on response-types.

Response matrix $\boldsymbol{R}$ stacks the response-types in supp $(\boldsymbol{S})$ as columns:

$$\boldsymbol{R} = [\boldsymbol{s}_1, \ldots, \boldsymbol{s}_{N_S}]. \tag{37}$$

Matrix $\boldsymbol{R}$ has dimension $N_Z \times N_S$. The entry in the ith row and nth column of $\boldsymbol{R}$ is denoted by $\boldsymbol{R}[i,n] = (T \mid Z = z_i, \boldsymbol{S} = \boldsymbol{s}_n), i \in \{1, \ldots, N_Z\}, n \in \{1, \ldots, N_S\}$. We use $\boldsymbol{R}[i,\cdot]$ to denote the ith row of $\boldsymbol{R}$ and $\boldsymbol{R}[\cdot,n]$ to denote the nth column of $\boldsymbol{R}$. IV relevance condition (5) prevents identical rows in $\boldsymbol{R}$.

We use $\boldsymbol{B}_t = \mathbf{1}[\boldsymbol{R} = t]$ to denote a binary matrix of the same dimension of $\boldsymbol{R}$ that takes value 1 if the respective element in $\boldsymbol{R}$ is equal to t and zero otherwise. An entry of $\boldsymbol{B}_t$ is given by $\boldsymbol{B}_t[i,n] = \mathbf{1}[T = t \mid Z = z_i, \boldsymbol{S} = \boldsymbol{s}_n]$. Let $\boldsymbol{B}_T$ be a binary matrix of dimension $(N_Z \cdot N_T) \times N_S$ generated by stacking $\boldsymbol{B}_t$ as t ranges over $\text{supp}\,(T)$: $\boldsymbol{B}_T = [\boldsymbol{B}'_{t_1}, \ldots, \boldsymbol{B}'_{t_{N_T}}]'$ and let $\boldsymbol{P}_Z$ be the $(N_Z \cdot N_T) \times 1$ vector that stacks the propensity scores $P_Z(t)$ across the treatment values: $\boldsymbol{P}_Z = [\boldsymbol{P}(t_1)', \ldots, \boldsymbol{P}(t_{N_T})']'$. In this notation, Eqs. (30) and (31) can be written in matrix form by the following equations:

$$\boldsymbol{P}_Z = \boldsymbol{B}_T \boldsymbol{P}_S, \tag{38}$$

$$\boldsymbol{Q}_Z(t) \odot \boldsymbol{P}_Z = \boldsymbol{B}_t \boldsymbol{Q}_S(t) \odot \boldsymbol{P}_S, \tag{39}$$

where $\odot$ denotes the Hadamard (element-wise) multiplication.

The response matrix $\boldsymbol{R}$ and the binary matrices $\boldsymbol{B}_t$, $t \in \text{supp}\,(T)$, are deterministic, as T is known given Z and $\boldsymbol{S}$. If $\boldsymbol{B}_t$ and $\boldsymbol{B}_T$ were invertible, $\boldsymbol{Q}_S(t)$ and $\boldsymbol{P}_S$ would be identified. However, such inverses do not always exist. In their place, we can use generalized inverses. Let $\boldsymbol{B}_T^+$ and $\boldsymbol{B}_t^+$ be the Moore-Penrose pseudo-inverses[10] of $\boldsymbol{B}_T$ and $\boldsymbol{B}_t$, $t \in \text{supp}\,(T)$. Under this notation, we can state the following result:

Theorem T-1 *The general solution for the system of linear equations in* (38) *and* (39):

$$\boldsymbol{P}_S = \boldsymbol{B}_T^+ \boldsymbol{P}_Z + \boldsymbol{K}_T \boldsymbol{\lambda} \tag{40}$$

and

$$\boldsymbol{Q}_S(t) \odot \boldsymbol{P}_S = \boldsymbol{B}_t^+ \boldsymbol{Q}_Z(t) + \boldsymbol{K}_t \tilde{\boldsymbol{\lambda}} \tag{41}$$

such that

$$\boldsymbol{K}_T = \boldsymbol{I}_{N_S} - \boldsymbol{B}_T^+ \boldsymbol{B}_T \quad \text{and} \quad \boldsymbol{K}_t = \boldsymbol{I}_{N_S} - \boldsymbol{B}_t^+ \boldsymbol{B}_t, \quad t \in \text{supp}\,(T), \tag{42}$$

where $\boldsymbol{I}_{N_S}$ *denotes an identity matrix of dimension* N_S, *and* $\boldsymbol{\lambda}$ *and* $\tilde{\boldsymbol{\lambda}}$ *are arbitrary* N_S*-dimensional vectors* (*with the same dimension as* $\boldsymbol{P}_S$).

Proof See Appendix A.1. □

[10] The Moore-Penrose inverse of a matrix $\boldsymbol{A}$ is denoted by $\boldsymbol{A}^+$ and is defined by the following four properties: (1) $\boldsymbol{A}\boldsymbol{A}^+\boldsymbol{A} = \boldsymbol{A}$; (2) $\boldsymbol{A}^+\boldsymbol{A}\boldsymbol{A}^+ = \boldsymbol{A}^+$; (3) $\boldsymbol{A}^+\boldsymbol{A}$ is symmetric; (4) $\boldsymbol{A}\boldsymbol{A}^+$ is symmetric. The Moore-Penrose matrix $\boldsymbol{A}^+$ of a real matrix $\boldsymbol{A}$ is unique and always exists (Magnus and Neudecker 1999).

Matrices $\boldsymbol{K}_T$ and $\boldsymbol{K}_t$ are orthogonal projection matrices that depend only on matrices $\boldsymbol{B}_T$ and $\boldsymbol{B}_t, t \in \text{supp}\,(T)$. Theorem T-1 is useful to provide the general conditions for identification of response probabilities and counterfactual means:

Corollary C-1 *In the IV model* (4)–(5), *if there exists a real-valued* N_S*-dimensional vector* λ *such that* $\lambda' \boldsymbol{K}_T = \boldsymbol{0}$, *then* $\lambda' \boldsymbol{P}_S$ *is identified. In addition, if there exists a real-valued* N_S*-dimensional vector* $\tilde{\lambda}$ *such that* $\tilde{\lambda}' \boldsymbol{K}_t = \boldsymbol{0}$, *then* $\tilde{\lambda}' \boldsymbol{Q}_S(t)$ *is identified.*

Proof See Heckman and Pinto (2018) or Appendix A.2. □

Corollary C-1 shows that the nonparametric identification of counterfactuals depends only on properties of the response matrix $\boldsymbol{R}$. If $\boldsymbol{B}_T$ had full column-rank, then $\boldsymbol{B}_T^+ \boldsymbol{B}_T = \boldsymbol{I}_{N_S}$ and $\boldsymbol{K}_T = \boldsymbol{0}$. In this case, each response-type probability is identified. Indeed, $\lambda' \boldsymbol{P}_S$ is identified for any real vector λ of dimension N_S including those that indicate each of the response-type probabilities.[11]

Binary matrix $\boldsymbol{B}_T$ contains each $\boldsymbol{B}_t, t \in \text{supp}\,(T)$. Thus, the conditions for identifying response-type probabilities are weaker than those for identifying counterfactual outcomes. In particular, a full-rank $\boldsymbol{B}_T$ does not imply that matrices $\boldsymbol{B}_t$, $t \in \text{supp}\,(T)$, are full-rank. Therefore, the identification of the response-type probabilities does not automatically identify corresponding mean counterfactual outcomes. Corollary C-2 formalizes this discussion.

Corollary C-2 *The following relationships hold for the IV model* (4)–(5):

$$\text{Vector } \boldsymbol{P}_S \text{ is point-identified} \Leftrightarrow \text{rank}\,(\boldsymbol{B}_T) = N_S, \tag{43}$$

$$\text{Vector } \boldsymbol{Q}_S(t) \text{ is point-identified} \Leftrightarrow \text{rank}\,(\boldsymbol{B}_t) = N_S. \tag{44}$$

Also, if (44) *holds, then* $E(\kappa(Y(t)))$ *is identified by* $\iota' \boldsymbol{B}_t^+ \boldsymbol{Q}_Z(t)$, *where* ι *is an* N_S*-dimensional vector of ones.*

Proof See Heckman and Pinto (2018) or Appendix A.3. □

Versions of Corollary C-2 are found in the literature on the identifiability of finite mixtures (see, e.g., Yakowitz and Spragins 1968 and B. L. S. P. Rao 1992). Given binary matrices $\boldsymbol{B}_T$ and $\boldsymbol{B}_t$, $t \in \{1, \ldots, N_T\}$, the problem of identifying $\boldsymbol{P}_S$ and $\boldsymbol{Q}_S(t)$ is equivalent to the problem of identifying finite mixtures of distributions where $\boldsymbol{B}_T$ and $\boldsymbol{B}_t$ play the roles of kernels of mixtures. Mixture components are the corresponding counterfactual outcomes conditional on the response-types, and mixture probabilities are the response-type probabilities.

[11] See Section A.4 of the Appendix for bounds on the response-type probabilities and counterfactual outcomes.

Understanding the Identification Challenge

Identification criteria (43) and (44) show that the identification of causal parameters depends solely on the properties of the response matrix $\boldsymbol{R}$. In particular, the identification of the counterfactual outcomes in $\boldsymbol{Q}_S(t)$ depends on the column-rank of the binary matrix $\boldsymbol{B}_t$. If the column-rank of $\boldsymbol{B}_t$ is N_S (full column-rank), then $\boldsymbol{B}_t^{+}\boldsymbol{B}_t = \boldsymbol{I}_{N_S}$ and $\boldsymbol{K}_t = \boldsymbol{0}$. In this case $\boldsymbol{\xi}'\boldsymbol{Q}_S(t)$ and $\boldsymbol{\xi}'\boldsymbol{P}_S$ are identified for any real vector $\boldsymbol{\xi}$ of dimension N_S, including all unit vectors with a value of 1 in the nth entry and 0 elsewhere. In summary, all counterfactual outcomes in $\boldsymbol{Q}_S(t)$ would be identified.

Identification criteria (43) and (44) pose a major identification problem. The column rank of any binary matrix $\boldsymbol{B}_t$ is less than or equal to its row-dimension N_Z. On the other hand, the dimension of $\boldsymbol{Q}_S(t)$ is the number of response-types N_S that usually far exceeds the number of IV-values N_Z. For instance, under no restrictions, the total number of potential response-types is $N_T^{N_Z}$. Thus, a requirement for generating any identification result on counterfactual outcomes is to reduce the number of response-types that the choice model admits.

A common approach to decreasing the number of response-types is to impose functional restrictions on the choice equation. Heckman and Pinto (2018) and Pinto (2021a)[12] adopt a different approach that relies on economic choice theory. They combine choice incentives with revealed preference analysis to generate choice restrictions that systematically eliminate potential response-types.

Using Rao's Orthogonal Design to Address Identification Problems Arising from Noncompliance in Social Experiments

We propose a novel application of Rao's orthogonal design (C. R. Rao 1946a, b, 1947, 1949). Rao's methodology is traditionally applied to investigate the effects of combinations of treatment factors. The method determines randomization groups exposed to an orthogonal arrangement of treatment factors.

Similar to Rao's work, ours uses an RCT setting. Our method differs from Rao's original methodology in two ways: (1) we consider the possibility of noncompliance; and (2) the orthogonal array design is not used to combine treatment factors but to determine *choice incentives* across a finite number of treatment alternatives.

We use revealed preference analysis to translate choice incentives into choice restrictions that eliminate response-types. This elimination process generates the response matrix $\boldsymbol{R}$, which contains all the necessary information to examine the nonparametric identification of causal parameters.

[12] Pinto, R. (2021a). Beyond intention to treat: Using the incentives in moving to opportunity to identify neighborhood effects [Unpublished manuscript]. Department of Economics, University of California, Los Angeles. https://www.rodrigopinto.net/_files/ugd/95d94d_90f491ec1afa45cf8ef1e9a77346c9a8.pdf.

Examining Choice Incentives Determined by an Orthogonal Array

Noncompliance in social experiments effectively transforms the original RCT into an IV model where each instrumental value represents a randomization arm. It implicitly adds a choice probability, which we explicitly model in the IV model. The experimenter cannot impose a treatment status upon participants but rather incentivizes them toward a treatment choice. In this setup, orthogonal arrays play the role of the *incentive matrix* of Pinto (2021a).[13] Each factor stands for a treatment choice and each run stands for a randomization arm that incentivizes one or several treatment alternatives.

We illustrate the method using the orthogonal array OA (4, 3, 2, 2) discussed in Section Introduction. This design can be understood as an RCT with four randomization arms $Z \in \{z_1, z_2, z_3, z_4\}$ and three treatment statuses $T \in \{t_1, t_2, t_3\}$, where z_1 denotes the control group that offers no incentive toward any choice, z_2 incentivizes participants toward choices t_1 and t_2, z_3 incentivizes them toward choices t_1 and t_3, and z_4 incentivizes them toward choices t_2 and t_3. This incentive pattern is described by an ordinal *incentive matrix*:

$$\text{Incentive Matrix } L = \begin{array}{c} \begin{matrix} t_1 & t_2 & t_3 \end{matrix} \\ \begin{bmatrix} 0 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix} \end{array} \begin{matrix} z_1 \\ z_2 \\ z_3 \\ z_4 \end{matrix} \tag{45}$$

Each column displays which choices are incentivized across all values of the instruments. The incentive matrix $\boldsymbol{L}$ in (45) is an orthogonal array of type OA (4, 3, 2, 2). The factors refer to treatment choices; the runs, to instrumental values.

Choice Restrictions

Classical revealed preference analysis can be used to translate choice incentives into choice restrictions. Pinto (2021a)[14] shows that the Weak Axiom of Revealed Preferences (WARP) and Normal Choice generate the *choice rule* described below:

$$\text{If } T_\omega(z) = t \text{ and } \boldsymbol{L}[z', t'] - \boldsymbol{L}[z, t'] \leq \boldsymbol{L}[z', t] - \boldsymbol{L}[z, t], \text{ then } T_\omega(z') \neq t'. \tag{46}$$

Choice rule (46) is intuitive. It states that if an agent ω chooses choice t under z, and the change from z to z' induces greater incentives toward t than toward t', then the same agent ω does not choose t' under z'.

[13] Pinto, R. (2021a). Beyond intention to treat: Using the incentives in moving to opportunity to identify neighborhood effects [Unpublished manuscript]. Department of Economics, University of California, Los Angeles. https://www.rodrigopinto.net/_files/ugd/95d94d_90f491ec1afa45cf8ef1e9a77346c9a8.pdf.

[14] Pinto, R. (2021a). Beyond intention to treat: Using the incentives in moving to opportunity to identify neighborhood effects [Unpublished manuscript]. Department of Economics, University of California, Los Angeles. https://www.rodrigopinto.net/_files/ugd/95d94d_90f491ec1afa45cf8ef1e9a77346c9a8.pdf.

Table 1 Applying Choice Rule (46) to $T_\omega(z_1) = t_1$

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

This table presents all the choice restrictions generated by applying the choice rule (46) 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 (45)

Table 2 Choice Restrictions Generated by Incentive Matrix (45)

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

This table presents all the choice restrictions generated by applying the choice rule (46) 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 (45)

Choice rules like (46) restrict $\boldsymbol{R}$. They enable analysts to translate any incentive matrix into a set of choice restrictions and generate a response matrix. A simple algorithm efficiently implements the task of moving from an incentive matrix to a response matrix. We now clarify this process.

Consider an agent ω that chooses t_1 if it were assigned to z_1; that is, $T_\omega(z_1) = t_1$. We seek to examine whether the agent would choose t_2 if it were assigned to z_2, z_3, or z_4.

The first row of Table 1 compares the incentive gains for choosing t_1 and t_2 if the instrument were to change from z_1 to z_2. The incentives to choose either t_1 or t_2 increase, which satisfies the incentive requirement of choice rule (46). Therefore, we can state that an agent that chooses t_1 under z_1 does not choose t_2 under z_2. This choice restriction is summarized as $T_\omega(z_1) = t_1 \Rightarrow T_\omega(z_2) \neq t_2$.

The second row compares the incentives to choose t_3 for the same instrumental change (z_1 to z_2). The incentive to choose t_1 increases, while the incentive to chose t_3 does not. Choice rule (46) applies and the agent does not switch to t_3; that is, $T_\omega(z_1) = t_1 \Rightarrow T_\omega(z_2) \neq t_3$.

The third and fourth rows of Table 1 compare the incentives for choosing t_1 versus t_2 (third row) and t_1 versus t_3 (fourth row) when the instrument changes from z_1 to z_3. The incentive to choose t_1 increases, while the incentives to choose either t_2 or t_3 do not. Choice rule (46) holds and the agent does not choose t_2 or t_3; namely, $T_\omega(z_1) = t_1 \Rightarrow T_\omega(z_3) \notin \{t_2, t_3\}$.

The last two rows investigate the instrumental change from z_1 to z_4. The incentives to choose t_2 or t_3 increase, while the incentive to choose t_1 does not. The incentive requirement of choice rule (46) is not satisfied, and therefore, no choice restriction is generated.

Table 2 presents all the choice restrictions generated by applying choice rule (46) to each combination of treatment pairs $(t, t') \in \{t_1, t_2, t_3\}^2$ and to each pair of instrumental values $(z, z') \in \{z_1, z_2, z_3, z_4\}^2$.

Generating the Response Matrix

The choice restrictions of Table 2 can be used to determine the set of admissible response-types that the response-vector $\boldsymbol{S} = [T(z_1), T(z_2), T(z_3), T(z_4)]'$ can take. The first panel of Table 2 examines the case where $T(z) = t_1$ for $z \in \{z_1, z_2, z_3, z_4\}$. The first restriction states that if $T(z_1) = t_1$, then $T(z_2) = T(z_3) = t_1$. Given $T(z_1) = t_1$, there are only three possible response-types that comply with this choice restriction: $\boldsymbol{s}_1 = [t_1, t_1, t_1, t_1]'$, $\boldsymbol{s}_2 = [t_1, t_1, t_1, t_2]'$, and $\boldsymbol{s}_3 = [t_1, t_1, t_1, t_3]'$. The second and third choice restrictions of Table 2 are subsumed by the first restriction. The fourth choice restriction implies that the only admissible response-type for which $T(z_4) = t_1$ is $\boldsymbol{s}_1 = [t_1, t_1, t_1, t_1]'$.

The second panel of Table 2 examines the case where $T(z) = t_2$ for $z \in \{z_1, z_2, z_3, z_4\}$. The third panel examines the case where $T(z) = t_3$ for $z \in \{z_1, z_2, z_3, z_4\}$. We apply the elimination analysis of the first panel to the second and third panels. There are only nine admissible response-types that comply with each of the 12 choice restrictions of Table 2. Those are displayed in the response matrix below:[15]

$$\text{Respose Matrix } R = \begin{array}{c} \begin{matrix} s_1 & s_2 & s_3 & s_4 & s_5 & s_6 & s_7 & s_8 & s_9 \end{matrix} \\ \begin{bmatrix} t_1 & t_1 & t_1 & t_2 & t_2 & t_2 & t_3 & t_3 & t_3 \\ t_1 & t_1 & t_1 & t_2 & t_2 & t_2 & t_1 & t_2 & t_3 \\ t_1 & t_1 & t_1 & t_1 & t_2 & t_3 & t_3 & t_3 & t_3 \\ t_1 & t_2 & t_3 & t_2 & t_2 & t_2 & t_3 & t_3 & t_3 \end{bmatrix} \begin{matrix} z_1 \\ z_2 \\ z_3 \\ z_4 \end{matrix} \end{array} \tag{47}$$

[15] Under no choice restrictions, each of the four counterfactual choices ($T(z_1)$, $T(z_2)$, $T(z_3)$, and $T(z_4)$) can take any of the three treatment values (t_1, t_2, or t_3). Thus, the total number of potential response-types is 81. The choice restrictions in Table 2 are able to eliminate 72 out of the 81 possible response-types. The nine response-types that survive this elimination process are displayed in (47).

Identification and Estimation

Theorem T-2 uses the identification criteria in C-1 to recover all causal parameters that are identified.

Theorem T-2 *The response matrix* (47) *enables the identification of the following causal parameters*:

1 *All response-type probabilities* $P(\boldsymbol{S} = \boldsymbol{s}_j)$; $j = 1, \ldots, 9$.
2 *The expectation* (*and distribution*) *of the following counterfactual outcomes*:

Response-Types	Treatment Choices		
	t_1	t_2	t_3
Always-Takers	$E(Y(t_1) \mid \boldsymbol{S} = \boldsymbol{s}_1)$	$E(Y(t_2) \mid \boldsymbol{S} = \boldsymbol{s}_5)$	$E(Y(t_3) \mid \boldsymbol{S} = \boldsymbol{s}_9)$
Switchers	$E(Y(t_1) \mid \boldsymbol{S} = \boldsymbol{s}_4)$	$E(Y(t_2) \mid \boldsymbol{S} = \boldsymbol{s}_2)$	$E(Y(t_3) \mid \boldsymbol{S} = \boldsymbol{s}_3)$
	$E(Y(t_1) \mid \boldsymbol{S} = \boldsymbol{s}_7)$	$E(Y(t_2) \mid \boldsymbol{S} = \boldsymbol{s}_8)$	$E(Y(t_3) \mid \boldsymbol{S} = \boldsymbol{s}_6)$
Partially Identified	$E(Y(t_1) \mid \boldsymbol{S} \in \{\boldsymbol{s}_2, \boldsymbol{s}_3\})$	$E(Y(t_2) \mid \boldsymbol{S} \in \{\boldsymbol{s}_4, \boldsymbol{s}_6\})$	$E(Y(t_3) \mid \boldsymbol{S} \in \{\boldsymbol{s}_7, \boldsymbol{s}_8\})$

Proof See Appendix A.5. □

The response matrix (47) enables the researcher to use well-known econometric methods to evaluate causal effects. For instance, the first row (z_1) and the last row (z_4) of the response matrix (47) differ for two response-types: $\boldsymbol{s}_2$ and $\boldsymbol{s}_3$ take the value t_1 for z_1 and the values t_2 and t_3 for z_4, respectively. It is easy to show that the 2SLS estimator that uses the t_1-indicator $D_{t_t} = \mathbf{1}[T = t_1]$ as the treatment and employs only the IV-values z_1 and z_4 evaluates the causal effect of choosing t_1 versus not choosing t_1 for response-types $\boldsymbol{s}_2$ and $\boldsymbol{s}_3$:

$$\frac{E\big(Y \mid Z = z_1\big) - E\big(Y \mid Z = z_4\big)}{P\big(T = t_1 \mid Z = z_1\big) - P\big(T = t_1 \mid Z = z_1\big)} = E\Big(Y(t_1) - Y(\bar{t}_1) \mid \boldsymbol{S} \in \{\boldsymbol{s}_2, \boldsymbol{s}_3\}\Big), \tag{48}$$

where $Y(\bar{t}_1)$ stands for the counterfactual outcome of not choosing t_1:

$$\begin{aligned} &E(Y(\bar{t}_1) \mid \boldsymbol{S} \in \{\boldsymbol{s}_2, \boldsymbol{s}_3\}) \\ &= \frac{E\big(Y(t_2) \mid \boldsymbol{S} = \boldsymbol{s}_2\big)P(\boldsymbol{S} = \boldsymbol{s}_2) + E\big(Y(t_3) \mid \boldsymbol{S} = \boldsymbol{s}_3\big)P(\boldsymbol{S} = \boldsymbol{s}_3)}{P(\boldsymbol{S} = \boldsymbol{s}_2) + P(\boldsymbol{S} = \boldsymbol{s}_3)}. \end{aligned} \tag{49}$$

We can make the same analogy for the 2SLS that uses the indicator D_{t_3} for treatment and employs data from z_1 and z_2. The 2SLS estimator evaluates the causal effect of choosing t_3 versus not choosing t_3 for response-types $\boldsymbol{s_7}$ and $\boldsymbol{s_8}$; that is,

Table 3 Choice Restrictions Generated by the Traditional Incentive Matrix (50)

1	$T_\omega(z_1) = t_1$	$\Rightarrow$	No Restriction
2	$T_\omega(z_2) = t_1$	$\Rightarrow$	$T_\omega(z_1) \notin \{t_2, t_3\}$ and $T_\omega(z_3) \neq t_2$ and $T_\omega(z_4) \notin \{t_2, t_3\}$
3	$T_\omega(z_3) = t_1$	$\Rightarrow$	$T_\omega(z_1) \notin \{t_2, t_3\}$ and $T_\omega(z_2) \neq t_3$ and $T_\omega(z_4) \notin \{t_2, t_3\}$
4	$T_\omega(z_4) = t_1$	$\Rightarrow$	$T_\omega(z_1) \notin \{t_2, t_3\}$ and $T_\omega(z_2) \neq t_3$ and $T_\omega(z_3) \neq t_2$
5	$T_\omega(z_1) = t_2$	$\Rightarrow$	$T_\omega(z_2) \notin \{t_1, t_3\}$ and $T_\omega(z_3) \neq t_1$ and $T_\omega(z_4) \notin \{t_1, t_3\}$
6	$T_\omega(z_2) = t_2$	$\Rightarrow$	No Restriction
7	$T_\omega(z_3) = t_2$	$\Rightarrow$	$T_\omega(z_1) \neq t_3$ and $T_\omega(z_2) \notin \{t_1, t_3\}$ and $T_\omega(z_4) \notin \{t_1, t_3\}$
8	$T_\omega(z_4) = t_2$	$\Rightarrow$	$T_\omega(z_1) \neq t_3$ and $T_\omega(z_2) \notin \{t_1, t_3\}$ and $T_\omega(z_3) \neq t_1$
9	$T_\omega(z_1) = t_3$	$\Rightarrow$	$T_\omega(z_2) \neq t_1$ and $T_\omega(z_3) \notin \{t_1, t_2\}$ and $T_\omega(z_4) \notin \{t_1, t_2\}$
10	$T_\omega(z_2) = t_3$	$\Rightarrow$	$T_\omega(z_1) \neq t_2$ and $T_\omega(z_3) \notin \{t_1, t_2\}$ and $T_\omega(z_4) \notin \{t_1, t_2\}$
11	$T_\omega(z_3) = t_3$	$\Rightarrow$	No Restriction
12	$T_\omega(z_4) = t_3$	$\Rightarrow$	$T_\omega(z_1) \neq t_2$ and $T_\omega(z_2) \neq t_1$ and $T_\omega(z_3) \notin \{t_1, t_2\}$

This table presents all the choice restrictions generated by applying the choice rule (46) to each 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 (50)

$E(Y(t_3) - Y(t_3^-) \mid \boldsymbol{S} \in \{\boldsymbol{s}_7, \boldsymbol{s}_8\})$. Finally, the 2SLS that uses the treatment indicator D_{t_3} and IV-values z_3 and z_4 evaluates the causal effect of choosing t_2 versus not choosing t_2 for response-types $\boldsymbol{s}_4$ and $\boldsymbol{s}_6$; namely, $E(Y(t_2) - Y(t_2^-) \mid \boldsymbol{S} \in \{\boldsymbol{s}_4, \boldsymbol{s}_6\})$.

Benefits of Orthogonal Designs

The benefits of orthogonal designs become more apparent when we compare their results to those of more traditional designs. The plethora of identification results generated by orthogonal designs stand in sharp contrast to the paucity of identification results of standard designs. For instance, consider a conventional experimental design consisting of a control group with no incentives and three randomization groups dedicated solely to each treatment alternative. Specifically, we would have that z_1 incentivizes participants toward t_1, z_2 incentivizes participants toward t_2, z_3 incentivizes participants toward t_3, and z_4 does not incentivize participants toward any choice. This incentive pattern is described by the *incentive matrix* $\boldsymbol{L}$ in (50).

$$\text{Traditional Design } L = \begin{array}{c} \begin{matrix} t_1 & t_2 & t_3 \end{matrix} \\ \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} \end{array} \begin{matrix} z_1 \\ z_2 \\ z_3 \\ z_4 \end{matrix} \tag{50}$$

We apply the same approach used to examine the orthogonal design of incentive matrix (45) to the traditional design of incentive matrix (50). Table 3 presents all the choice restrictions generated by applying the choice rule (46) to each

Table 4 Causal Parameters Identified by Response Matrix (51)

1. The identified response-type probabilities are:

$P(\boldsymbol{S} \in \{s_1, s_2\})$	$P(\boldsymbol{S} \in \{s_1, s_3\})$	$P(\boldsymbol{S} \in \{s_2, s_5\})$	$P(\boldsymbol{S} \in \{s_3, s_5\})$
$P(\boldsymbol{S} \in \{s_4, s_6\})$	$P(\boldsymbol{S} \in \{s_4, s_9\})$	$P(\boldsymbol{S} \in \{s_6, s_{10}\})$	$P(\boldsymbol{S} \in \{s_7, s_8\})$
$P(\boldsymbol{S} \in \{s_7, s_{11}\})$	$P(\boldsymbol{S} \in \{s_8, s_{12}\})$	$P(\boldsymbol{S} \in \{s_9, s_{10}\})$	$P(\boldsymbol{S} \in \{s_{11}, s_{12}\})$

2. The expectation (and distribution) of the following counterfactual outcomes are identified:

t_1	t_2	t_3
$\mathrm{E}(Y(t_1) \mid S \in \{s_1, s_2\})$	$\mathrm{E}(Y(t_2) \mid S \in \{s_4, s_6\})$	$\mathrm{E}(Y(t_3) \mid S \in \{s_7, s_8\})$
$\mathrm{E}(Y(t_1) \mid S \in \{s_1, s_3\})$	$\mathrm{E}(Y(t_2) \mid S \in \{s_4, s_9\})$	$\mathrm{E}(Y(t_3) \mid S \in \{s_7, s_{11}\})$
$\mathrm{E}(Y(t_1) \mid S \in \{s_2, s_5\})$	$\mathrm{E}(Y(t_2) \mid S \in \{s_6, s_{10}\})$	$\mathrm{E}(Y(t_3) \mid S \in \{s_8, s_{12}\})$
$\mathrm{E}(Y(t_1) \mid S \in \{s_3, s_5\})$	$\mathrm{E}(Y(t_2) \mid S \in \{s_9, s_{10}\})$	$\mathrm{E}(Y(t_3) \mid S \in \{s_{11}, s_{12}\})$
$\mathrm{E}(Y(t_1) \mid S \in \{s_4, s_6, s_7, s_8\})$	$\mathrm{E}(Y(t_2) \mid S \in \{s_3, s_5, s_7, s_{11}\})$	$\mathrm{E}(Y(t_3) \mid S \in \{s_2, s_5, s_6, s_{10}\})$

This table presents all the causal parameters that are identified by response matrix (51)

combination of treatment pairs $(t, t') \in \{t_1, t_2, t_3\}^2$ and to each pair of instrumental values $(z, z') \in \{z_1, z_2, z_3, z_4\}^2$ of incentive matrix (50).

The choice restrictions of Table 3 eliminate 69 out of the 81 possible response-types. The 12 admissible response-types that comply with all the choice restrictions in Table 3 are presented in the response matrix below:

$$R = \begin{array}{c} \begin{matrix} s_1 & s_2 & s_3 & s_4 & s_5 & s_6 & s_7 & s_8 & s_9 & s_{10} & s_{11} & s_{12} \end{matrix} \\ \begin{bmatrix} t_1 & t_1 & t_1 & t_1 & t_1 & t_1 & t_1 & t_1 & t_2 & t_2 & t_3 & t_3 \\ t_1 & t_1 & t_1 & t_2 & t_2 & t_2 & t_2 & t_3 & t_2 & t_2 & t_2 & t_3 \\ t_1 & t_3 & t_2 & t_2 & t_3 & t_3 & t_3 & t_3 & t_2 & t_3 & t_3 & t_3 \\ t_1 & t_1 & t_1 & t_2 & t_1 & t_2 & t_3 & t_3 & t_2 & t_2 & t_3 & t_3 \end{bmatrix} \begin{matrix} z_1 \\ z_2 \\ z_3 \\ z_4 \end{matrix} \end{array} \tag{51}$$

Table 4 presents all the response-type probabilities and counterfactual outcomes that are identified by the response matrix (51). Response matrix (51) does not generate a single point-identified response-type probability. The matrix does not generate any point-identified counterfactual outcomes either. By choosing an orthogonal design for the incentive matrix, we secure the identification of causal parameters. Using a traditional design, we do not.

Appendix B applies our analysis to the study of Latin squares. We refer to Pinto and Navjeevan (2022)[16] for further discussion on how economic incentives shape choice restrictions in the IV model with multiple choices and heterogeneous agents.

[16] Pinto, R., and Navjeevan, M. (2022). Ordered, unordered and minimal monotonicity criteria [Unpublished manuscript]. Department of Economics, University of California, Los Angeles. https://www.rodrigopinto.net/_files/ugd/95d94d_1405f5376ae449a9b07f3bd3f37db161.pdf.

Conclusion

This paper provides a novel application of Rao's fundamental work on the design of experiments using orthogonal arrays. Rao's seminal ideas are widely used to determine efficient arrangements of treatment factors in RCTs. His method is well suited for experiments where the analyst can reliably assign treatment factors to randomization units. Unfortunately, social scientists can seldom impose treatment statuses. Most social experiments are consequently plagued by noncompliance, which undermines the random assignment of treatment statuses.

We repurpose Rao's original ideas to address the common challenges that noncompliance generates. We use a novel framework whereby orthogonal arrays denote a pattern of *choice incentives*. We combine the IV framework of Heckman and Pinto (2018) with the recently developed econometric tools in Pinto (2021a, b),[17] and Pinto and Navjeevan (2022)[18] to translate choice incentives into choice restrictions. These restrictions determine the set of economically justifiable counterfactual choices, which, in turn, enable the identification of causal parameters. We then show the benefits of using orthogonal arrays (rather than traditional approaches) for identifying causal parameters.

Our method broadly applies to IV models with multiple treatments, categorical instruments, and heterogeneous agents. We establish a tight link between the problem of the unobserved mixture of distributions and the identification of counterfactuals. We explore the notion of a response matrix. The matrix contains all the necessary information to examine the nonparametric identification of model counterfactuals. We apply mixture model methods to matrices to prove the identification of causal parameters.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s40953-022-00307-w.

Acknowledgements This research was supported by a MERIT award from the Eunice Kennedy Shriver National Institutes of Child Health and Human Development under award number R37HD06572 and a grant from a private donor. A web appendix (https://cehd.uchicago.edu/causal-models-choice-treat-appx) contains proofs of propositions.

Funding All funding sources are disclosed in the acknowledgments.

Declaration

Conflict of interest The authors declare that they have no competing interests that influenced the research or writing of this manuscript.

[17] Pinto, R. (2021a). Beyond intention to treat: Using the incentives in moving to opportunity to identify neighborhood effects [Unpublished manuscript]. Department of Economics, University of California, Los Angeles. https://www.rodrigopinto.net/_files/ugd/95d94d_90f491ec1afa45cf8ef1e9a77346c9a8.pdf. Pinto, R. (2021b). Economics of monotonicity conditions [Unpublished manuscript]. Departmentof Economics, University of California, Los Angeles.

[18] Pinto, R., and Navjeevan, M. (2022). Ordered, unordered and minimal monotonicity criteria [Unpublished manuscript]. Department of Economics, University of California, Los Angeles. https://www.rodrigopinto.net/_files/ugd/95d94d_1405f5376ae449a9b07f3bd3f37db161.pdf.

References

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.

Balke, A.A., and J. Pearl. 1993. *Nonparametric bounds on causal effects from partial compliance data* (Tech. Rep. No. R-199). University of California, Los Angeles.

Becker, G.S. 1962. Irrational behavior and economic theory. *Journal of Political Economy* 70: 1–13.

Frangakis, C.E., and D. Rubin. 2002. Principal stratification in causal inference. *Biometrics* 58 (1): 21–29.

Frisch, R. 1938. Autonomy of economic relations: Statistical versus theoretical relations in economic macrodynamics. Paper given at League of Nations. Reprinted in D.F. Hendry and M.S. Morgan (1995), The Foundations of Econometric Analysis, Cambridge University Press.

Haavelmo, T. 1943. The statistical implications of a system of simultaneous equations. *Econometrica* 11 (1): 1–12.

Haavelmo, T. 1944. The probability approach in econometrics. *Econometrica* 12 (Supplement), iii–iv and 1–115.

Heckman, J.J., and R. Pinto. 2015. Causal analysis after Haavelmo. *Econometric Theory* 31 (1): 115–151.

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

Heckman, J.J., and R. Robb. 1985. Alternative methods for evaluating the impact of interventions: An overview. *Journal of Econometrics* 30 (1–2): 239–267.

Keane, M.P., and K.I. Wolpin. 1997. The career decisions of young men. *Journal of Political Economy* 105 (3): 473–522.

Magnus, J., and H. Neudecker. 1999. *Matrix differential calculus with applications in statistics and econometrics*, 2nd ed. New York: Wiley.

McFadden, D. 1981. Econometric models of probabilistic choice. In *Structural analysis of discrete data with econometric applications*, ed. C. Manski and D. McFadden, 198–272. Cambridge, MA: MIT Press.

Pinto, R., and J.J. Heckman. 2021. The econometric model for causal policy analysis (Forthcoming, Annual Review of Economics).

Powell, J.L. 1994. Estimation of semiparametric models. In *Handbook of econometrics*, vol. 4, ed. R. Engle and D. McFadden, 2443–2521. Amsterdam: Elsevier.

Quandt, R.E. 1958. The estimation of the parameters of a linear regression system obeying two separate regimes. *Journal of the American Statistical Association* 53 (284): 873–880.

Quandt, R.E. 1972. A new approach to estimating switching regressions. *Journal of the American Statistical Association* 67 (338): 306–310.

Rao, C.R. 1943. Researches in the theory of the design of experiments and distribution problems connected with bivariate and multivariate populations. Thesis submitted to Calcutta University in lieu of 7th and 8th practical papers of the master's examination in statistics.

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. Retrieved from https://ci.nii.ac.jp/naid/10010345773/en/.

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. https://doi.org/10.2307/2983576.

Rao, C.R. 1949. On a class of arrangements. *Proceedings of the Edinburgh Mathematical Society* 8 (3): 119–125. https://doi.org/10.1017/S0013091500002650.

Rao, B.L.S.P. 1992. Identifiability for mixtures of distributions. *Identifiability in stochastic models: Characterization of probability distributions*, 183–228. Boston, MA: Academic Press.

Reiersöl, O. 1945. Confluence analysis by means of instrumental sets of variables. *Arkiv för Matematik, Astronomi och Fysik* 32A (4): 1–119.

Robins, J.M., and S. Greenland. 1992. Identifiability and exchangeability for direct and indirect effects. *Epidemiology* 3 (2): 143–155.

Rosenbaum, P.R., and D.B. Rubin. 1983. The central role of the propensity score in observational studies for causal effects. *Biometrika* 70 (1): 41–55.

Stinson, D. 2004. *Combinatorial designs: constructions and analysis*. Berlin: Springer.

Thaler, R.H. 2016. *Misbehaving: the making of behavioral economics*. New York: W. W. Norton & Company.

Theil, H. 1958. *Economic forecasts and policy* (No. 15). Amsterdam: North Holland Publishing Company.

Theil, H. 1971. *Principles of econometrics*. New York: Wiley.

Yakowitz, S.J., and J.D. Spragins. 1968. On the identifiability of finite mixtures. *Annals of Mathematical Statistics* 39 (1): 209–214.