Web Appendix for

Causal Inference of Social Experiments Using Orthogonal Designs

James J. Heckman and Rodrigo Pinto

April 13, 2022

Contents

A Proofs of Lemmas and Theorems

A.1 Proof of Theorem T-1:

Proof of T-1:

Proof. It suffices to prove that a general solution for $\boldsymbol{x}$ in the system of linear equations represented by $\boldsymbol{b} = \boldsymbol{B}\boldsymbol{x} \Rightarrow \boldsymbol{x}$ is given by:

$$\boldsymbol{b} = \boldsymbol{B}\boldsymbol{x} \Rightarrow \boldsymbol{x} = \boldsymbol{B}^{+}\boldsymbol{b} + (\boldsymbol{I} - \boldsymbol{B}^{+}\boldsymbol{B})\boldsymbol{\lambda} \tag{A.1}$$

where $\boldsymbol{\lambda}$ is an arbitrary real-valued $|\boldsymbol{b}|$-dimension vector, $\boldsymbol{I}$ is an identity matrix of the same dimension and $\boldsymbol{B}^{+}$ is the Moore—Penrose Pseudoinverse of matrix $\boldsymbol{B}$.[1]

In this proof we use the definition of the Moore—Penrose Pseudoinverse $\boldsymbol{B}^{+}$ and the fact that the matrix $\boldsymbol{B}^{+}$ is unique for a real-valued matrix $\boldsymbol{B}$. Matrix $\boldsymbol{B}^{+}$ has the following properties: (1) $\boldsymbol{B}\boldsymbol{B}^{+}\boldsymbol{B} = \boldsymbol{B}$; (2) $\boldsymbol{B}^{+}\boldsymbol{B}\boldsymbol{B}^{+} = \boldsymbol{B}^{+}$; (3) $\boldsymbol{B}^{+}\boldsymbol{B} = (\boldsymbol{B}^{+}\boldsymbol{B})'$; and (4) $\boldsymbol{B}\boldsymbol{B}^{+} = (\boldsymbol{B}\boldsymbol{B}^{+})'$. Properties (2)–(3) imply that $\boldsymbol{Q} = \boldsymbol{B}^{+}\boldsymbol{B}$ is an orthogonal projection operator, so $\boldsymbol{Q}^2 = \boldsymbol{Q}$ and $\boldsymbol{Q}' = \boldsymbol{Q}$:

$$\boldsymbol{Q}^2 = \boldsymbol{B}^{+}\boldsymbol{B}\boldsymbol{B}^{+}\boldsymbol{B} = \boldsymbol{B}^{+}\boldsymbol{B} = \boldsymbol{Q} \qquad \text{due to property (2)}$$

$$\boldsymbol{Q}' = (\boldsymbol{B}^{+}\boldsymbol{B})' = \boldsymbol{B}^{+}\boldsymbol{B} = \boldsymbol{Q} \qquad \text{due to property (3).}$$

Any vector $\boldsymbol{x}$ can be decomposed by a orthogonal $\boldsymbol{Q}$ projection as: $\boldsymbol{x} = \boldsymbol{Q}\boldsymbol{x} + (\boldsymbol{I} - \boldsymbol{Q})\boldsymbol{x}$. In our case, we have that $\boldsymbol{x} = \boldsymbol{B}^{+}\boldsymbol{B}\boldsymbol{x} + (\boldsymbol{I} - \boldsymbol{B}^{+}\boldsymbol{B})\boldsymbol{x}$. If vector $\boldsymbol{x}$ is a solution to the system

[1] We refer to Magnus and Neudecker (1999) for a general discussion of linear systems.

$\boldsymbol{b} = \boldsymbol{Bx}$, then it must be that:

$$\boldsymbol{Bx} = \boldsymbol{b} \Rightarrow \boldsymbol{x} = \boldsymbol{B}^{+}\boldsymbol{b} + (\boldsymbol{I} - \boldsymbol{B}^{+}\boldsymbol{B})\boldsymbol{x}$$

Moreover $\boldsymbol{b} = \boldsymbol{Bx} \Rightarrow \boldsymbol{b} = \boldsymbol{B}\big(\boldsymbol{B}^{+}\boldsymbol{b} + \boldsymbol{B}(\boldsymbol{I} - \boldsymbol{B}^{+}\boldsymbol{B})\boldsymbol{x}\big)$

But: $\boldsymbol{B}(\boldsymbol{I} - \boldsymbol{B}^{+}\boldsymbol{B}) = \boldsymbol{0}$ due to property (4) of $\boldsymbol{B}^{+}$

Thus : $\boldsymbol{B}(\boldsymbol{I} - \boldsymbol{B}^{+}\boldsymbol{B})\boldsymbol{\lambda} = \boldsymbol{0}$ for any real valued $\boldsymbol{\lambda}$

$$\Rightarrow \boldsymbol{b} = \boldsymbol{B}\big(\boldsymbol{B}^{+}\boldsymbol{b} + (\boldsymbol{I} - \boldsymbol{B}^{+}\boldsymbol{B})\boldsymbol{\lambda}\big)$$

$\therefore \widetilde{\boldsymbol{x}} = \boldsymbol{B}^{+}\boldsymbol{b} + (\boldsymbol{I} - \boldsymbol{B}^{+}\boldsymbol{B})\boldsymbol{\lambda}$ is also a solution as $\boldsymbol{b} = \boldsymbol{B}\tilde{\boldsymbol{x}}$ holds.

Thus $\widetilde{\boldsymbol{x}} = \boldsymbol{B}^{+}\boldsymbol{b} + \boldsymbol{K}\boldsymbol{\lambda}$ such that $\boldsymbol{K} = (\boldsymbol{I} - \boldsymbol{B}^{+}\boldsymbol{B})$ is a general solution.

□

A.2 Proof of Corollary C-1:

We use Theorem **T-1** to prove Corollary **C-1**.

Proof of C-1:

Proof. We apply the general solution for the matrix form of a system of linear equations to Equation (38) in the text. This generates $\boldsymbol{P_S} = \boldsymbol{B}_T^{+}\boldsymbol{P_Z} + \boldsymbol{K}_T\boldsymbol{\lambda}$. By hypothesis $\xi'\boldsymbol{K}_T = 0$, and thus $\xi'\boldsymbol{P}_S = \boldsymbol{\xi}'\boldsymbol{B}_T^{+}\boldsymbol{P_Z}$, which makes $\boldsymbol{\xi}'\boldsymbol{P_S}$ identified. By the same reasoning, $\boldsymbol{Q_S}(t) = \boldsymbol{B}_T^{+}\boldsymbol{Q}_Z(t) + \boldsymbol{K}_t\boldsymbol{\lambda}$. Thus $\zeta'\boldsymbol{K}_t = \boldsymbol{0}$ implies that $\boldsymbol{\zeta}'\boldsymbol{Q_S}(t) = \boldsymbol{\zeta}'\boldsymbol{B}_T^{+}\boldsymbol{Q}_Z(t)$ is identified. □

A.3 Proof of Corollary C-2:

Proof. According to C-1, Vector $\boldsymbol{P}_S$ is point-identified if and only if $\boldsymbol{\xi}'\boldsymbol{K}_T = \boldsymbol{0}$ for any $\boldsymbol{\xi}'$. Thus it must be the case that $\boldsymbol{K}_T = \boldsymbol{0}$. Since $\boldsymbol{K}_T = (\boldsymbol{I}_{N_{\boldsymbol{S}}} - \boldsymbol{B}_T^{+}\boldsymbol{B}_T)$, $\boldsymbol{K}_T = \boldsymbol{0}$ if and only if $\boldsymbol{I}_{N_{\boldsymbol{S}}} = \boldsymbol{B}_T^{+}\boldsymbol{B}_T$ which holds if and only if $\text{rank}(\boldsymbol{B}_T) = N_{\boldsymbol{S}}$, that is, $\boldsymbol{B}_T$ has full column-rank. From Theorem C-1, $\boldsymbol{P_S}$ is identified from $\boldsymbol{B}_T^{+}\boldsymbol{P_Z}$ if and only if $\text{rank}(\boldsymbol{B}_T) = N_{\boldsymbol{S}}$. The second equation follows from the same rationale. $\boldsymbol{K}_t = \boldsymbol{0}$ if and only if $\text{rank}(\boldsymbol{B}_t) = N_{\boldsymbol{S}}$. According to Theorem C-1, if $\boldsymbol{K}_t = \boldsymbol{0}$, then $\boldsymbol{Q_S}(t) = \boldsymbol{B}_t^{+}\boldsymbol{Q}_Z(t)$, and thereby $E(\kappa(Y(t)))$

can be expressed as:

$$
\begin{aligned}
E(\kappa(Y(t))) &= \sum_{n=1}^{N_{\boldsymbol{S}}} E(\kappa(Y(t)) \mid \boldsymbol{S} = \boldsymbol{s}_n) P(\boldsymbol{S} = \boldsymbol{s}_n) \\
&= \iota'_{N_{\boldsymbol{S}}} \boldsymbol{Q}_{\boldsymbol{S}}(t), \\
&= \iota'_{N_{\boldsymbol{S}}} \boldsymbol{B}_t^{+} \boldsymbol{Q}_Z(t),
\end{aligned}
$$

where $\iota_{N_{\boldsymbol{S}}}$ is a $N_{\boldsymbol{S}}$-dimensional vector of $1s$. □

A.4 Bounds for Response-Type Probabilities and Counterfactual Outcomes

Lemma **L-1** below uses linear Equations (39)–(38) and Theorem **T-1** to generate simple bounds for response-type probabilities and counterfactual outcomes:

Lemma L-1. *For the IV model (4)–(5), bounds for response-type probabilities $\boldsymbol{P}_{\boldsymbol{S}}$ given a response matrix $\boldsymbol{R}$ are given by:*

$$
\boldsymbol{P}_{\boldsymbol{S}} \in \left[\max\left(\boldsymbol{0}_{N_{\boldsymbol{S}}}, \boldsymbol{B}_T^{+} \boldsymbol{P}_Z + \min_{\boldsymbol{\lambda} \in \mathbb{R}^{N_{\boldsymbol{S}}}} \left(\boldsymbol{K}_T \boldsymbol{\lambda} \right) \right), \min\left(\boldsymbol{\iota}_{N_{\boldsymbol{S}}}, \boldsymbol{B}_T^{+} \boldsymbol{P}_Z + \max_{\boldsymbol{\lambda} \in \mathbb{R}^{N_{\boldsymbol{S}}}} \left(\boldsymbol{K}_T \boldsymbol{\lambda} \right) \right) \right], \tag{A.2a}
$$

where $\boldsymbol{\lambda}$ is an arbitrary real-valued vector of dimension $N_{\boldsymbol{S}}$. Bounds on $\boldsymbol{\lambda}$ come from the fact that $P_{\boldsymbol{S}}$ is a vector with probabilities defined on the unit simplex. Bounds for the expectation of outcomes by strata are given by:

$$
\left(\boldsymbol{B}_t^{+} \boldsymbol{Q}_Z(t) + \min_{\boldsymbol{\xi} \in \mathbb{R}^{N_{\boldsymbol{S}}}} \left(\boldsymbol{K}_t \boldsymbol{\xi} \right) \right) \leq \boldsymbol{Q}_{\boldsymbol{S}}(t) \leq \left(\boldsymbol{B}_t^{+} \boldsymbol{Q}_Z(t) + \max_{\boldsymbol{\xi} \in \mathbb{R}^{N_{\boldsymbol{S}}}} \left(\boldsymbol{K}_t \boldsymbol{\xi} \right) \right).[2] \tag{A.2b}
$$

where $\boldsymbol{\xi}$ is an arbitrary real-valued vector of dimension $N_{\boldsymbol{S}}$.

Proof. Equations (A.2a) and (A.2b) follow directly from the application of the general linear solution in Theorem **T-1** of the system of linear equations of Equations (38) and (39) respectively. The admissible ranges of λ in Equation (A.2a) comes from using the fact that $\boldsymbol{P}_{\boldsymbol{S}}$ are probabilities. □

[2] These bounds are not sharp because we do not use the full distribution of the data generating process in constructing them.

A.5 Proof of Theorem T-2

Proof. The identification of counterfactual outcomes stems from the identification criteria (43) and (44). Namely, $\boldsymbol{\xi}'\boldsymbol{Q}_S(t)$ is identified if and only if there exists a 9×1 vector $\boldsymbol{\xi}$ such that $\boldsymbol{\xi}'\boldsymbol{K}_t = \boldsymbol{0}'$, where $\boldsymbol{K}_t \equiv \boldsymbol{I}_9 - \boldsymbol{B}_t^+\boldsymbol{B}_t$; $t \in \{t_1, t_2, t_3\}$ and $\boldsymbol{B}_t = \boldsymbol{1}[\boldsymbol{R} = t]$. Matrix $\boldsymbol{K}_t$ is symmetric, thus we can adopt the equivalence criteria $\boldsymbol{K}_t\boldsymbol{\xi} = \boldsymbol{0}$ instead of $\boldsymbol{\xi}'\boldsymbol{K}_t = \boldsymbol{0}'$. Note that $\boldsymbol{K}_t$ is a $N_S \times N_S$ matrix and if the $\boldsymbol{s}$-column in $\boldsymbol{K}_t$ has only zeros, Then $E(Y(t) \mid \boldsymbol{S} = \boldsymbol{s})$ is identified. Equation (A.3) investigates the counterfactual outcomes for t_1:

$$\boldsymbol{B}_{t_1} = \begin{array}{c} \begin{matrix} \boldsymbol{s}_1 & \boldsymbol{s}_2 & \boldsymbol{s}_3 & \boldsymbol{s}_4 & \boldsymbol{s}_5 & \boldsymbol{s}_6 & \boldsymbol{s}_7 & \boldsymbol{s}_8 & \boldsymbol{s}_9 \end{matrix} \\ \begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \end{array} \Rightarrow \boldsymbol{K}_{t_1} = \frac{1}{2} \begin{array}{c} \begin{matrix} \boldsymbol{s}_1 & \boldsymbol{s}_2 & \boldsymbol{s}_3 & \boldsymbol{s}_4 & \boldsymbol{s}_5 & \boldsymbol{s}_6 & \boldsymbol{s}_7 & \boldsymbol{s}_8 & \boldsymbol{s}_9 \end{matrix} \\ \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2 \end{bmatrix} \end{array} \tag{A.3}$$

Columns $\boldsymbol{s}_1$, $\boldsymbol{s}_4$, and $\boldsymbol{s}_7$ in $\boldsymbol{K}_1$ are zero vectors. Thus $E(Y(t_1) \mid \boldsymbol{S} = \boldsymbol{s}_1)$, $E(Y(t_1) \mid \boldsymbol{S} = \boldsymbol{s}_4)$, and $E(Y(t_1) \mid \boldsymbol{S} = \boldsymbol{s}_7)$ are identified. Moreover $\boldsymbol{K}_t\boldsymbol{\xi} = \boldsymbol{0}$ for $\boldsymbol{\xi} = [0,1,1,0,0,0,0,0,0]'$. That is to say that $\boldsymbol{\xi}[\boldsymbol{s}] = 1; \boldsymbol{s} \in \{\boldsymbol{s}_2, \boldsymbol{s}_3\}$ and $\boldsymbol{\xi}[\boldsymbol{s}] = 0$ otherwise. Thus, according to the identification criteria (43) and (44), $E(Y(t_1) \mid \boldsymbol{S} = \boldsymbol{s}_2)P(\boldsymbol{S} = \boldsymbol{s}_2) + E(Y(t_1) \mid \boldsymbol{S} = \boldsymbol{s}_3)P(\boldsymbol{S} = \boldsymbol{s}_3)$ is identified. Moreover, $P(\boldsymbol{S} = \boldsymbol{s}_2) + P(\boldsymbol{S} = \boldsymbol{s}_3)$ is also identified (by setting Y to one). Thus,

$$\frac{E(Y(t_1) \mid \boldsymbol{S} = \boldsymbol{s}_2)P(\boldsymbol{S} = \boldsymbol{s}_2) + E(Y(t_1) \mid \boldsymbol{S} = \boldsymbol{s}_3)P(\boldsymbol{S} = \boldsymbol{s}_3)}{P(\boldsymbol{S} = \boldsymbol{s}_2) + P(\boldsymbol{S} = \boldsymbol{s}_3)} = E(Y(t_1) \mid \boldsymbol{S} \in \{\boldsymbol{s}_2, \boldsymbol{s}_3\})$$

is identified. Equation (A.4) investigates the counterfactual outcomes for t_2:

$$\boldsymbol{B}_{t_2} = \begin{array}{c} \begin{matrix} \boldsymbol{s}_1 & \boldsymbol{s}_2 & \boldsymbol{s}_3 & \boldsymbol{s}_4 & \boldsymbol{s}_5 & \boldsymbol{s}_6 & \boldsymbol{s}_7 & \boldsymbol{s}_8 & \boldsymbol{s}_9 \end{matrix} \\ \begin{bmatrix} 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \end{bmatrix} \end{array} \Rightarrow \boldsymbol{K}_{t_2} = \frac{1}{2} \begin{array}{c} \begin{matrix} \boldsymbol{s}_1 & \boldsymbol{s}_2 & \boldsymbol{s}_3 & \boldsymbol{s}_4 & \boldsymbol{s}_5 & \boldsymbol{s}_6 & \boldsymbol{s}_7 & \boldsymbol{s}_8 & \boldsymbol{s}_9 \end{matrix} \\ \begin{bmatrix} 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2 \end{bmatrix} \end{array} \tag{A.4}$$

Columns $\boldsymbol{s}_2$, $\boldsymbol{s}_5$, and $\boldsymbol{s}_8$ in $\boldsymbol{K}_2$ are zero vectors. Thus $E(Y(t_2) \mid \boldsymbol{S} = \boldsymbol{s}_2)$, $E(Y(t_2) \mid \boldsymbol{S} = \boldsymbol{s}_5)$, and $E(Y(t_2) \mid \boldsymbol{S} = \boldsymbol{s}_8)$ are identified. Moreover $\boldsymbol{K}_t\boldsymbol{\xi} = \boldsymbol{0}$ for $\boldsymbol{\xi} = [0,0,0,1,0,1,0,0,0]'$. That is to say that $\boldsymbol{\xi}[\boldsymbol{s}] = 1; \boldsymbol{s} \in \{\boldsymbol{s}_4, \boldsymbol{s}_6\}$ and $\boldsymbol{\xi}[\boldsymbol{s}] = 0$ otherwise. Thus, according to the identification criteria (43) and (44), $E(Y(t_2) \mid \boldsymbol{S} \in \{\boldsymbol{s}_4, \boldsymbol{s}_6\})$ is identified. Equation (A.5) investigates the counterfactual outcomes for t_3:

$$\boldsymbol{B}_{t_3} = \begin{array}{c} \begin{matrix} \boldsymbol{s}_1 & \boldsymbol{s}_2 & \boldsymbol{s}_3 & \boldsymbol{s}_4 & \boldsymbol{s}_5 & \boldsymbol{s}_6 & \boldsymbol{s}_7 & \boldsymbol{s}_8 & \boldsymbol{s}_9 \end{matrix} \\ \begin{bmatrix} 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \end{bmatrix} \end{array} \Rightarrow \boldsymbol{K}_{t_3} = \frac{1}{2} \begin{array}{c} \begin{matrix} \boldsymbol{s}_1 & \boldsymbol{s}_2 & \boldsymbol{s}_3 & \boldsymbol{s}_4 & \boldsymbol{s}_5 & \boldsymbol{s}_6 & \boldsymbol{s}_7 & \boldsymbol{s}_8 & \boldsymbol{s}_9 \end{matrix} \\ \begin{bmatrix} 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \end{array} \tag{A.5}$$

Columns $\boldsymbol{s}_3$, $\boldsymbol{s}_6$, and $\boldsymbol{s}_9$ in $\boldsymbol{K}_3$ are zero vectors. Thus $E(Y(t_2) \mid \boldsymbol{S} = \boldsymbol{s}_3)$, $E(Y(t_2) \mid \boldsymbol{S} = \boldsymbol{s}_6)$, and $E(Y(t_2) \mid \boldsymbol{S} = \boldsymbol{s}_9)$ are identified. Moreover $\boldsymbol{K}_t\boldsymbol{\xi} = \boldsymbol{0}$ for $\boldsymbol{\xi} = [0,0,0,1,0,2,0,0,0]'$. That is to say that $\boldsymbol{\xi}[\boldsymbol{s}] = 1; \boldsymbol{s} \in \{\boldsymbol{s}_4, \boldsymbol{s}_6\}$ and $\boldsymbol{\xi}[\boldsymbol{s}] = 0$ otherwise. Thus, according to the identification criteria (43) and (44), $E(Y(t_2) \mid \boldsymbol{S} \in \{\boldsymbol{s}_4, \boldsymbol{s}_6\})$ is identified.

We can investigate the identification of response-type probabilities by setting the outcome to one. The identification results regarding matrix $\boldsymbol{K}_{t_1}$ assures the identification of three response-type probabilities: $P(\boldsymbol{S} = \boldsymbol{s}_1)$, $P(\boldsymbol{S} = \boldsymbol{s}_4)$, and $P(\boldsymbol{S} = \boldsymbol{s}_7)$. Matrix $\boldsymbol{K}_{t_2}$ identifies three additional response-type probabilities: $P(\boldsymbol{S} = \boldsymbol{s}_2)$, $P(\boldsymbol{S} = \boldsymbol{s}_5)$, and $P(\boldsymbol{S} = \boldsymbol{s}_8)$. Matrix $\boldsymbol{K}_{t_2}$ identifies the remaining response-type probabilities: $P(\boldsymbol{S} = \boldsymbol{s}_3)$, $P(\boldsymbol{S} = \boldsymbol{s}_6)$, and $P(\boldsymbol{S} = \boldsymbol{s}_9)$. □

B Application to Latin Squares

Different incentive designs generate choice model with distinct properties. It is possible to tailor the incentives to generate choice models with desirable properties. It is often the case that well-known design generate non-trivial model properties. We examine the Latin square design of the incentive matrix (B.1) to illustrate this fact.

A Latin square is a matrix whose elements occurs exactly once in each row and column. The matrix is necessarily square and the number of unique elements is also the matrix dimension. Incentive matrix (B.1) is an example of Latin square of dimension three:

$$\textbf{Latin Square Incentive Matrix } \boldsymbol{L} = \begin{matrix} & t_1 & t_2 & t_3 & \\ & \left[\begin{matrix} 3 \\ 2 \\ 1 \end{matrix}\right. & \begin{matrix} 1 \\ 3 \\ 2 \end{matrix} & \left.\begin{matrix} 2 \\ 1 \\ 3 \end{matrix}\right] & \begin{matrix} z_1 \\ z_2 \\ z_3 \end{matrix} \end{matrix} \tag{B.1}$$

We can represent the Latin square as an orthogonal array by expressing each of its entries as a triple (r, c, e), where r is the row, c is the column, and e is the entry value. Matrix (B.2) displays the orthogonal representation of the incentive matrix (B.1). The representation constitutes an orthogonal array of the type $OA(9, 3, 3, 2)$.

$$\begin{array}{ccc} r & c & e \end{array}$$
$$\begin{bmatrix} 1 & 1 & 3 \\ 1 & 2 & 1 \\ 1 & 3 & 2 \\ 2 & 1 & 2 \\ 2 & 2 & 3 \\ 2 & 3 & 1 \\ 3 & 1 & 1 \\ 3 & 2 & 2 \\ 3 & 3 & 3 \end{bmatrix} \tag{B.2}$$

Each column of incentive matrix (B.1) presents a ranking of the incentives induced by IV-values towards a treatment choice. The first column implies that z_1 offers the highest value towards t_1 follows by z_2 and the z_3. The ranking for t_2 is that z_2 offers the highest incentive followed by z_3 and z_1. We can apply choice rule (46) to incentive matrix (B.1) in the same fashion we examined the incentive matrix (45). The incentive matrix generates nine choice restrictions listed in Table B.1.

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

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

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 $(z, z') \in \{z_1, z_2, z_3\}$ of the incentive matrix (B.1).

There are seven response-types that survive the choice restrictions in Tale B.1. The surviving response-types are stacked in the response matrix (B.3).

$$\boldsymbol{R} = \begin{array}{c} \begin{matrix} \boldsymbol{s}_1 & \boldsymbol{s}_2 & \boldsymbol{s}_3 & \boldsymbol{s}_4 & \boldsymbol{s}_5 & \boldsymbol{s}_6 & \boldsymbol{s}_7 \end{matrix} \\ \begin{bmatrix} t_1 & t_1 & t_1 & t_1 & t_2 & t_3 & t_3 \\ t_1 & t_1 & t_2 & t_2 & t_2 & t_2 & t_3 \\ t_1 & t_3 & t_2 & t_3 & t_2 & t_3 & t_3 \end{bmatrix} \begin{matrix} z_1 \\ z_2 \\ z_3 \end{matrix} \end{array} \tag{B.3}$$

Response matrix (B.3) has a useful property: it is possible to reorder rows and columns such that the values of any of the neighborhood choices $t \in \{t_1, t_2, t_3\}$ lie in the *lower triangular* portion of the matrix. The reordered matrices are displayed in (B.4)–(B.6).

$$\begin{array}{c} \begin{matrix} \boldsymbol{s}_1 & \boldsymbol{s}_2 & \boldsymbol{s}_3 & \boldsymbol{s}_4 & \boldsymbol{s}_5 & \boldsymbol{s}_6 & \boldsymbol{s}_7 \end{matrix} \\ \begin{bmatrix} t_1 & t_3 & t_2 & t_3 & t_2 & t_3 & t_3 \\ t_1 & t_1 & t_2 & t_2 & t_2 & t_2 & t_3 \\ t_1 & t_1 & t_1 & t_1 & t_2 & t_3 & t_3 \end{bmatrix} \begin{matrix} z_3 \\ z_2 \\ z_1 \end{matrix} \end{array} \tag{B.4}$$

$$\begin{array}{c} \begin{matrix} \boldsymbol{s}_3 & \boldsymbol{s}_4 & \boldsymbol{s}_5 & \boldsymbol{s}_6 & \boldsymbol{s}_7 & \boldsymbol{s}_2 & \boldsymbol{s}_1 \end{matrix} \\ \begin{bmatrix} t_2 & t_1 & t_1 & t_3 & t_3 & t_1 & t_1 \\ t_2 & t_2 & t_3 & t_3 & t_3 & t_3 & t_1 \\ t_2 & t_2 & t_2 & t_2 & t_3 & t_1 & t_1 \end{bmatrix} \begin{matrix} z_1 \\ z_3 \\ z_2 \end{matrix} \end{array} \tag{B.5}$$

$$\begin{array}{c} \begin{matrix} \boldsymbol{s}_7 & \boldsymbol{s}_6 & \boldsymbol{s}_2 & \boldsymbol{s}_4 & \boldsymbol{s}_5 & \boldsymbol{s}_1 & \boldsymbol{s}_3 \end{matrix} \\ \begin{bmatrix} t_3 & t_2 & t_1 & t_2 & t_2 & t_1 & t_2 \\ t_3 & t_3 & t_1 & t_1 & t_2 & t_1 & t_1 \\ t_3 & t_3 & t_3 & t_3 & t_2 & t_1 & t_2 \end{bmatrix} \begin{matrix} z_2 \\ z_1 \\ z_3 \end{matrix} \end{array} \tag{B.6}$$

Heckman and Pinto (2018) show that the triangular property of the response matrix in (B.4)–(B.6) is a necessary and sufficient criteria for "unordered monotonicity" defined in their paper to hold. That condition is a restriction on $\boldsymbol{R}$ and states that a change in the instrument cannot induce some agents towards a choice while inducing others against the same choice. Formally, unordered monotonicity states that for any pair of instrumental values z, z' and any choice t, we must have that:

$$\mathbf{1}[T_\omega(z) = t] \geq \mathbf{1}[T_\omega(z') = t] \text{ for all } \omega \in \Omega \text{ or } \mathbf{1}[T_\omega(z) = t] \leq \mathbf{1}[T_\omega(z') = t] \text{ for all } \omega \in \Omega. \tag{B.7}$$

The unordered monotonicity in our three-choice and three-valued instrument consists of nine inequalities—one for each combination of $t \in \{t_1, t_2, t_3\}$ and $(z, z') \in \{z_1, z_2, z_3\}$. Table B.2 lists the (unique) set of inequalities capable to generate response matrix (B.3). These inequalities are equivalent to the nine choice restrictions of Table B.1 as both generate the same response matrix.

Table B.2: Unordered Monotonicity Conditions that Generate Response Matrix (B.1)

	Z-pairs	T	Unordered Monotonicity Conditions
Relation 1	(z_1, z_2)	t_1	$\mathbf{1}[T_\omega(z_1) = t_1] \;\geq\; \mathbf{1}[T_\omega(z_2) = t_1]$
Relation 2	(z_2, z_3)	t_1	$\mathbf{1}[T_\omega(z_2) = t_1] \;\geq\; \mathbf{1}[T_\omega(z_3) = t_1]$
Relation 3	(z_3, z_1)	t_1	$\mathbf{1}[T_\omega(z_3) = t_1] \;\leq\; \mathbf{1}[T_\omega(z_1) = t_1]$
Relation 4	(z_1, z_2)	t_2	$\mathbf{1}[T_\omega(z_1) = t_2] \;\leq\; \mathbf{1}[T_\omega(z_2) = t_2]$
Relation 5	(z_2, z_3)	t_2	$\mathbf{1}[T_\omega(z_2) = t_2] \;\geq\; \mathbf{1}[T_\omega(z_3) = t_2]$
Relation 6	(z_3, z_1)	t_2	$\mathbf{1}[T_\omega(z_3) = t_2] \;\geq\; \mathbf{1}[T_\omega(z_1) = t_2]$
Relation 7	(z_1, z_2)	t_3	$\mathbf{1}[T_\omega(z_1) = t_3] \;\geq\; \mathbf{1}[T_\omega(z_2) = t_3]$
Relation 8	(z_2, z_3)	t_3	$\mathbf{1}[T_\omega(z_2) = t_3] \;\leq\; \mathbf{1}[T_\omega(z_3) = t_3]$
Relation 9	(z_3, z_1)	t_3	$\mathbf{1}[T_\omega(z_3) = t_3] \;\leq\; \mathbf{1}[T_\omega(z_1) = t_3]$

Heckman and Pinto (2018) show that unordered monotonicity renders the identification of several counterfactual outcomes. They also show that the condition is equivalent to the proposition that each choice indicator can be expressed as a separable inequality of the propensity score and a uniformity distributed random variable that depends on the unobserved confounder $\boldsymbol{V}$, that is, $\mathbf{1}[Y = t] = \mathbf{1}[P(T = t \mid Z) \geq U_t]; Ut \sim Unif[0, 1]$.

References

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

MAGNUS, J. AND H. NEUDECKER (1999): *Matrix Differential Calculus with Applications in Statistics and Econometrics*, Wiley, 2 ed.