Web Appendix

A Appendix - Proofs

Each proofs are organized by sections described in the following sections.

A.1 Proof of Lemma L-1:

Proof. Lemma **L-1** comes as a direct consequence of the Weak Axiom of Revealed Preference (WARP). Some notation is necessary to assess the WARP. If a bundle (t, g) is chosen over (t', g') by agent ω when both were available, i.e. $g, g' \in \mathcal{B}_\omega$, where $\mathcal{B}_\omega$ denotes the budget set for consumption goods of agent ω, then bundle (t, g) is said to be *(directly) strictly revealed preferred* to (t', g') by ω. Notationally, we write $(t, g) \succ^d_\omega (t', g')$. The Weak Axiom of Revealed Preference (WARP) states that if (t, g) is revealed preferred to (t', g') then it cannot be the case that (t', g') is revealed preferred to (t, g). Notationally, we write:

$$\textbf{WARP:} \quad (t, g) \succ^d_\omega (t', g') \Rightarrow (t', g') \nsucc^d_\omega (t, g). \tag{87}$$

Let the neighborhood choice of family ω under a voucher $z \in \{z_c, z_8, z_e\}$ be $t \in \{t_l, t_m, t_h\}$, namely, $T_\omega(z) = t$. Thus there must exist an unobserved bundle (t, g^*) where $g^* \in \mathcal{B}_\omega(z, t)$ that is revealed preferred to each of the bundles $\{(t', g'); g' \in \mathcal{B}_\omega(z, t')\}$ for any $t' \neq t$. This means that the bundle is (t, g^*) is strictly revealed preferred to each bundle $\{(t', g'); g' \in \mathcal{B}_\omega(z, t')\}$. In summary we have that:

$$T_\omega(z) = t \Rightarrow \exists g^* \in \mathcal{B}_\omega(z, t) \text{ such that } (t, g^*) \succ^d_\omega (t', g') \,\forall\, g' \in \mathcal{B}_\omega(z, t'). \tag{88}$$

Now consider a voucher change from z to z' such that the consumption budget associated with t does not decrease, i.e., $\mathcal{B}_\omega(z, t) \subseteq \mathcal{B}_\omega(z', t)$. This condition assures that the bundle (t, g^*) is still available under z'. Also, suppose that the consumption budget associated with t' does not increase, i.e., $\mathcal{B}_\omega(z, t') \supseteq \mathcal{B}_\omega(z', t')$. This condition assures that all bundles (t', g') available under z' were also available under z. Thus, according to WARP, it cannot be the case that a bundle $\{(t', g'); g' \in \mathcal{B}_\omega(z', t')\}$ is directed revealed preferred to the bundle (t, g^*) in $\mathcal{B}_\omega(z', t)$. As a consequence, family ω cannot choose t' under z', i.e. $T_\omega(z') \neq t'$. Equation (89) summarizes the revealed preference analysis:

$$\text{If} \quad T_\omega(z) = t \quad \text{and} \quad \mathcal{B}_\omega(z, t) \subseteq \mathcal{B}_\omega(z', t) \;,\; \mathcal{B}_\omega(z, t') \supseteq \mathcal{B}_\omega(z', t') \quad \text{then} \quad T_\omega(z') \neq t'. \tag{89}$$

Moreover, we have that $\boldsymbol{L}[z, t] \leq \boldsymbol{L}[z', t] \Rightarrow \mathcal{B}_\omega(z, t) \subset \mathcal{B}_\omega(z', t)$ (equation (5)). Thus the budget set relation $\mathcal{B}_\omega(z, t) \subseteq \mathcal{B}_\omega(z', t)$ in (89) can be replaced by the incentive inequality $\boldsymbol{L}[z, t] \leq \boldsymbol{L}[z', t]$ (or equivalently $0 \leq \boldsymbol{L}[z', t] - \boldsymbol{L}[z, t]$) On the other hand, the budget set relation $\mathcal{B}_\omega(z, t') \supseteq \mathcal{B}_\omega(z', t')$ can be replaced by $\boldsymbol{L}[z, t'] \geq \boldsymbol{L}[z', t']$ (or equivalently $\boldsymbol{L}[z', t'] - \boldsymbol{L}[z, t'] \leq 0$). This replacement generates expression (90), as desired.

$$\text{If} \quad T_\omega(z) = t \quad \text{and} \quad \boldsymbol{L}[z', t'] - \boldsymbol{L}[z, t'] \leq 0 \leq \boldsymbol{L}[z', t] - \boldsymbol{L}[z, t] \quad \text{then} \quad T_\omega(z') \neq t'. \tag{90}$$

□

A.2 Proof of Lemma L-2:

Proof. Table (A.1) applies the WARP choice rule of Lemma **L-1**, that is:

$$\text{If} \quad T_\omega(z) = t \quad \text{and} \quad \boldsymbol{L}[z', t'] - \boldsymbol{L}[z, t'] \leq 0 \leq \boldsymbol{L}[z', t] - \boldsymbol{L}[z, t] \quad \text{then} \quad T_\omega(z') \neq t'.$$

to the MTO incentive matrix (2) that is given by :

$$\boldsymbol{L} = \begin{array}{c} \begin{matrix} t_h & t_m & t_l \end{matrix} \\ \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix} \end{array} \begin{matrix} z_c \\ z_8 \\ z_e \end{matrix}$$

The symbol $\not\leq$ in the table means that the values in the left-hand side of the inequality is **not** less or equal than the value in the right-hand side of the inequality and thereby the choice rule does not apply.

Table A.1: Choice Restrictions Due to WARP

Revealed Choice	Incentive Inequalities					Choice Statement
$T_\omega(z_c) = t_h,$	$\boldsymbol{L}[z_8, t_m] - \boldsymbol{L}[z_c, t_m] = 1$	$\not\leq$	0	$\leq$	$0 = \boldsymbol{L}[z_8, t_h] - \boldsymbol{L}[z_c, t_h]$	–
$T_\omega(z_c) = t_h,$	$\boldsymbol{L}[z_e, t_m] - \boldsymbol{L}[z_c, t_m] = 0$	$\leq$	0	$\leq$	$0 = \boldsymbol{L}[z_e, t_h] - \boldsymbol{L}[z_c, t_h]$	$T_\omega(z_e) \neq t_m$
$T_\omega(z_c) = t_h,$	$\boldsymbol{L}[z_8, t_l] - \boldsymbol{L}[z_c, t_l] = 1$	$\not\leq$	0	$\leq$	$0 = \boldsymbol{L}[z_8, t_h] - \boldsymbol{L}[z_c, t_h]$	–
$T_\omega(z_c) = t_h,$	$\boldsymbol{L}[z_e, t_l] - \boldsymbol{L}[z_c, t_l] = 1$	$\not\leq$	0	$\leq$	$0 = \boldsymbol{L}[z_e, t_h] - \boldsymbol{L}[z_c, t_h]$	–
$T_\omega(z_c) = t_m,$	$\boldsymbol{L}[z_8, t_h] - \boldsymbol{L}[z_c, t_h] = 0$	$\leq$	0	$\leq$	$1 = \boldsymbol{L}[z_8, t_m] - \boldsymbol{L}[z_c, t_m]$	$T_\omega(z_8) \neq t_h$
$T_\omega(z_c) = t_m,$	$\boldsymbol{L}[z_e, t_h] - \boldsymbol{L}[z_c, t_h] = 0$	$\leq$	0	$\leq$	$0 = \boldsymbol{L}[z_e, t_m] - \boldsymbol{L}[z_c, t_m]$	$T_\omega(z_e) \neq t_h$
$T_\omega(z_c) = t_m,$	$\boldsymbol{L}[z_8, t_l] - \boldsymbol{L}[z_c, t_l] = 1$	$\not\leq$	0	$\leq$	$1 = \boldsymbol{L}[z_8, t_m] - \boldsymbol{L}[z_c, t_m]$	–
$T_\omega(z_c) = t_m,$	$\boldsymbol{L}[z_e, t_l] - \boldsymbol{L}[z_c, t_l] = 1$	$\not\leq$	0	$\leq$	$0 = \boldsymbol{L}[z_e, t_m] - \boldsymbol{L}[z_c, t_m]$	–
$T_\omega(z_c) = t_l,$	$\boldsymbol{L}[z_8, t_h] - \boldsymbol{L}[z_c, t_h] = 0$	$\leq$	0	$\leq$	$1 = \boldsymbol{L}[z_8, t_l] - \boldsymbol{L}[z_c, t_l]$	$T_\omega(z_8) \neq t_h$
$T_\omega(z_c) = t_l,$	$\boldsymbol{L}[z_e, t_h] - \boldsymbol{L}[z_c, t_h] = 0$	$\leq$	0	$\leq$	$1 = \boldsymbol{L}[z_e, t_l] - \boldsymbol{L}[z_c, t_l]$	$T_\omega(z_e) \neq t_h$
$T_\omega(z_c) = t_l,$	$\boldsymbol{L}[z_8, t_m] - \boldsymbol{L}[z_c, t_m] = 1$	$\not\leq$	0	$\leq$	$1 = \boldsymbol{L}[z_8, t_l] - \boldsymbol{L}[z_c, t_l]$	–
$T_\omega(z_c) = t_l,$	$\boldsymbol{L}[z_e, t_m] - \boldsymbol{L}[z_c, t_m] = 0$	$\leq$	0	$\leq$	$1 = \boldsymbol{L}[z_e, t_l] - \boldsymbol{L}[z_c, t_l]$	$T_\omega(z_e) \neq t_m$
$T_\omega(z_8) = t_h,$	$\boldsymbol{L}[z_c, t_m] - \boldsymbol{L}[z_8, t_m] = -1$	$\leq$	0	$\leq$	$0 = \boldsymbol{L}[z_c, t_h] - \boldsymbol{L}[z_8, t_h]$	$T_\omega(z_c) \neq t_m$
$T_\omega(z_8) = t_h,$	$\boldsymbol{L}[z_e, t_m] - \boldsymbol{L}[z_8, t_m] = -1$	$\leq$	0	$\leq$	$0 = \boldsymbol{L}[z_e, t_h] - \boldsymbol{L}[z_8, t_h]$	$T_\omega(z_e) \neq t_m$
$T_\omega(z_8) = t_h,$	$\boldsymbol{L}[z_c, t_l] - \boldsymbol{L}[z_8, t_l] = -1$	$\leq$	0	$\leq$	$0 = \boldsymbol{L}[z_c, t_h] - \boldsymbol{L}[z_8, t_h]$	$T_\omega(z_c) \neq t_l$
$T_\omega(z_8) = t_h,$	$\boldsymbol{L}[z_e, t_l] - \boldsymbol{L}[z_8, t_l] = 0$	$\leq$	0	$\leq$	$0 = \boldsymbol{L}[z_e, t_h] - \boldsymbol{L}[z_8, t_h]$	$T_\omega(z_e) \neq t_l$
$T_\omega(z_8) = t_m,$	$\boldsymbol{L}[z_c, t_h] - \boldsymbol{L}[z_8, t_h] = 0$	$\leq$	0	$\not\leq$	$-1 = \boldsymbol{L}[z_c, t_m] - \boldsymbol{L}[z_8, t_m]$	–
$T_\omega(z_8) = t_m,$	$\boldsymbol{L}[z_e, t_h] - \boldsymbol{L}[z_8, t_h] = 0$	$\leq$	0	$\not\leq$	$-1 = \boldsymbol{L}[z_e, t_m] - \boldsymbol{L}[z_8, t_m]$	–
$T_\omega(z_8) = t_m,$	$\boldsymbol{L}[z_c, t_l] - \boldsymbol{L}[z_8, t_l] = -1$	$\leq$	0	$\not\leq$	$-1 = \boldsymbol{L}[z_c, t_m] - \boldsymbol{L}[z_8, t_m]$	–
$T_\omega(z_8) = t_m,$	$\boldsymbol{L}[z_e, t_l] - \boldsymbol{L}[z_8, t_l] = 0$	$\leq$	0	$\not\leq$	$-1 = \boldsymbol{L}[z_e, t_m] - \boldsymbol{L}[z_8, t_m]$	–
$T_\omega(z_8) = t_l,$	$\boldsymbol{L}[z_c, t_h] - \boldsymbol{L}[z_8, t_h] = 0$	$\leq$	0	$\not\leq$	$-1 = \boldsymbol{L}[z_c, t_l] - \boldsymbol{L}[z_8, t_l]$	–
$T_\omega(z_8) = t_l,$	$\boldsymbol{L}[z_e, t_h] - \boldsymbol{L}[z_8, t_h] = 0$	$\leq$	0	$\leq$	$0 = \boldsymbol{L}[z_e, t_l] - \boldsymbol{L}[z_8, t_l]$	$T_\omega(z_e) \neq t_h$
$T_\omega(z_8) = t_l,$	$\boldsymbol{L}[z_c, t_m] - \boldsymbol{L}[z_8, t_m] = -1$	$\leq$	0	$\not\leq$	$-1 = \boldsymbol{L}[z_c, t_l] - \boldsymbol{L}[z_8, t_l]$	–
$T_\omega(z_8) = t_l,$	$\boldsymbol{L}[z_e, t_m] - \boldsymbol{L}[z_8, t_m] = -1$	$\leq$	0	$\leq$	$0 = \boldsymbol{L}[z_e, t_l] - \boldsymbol{L}[z_8, t_l]$	$T_\omega(z_e) \neq t_m$
$T_\omega(z_e) = t_h,$	$\boldsymbol{L}[z_c, t_m] - \boldsymbol{L}[z_e, t_m] = 0$	$\leq$	0	$\leq$	$0 = \boldsymbol{L}[z_c, t_h] - \boldsymbol{L}[z_e, t_h]$	$T_\omega(z_c) \neq t_m$
$T_\omega(z_e) = t_h,$	$\boldsymbol{L}[z_8, t_m] - \boldsymbol{L}[z_e, t_m] = 1$	$\not\leq$	0	$\leq$	$0 = \boldsymbol{L}[z_8, t_h] - \boldsymbol{L}[z_e, t_h]$	–
$T_\omega(z_e) = t_h,$	$\boldsymbol{L}[z_c, t_l] - \boldsymbol{L}[z_e, t_l] = -1$	$\leq$	0	$\leq$	$0 = \boldsymbol{L}[z_c, t_h] - \boldsymbol{L}[z_e, t_h]$	$T_\omega(z_c) \neq t_l$
$T_\omega(z_e) = t_h,$	$\boldsymbol{L}[z_8, t_l] - \boldsymbol{L}[z_e, t_l] = 0$	$\leq$	0	$\leq$	$0 = \boldsymbol{L}[z_8, t_h] - \boldsymbol{L}[z_e, t_h]$	$T_\omega(z_8) \neq t_l$
$T_\omega(z_e) = t_m,$	$\boldsymbol{L}[z_c, t_h] - \boldsymbol{L}[z_e, t_h] = 0$	$\leq$	0	$\leq$	$0 = \boldsymbol{L}[z_c, t_m] - \boldsymbol{L}[z_e, t_m]$	$T_\omega(z_c) \neq t_h$
$T_\omega(z_e) = t_m,$	$\boldsymbol{L}[z_8, t_h] - \boldsymbol{L}[z_e, t_h] = 0$	$\leq$	0	$\leq$	$1 = \boldsymbol{L}[z_8, t_m] - \boldsymbol{L}[z_e, t_m]$	$T_\omega(z_8) \neq t_h$
$T_\omega(z_e) = t_m,$	$\boldsymbol{L}[z_c, t_l] - \boldsymbol{L}[z_e, t_l] = -1$	$\leq$	0	$\leq$	$0 = \boldsymbol{L}[z_c, t_m] - \boldsymbol{L}[z_e, t_m]$	$T_\omega(z_c) \neq t_l$
$T_\omega(z_e) = t_m,$	$\boldsymbol{L}[z_8, t_l] - \boldsymbol{L}[z_e, t_l] = 0$	$\leq$	0	$\leq$	$1 = \boldsymbol{L}[z_8, t_m] - \boldsymbol{L}[z_e, t_m]$	$T_\omega(z_8) \neq t_l$
$T_\omega(z_e) = t_l,$	$\boldsymbol{L}[z_c, t_h] - \boldsymbol{L}[z_e, t_h] = 0$	$\leq$	0	$\not\leq$	$-1 = \boldsymbol{L}[z_c, t_l] - \boldsymbol{L}[z_e, t_l]$	–
$T_\omega(z_e) = t_l,$	$\boldsymbol{L}[z_8, t_h] - \boldsymbol{L}[z_e, t_h] = 0$	$\leq$	0	$\leq$	$0 = \boldsymbol{L}[z_8, t_l] - \boldsymbol{L}[z_e, t_l]$	$T_\omega(z_8) \neq t_h$
$T_\omega(z_e) = t_l,$	$\boldsymbol{L}[z_c, t_m] - \boldsymbol{L}[z_e, t_m] = 0$	$\leq$	0	$\not\leq$	$-1 = \boldsymbol{L}[z_c, t_l] - \boldsymbol{L}[z_e, t_l]$	–
$T_\omega(z_e) = t_l,$	$\boldsymbol{L}[z_8, t_m] - \boldsymbol{L}[z_e, t_m] = 1$	$\not\leq$	0	$\leq$	$0 = \boldsymbol{L}[z_8, t_l] - \boldsymbol{L}[z_e, t_l]$	–

This table applies the WARP choice rule of Lemma **L-1**, that is:

$$\text{If} \quad T_\omega(z) = t \quad \text{and} \quad \boldsymbol{L}[z', t'] - \boldsymbol{L}[z, t'] \leq 0 \leq \boldsymbol{L}[z', t] - \boldsymbol{L}[z, t] \quad \text{then} \quad T_\omega(z') \neq t'.$$

to the MTO incentive matrix (2) that is given by :

$$\boldsymbol{L} = \begin{array}{c} \\ z_c \\ z_8 \\ z_e \end{array} \begin{array}{c} \begin{array}{ccc} t_h & t_m & t_l \end{array} \\ \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix} \end{array}$$

The symbol $\not\leq$ means that the values in the left-hand side of the inequality is **not** less or equal than the value in the righhand side of the inequality and thereby the choice rule does not apply.

Table A.2 list the choice restrictions due to WARP. These choice restrictions are summarized into the six choice restrictions displayed in the lemma. The first column of the table also indicates

the choice restrictions used to generate the six choice restrictions mentioned in the lemma. The choice restrictions 7 and 8 of Table (A.2) are irrelevant given the previous restrictions. That is to say that these two restrictions do not eliminate any further response-type that was already eliminated by the previous restrictions.

Table A.2: List of Choice Restrictions Due to WARP

Number	**Choice Restriction**		
1	$T_\omega(z_c) = t_l$	$\Rightarrow$	$T_\omega(z_8) \neq t_h$
1	$T_\omega(z_c) = t_l$	$\Rightarrow$	$T_\omega(z_e) \neq t_h$
1	$T_\omega(z_c) = t_l$	$\Rightarrow$	$T_\omega(z_e) \neq t_m$
2	$T_\omega(z_c) = t_m$	$\Rightarrow$	$T_\omega(z_8) \neq t_h$
2	$T_\omega(z_c) = t_m$	$\Rightarrow$	$T_\omega(z_e) \neq t_h$
3	$T_\omega(z_e) = t_m$	$\Rightarrow$	$T_\omega(z_c) \neq t_h$
3	$T_\omega(z_e) = t_m$	$\Rightarrow$	$T_\omega(z_8) \neq t_h$
3	$T_\omega(z_e) = t_m$	$\Rightarrow$	$T_\omega(z_c) \neq t_l$
3	$T_\omega(z_e) = t_m$	$\Rightarrow$	$T_\omega(z_8) \neq t_l$
4	$T_\omega(z_e) = t_h$	$\Rightarrow$	$T_\omega(z_c) \neq t_m$
4	$T_\omega(z_e) = t_h$	$\Rightarrow$	$T_\omega(z_c) \neq t_l$
4	$T_\omega(z_e) = t_h$	$\Rightarrow$	$T_\omega(z_8) \neq t_l$
5	$T_\omega(z_8) = t_h$	$\Rightarrow$	$T_\omega(z_c) \neq t_m$
5	$T_\omega(z_8) = t_h$	$\Rightarrow$	$T_\omega(z_e) \neq t_m$
5	$T_\omega(z_8) = t_h$	$\Rightarrow$	$T_\omega(z_c) \neq t_l$
5	$T_\omega(z_8) = t_h$	$\Rightarrow$	$T_\omega(z_e) \neq t_l$
6	$T_\omega(z_8) = t_l$	$\Rightarrow$	$T_\omega(z_e) \neq t_h$
6	$T_\omega(z_8) = t_l$	$\Rightarrow$	$T_\omega(z_e) \neq t_m$
7	$T_\omega(z_e) = t_l$	$\Rightarrow$	$T_\omega(z_8) \neq t_h$
8	$T_\omega(z_c) = t_h$	$\Rightarrow$	$T_\omega(z_e) \neq t_m$

Choice restrictions generated by the WARP choice rule of Lemma **L-1**.

□

A.3 Proof of Theorem T-1:

Proof. The unordered monotonicity criteria (13) states that for any choice $t \in \text{supp}(T)$, and for any $z, z' \in \text{supp}(Z)$,

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

Each $\omega \in \Omega$ is associated with a single response type $\boldsymbol{s} \in \text{supp}(\boldsymbol{S})$ because $\boldsymbol{S}$ is a balancing score. Thus we can use the definition of binary matrices $\boldsymbol{B}_t[z, \boldsymbol{s}] \equiv \big(\mathbf{1}[T_\omega = t] | \boldsymbol{S} = \boldsymbol{s}, Z = z\big)$ to restate the monotonicity criteria as:

$$\boldsymbol{B}_t[z, \boldsymbol{s}] \geq \boldsymbol{B}_t[z', \boldsymbol{s}] \; \forall \; \boldsymbol{s} \in \text{supp}(\boldsymbol{S}) \text{ or } \boldsymbol{B}_t[z, \boldsymbol{s}] \leq \boldsymbol{B}_t[z', \boldsymbol{s}] \; \forall \; \boldsymbol{s} \in \text{supp}(\boldsymbol{S}) \text{ and } \forall \, t \in \text{supp}(T). \tag{91}$$

Now consider a change of instrumental values $z \to z'$, $\boldsymbol{B}_t[z, \boldsymbol{s}] = 1 \& \boldsymbol{B}_t[z', \boldsymbol{s}] = 0$. Equation (91) implies that there is no response-types $\boldsymbol{s}' \in \text{supp}(\boldsymbol{S})$ such that $\boldsymbol{B}_t[z, \boldsymbol{s}'] = 0 \& \boldsymbol{B}_t[z', \boldsymbol{s}'] = 1$. Equation (91) can be equivalently stated as:

$$\text{for any } t \in \text{supp}(T) \text{ and any } \boldsymbol{s}, \boldsymbol{s}' \in \text{supp}(\boldsymbol{S}), \text{ it cannot be the case that} \tag{92}$$

$$\boldsymbol{B}_t[z, \boldsymbol{s}] = 1 \; \& \; \boldsymbol{B}_t[z', \boldsymbol{s}] = 0 \text{ and } \boldsymbol{B}_t[z, \boldsymbol{s}] = 1 \; \& \; \boldsymbol{B}_t[z', \boldsymbol{s}] = 0. \tag{93}$$

Condition (92)–(93) is only binding for distinct and non-zero vectors $\boldsymbol{B}_t[\cdot, \boldsymbol{s}], \boldsymbol{B}_t[\cdot, \boldsymbol{s}']$ differ. Thus we can replace $\boldsymbol{B}_t$ by $\boldsymbol{C}_t$ because matrix $\boldsymbol{C}_t$ is simply the collection of non-zero distinct vectors in the matrix $\boldsymbol{B}_t$. Let $i \in \{1, ..., r_t\}$ and $j \in \{1, ..., c_t\}$ be the row and column indexes for a matrix $\boldsymbol{C}_t$. In this notation, unordered monotonicity can be equivalently stated as:

$$\text{for any } t \in \text{supp}(T) \text{ and any } i, i' \in \{1, ..., r_t\} \text{ and any } j, j' \in \{1, ..., c_t\}, \tag{94}$$

$$\text{it cannot be that } \boldsymbol{C}_t[i, j] = 1 \; \& \; \boldsymbol{C}_t[i', j] = 0 \text{ and } \boldsymbol{C}_t[i, j'] = 1 \; \& \; \boldsymbol{C}_t[i', j'] = 0. \tag{95}$$

Let $\boldsymbol{C}_t[\cdot, j]$ and $\boldsymbol{C}_t[\cdot, j']$ denote two columns of matrix $\boldsymbol{C}_t$. For condition (94)–eq:Cmono2 to be violated, two conditions must occur:

1. There must exist a row i such that $\big[\boldsymbol{C}_t[i, j], \boldsymbol{C}_t[i, j']$ that takes value $[1, 0]$;
2. There must exist another row i' such that $\big[\boldsymbol{C}_t[i', j], \boldsymbol{C}_t[i', j'] = [0, 1]$

The first condition occurs only if $\boldsymbol{C}_t[\cdot, j]' \cdot \big(\boldsymbol{\iota} - \boldsymbol{C}_t[\cdot, j']\big) > 0$. The second condition occurs only if $\big(\boldsymbol{\iota} - \boldsymbol{C}_t[\cdot, j]\big)' \cdot \boldsymbol{C}_t[\cdot, j'] > 0$. where $\boldsymbol{\iota}$ denotes a vector of elements ones that has the row-dimension of $\boldsymbol{C}_t$. Therefore, unordered monotonicity criteria can be equivalently stated as:

$$\text{for any } t \in \text{supp}(T) \text{ and any } i, i' \in \{1, ..., r_t\} \text{ and any } j, j' \in \{1, ..., c_t\}, \tag{96}$$

$$\Big(\boldsymbol{C}_t[\cdot, j]' \cdot \big(\boldsymbol{\iota} - \boldsymbol{C}_t[\cdot, j']\big)\Big) \cdot \Big(\big(\boldsymbol{\iota} - \boldsymbol{C}_t[\cdot, j]\big)' \cdot \boldsymbol{C}_t[\cdot, j']\Big) = 0. \tag{97}$$

Let the $\boldsymbol{K}_t$ be a square matrix of dimension $r_t \times r_t$ whose elements are given by:

$$\boldsymbol{K}_t[j, j'] = \Big(\boldsymbol{C}_t[\cdot, j]' \cdot \big(\boldsymbol{\iota} - \boldsymbol{C}_t[\cdot, j']\big)\Big) \cdot \Big(\big(\boldsymbol{\iota} - \boldsymbol{C}_t[\cdot, j]\big)' \cdot \boldsymbol{C}_t[\cdot, j']\Big).$$

Given that $\boldsymbol{K}_t[j, j'] \geq 0$, unordered monotonicity holds if and only if the following condition holds:

$$\sum_{t \in \text{supp}(T)} \sum_{j=1}^{r_t} \sum_{j'=1}^{r_t} \boldsymbol{K}_t[j, j'] = 0 \text{ which is equivalent to } \sum_{t \in \text{supp}(T)} \boldsymbol{\iota}' \boldsymbol{K}_t \boldsymbol{\iota} = 0.$$

The condition of the theorem is obtained by expressing $\boldsymbol{K}_t$ by the matrix multiplication (98).

$$\boldsymbol{K}_t = \Big(\boldsymbol{C}_t' \cdot \big(\boldsymbol{\iota}\boldsymbol{\iota}' - \boldsymbol{C}_t\big)\Big) \odot \Big(\big(\boldsymbol{\iota}\boldsymbol{\iota}' - \boldsymbol{C}_t\big)' \cdot \boldsymbol{C}_t\Big). \tag{98}$$

□

A.4 Proof of Property P-1:

Proof. Table A.3 describes the elimination process based on the nine relations of the monotonicity property **P-1**. These surviving response-types generate the MTO response matrix of Lemma **L-3**. □

A.5 Proof of Lemma L-4:

Proof. The proof consists of showing that the reverse of each the monotonicity relations in **P-1** is violated by at least one response-type in the response matrix of **L-3**. For convenience, the MTO response matrix is presented below:

$$\begin{array}{c} \\ \boldsymbol{R} = \end{array} \begin{array}{c} \begin{array}{ccccccc} \boldsymbol{s}_1 & \boldsymbol{s}_2 & \boldsymbol{s}_3 & \boldsymbol{s}_4 & \boldsymbol{s}_5 & \boldsymbol{s}_6 & \boldsymbol{s}_7 \end{array} \\ \begin{bmatrix} t_h & t_m & t_l & t_h & t_h & t_m & t_h \\ t_h & t_m & t_l & t_m & t_l & t_m & t_m \\ t_h & t_m & t_l & t_l & t_l & t_l & t_h \end{bmatrix} \end{array} \begin{array}{c} \\ T_\omega(z_c) \\ T_\omega(z_8) \\ T_\omega(z_e) \end{array}$$

The table below *reverses* each of the direction of each of the monotonicity inequalities listed in **P-1**. The last column of the table displays the response-type that do not comply with the change in each monotonicity relation. Each reversed monotonicity relation is associated with at least one response-type. This means that for each monotonicity relation in **P-1** we can find a response-type that would be ellminitated if the monotonicity relation were to be changed and thereby the MTO response matrix above could not be generated.

	Values of Z-pairs	Values of T	**Reverse** of Unordered Monotonicity Relations			Response-type Violations
Monotonicity Relation 1	(z_c, z_8)	t_h	$\mathbf{1}[T_\omega(z_c) = t_h]$	$\leq$	$\mathbf{1}[T_\omega(z_8) = t_h]$	$\boldsymbol{s}_4, \boldsymbol{s}_5, \boldsymbol{s}_7$
Monotonicity Relation 2	(z_8, z_e)	t_h	$\mathbf{1}[T_\omega(z_8) = t_h]$	$\geq$	$\mathbf{1}[T_\omega(z_e) = t_h]$	$\boldsymbol{s}_7$
Monotonicity Relation 3	(z_e, z_c)	t_h	$\mathbf{1}[T_\omega(z_e) = t_h]$	$\geq$	$\mathbf{1}[T_\omega(z_c) = t_h]$	$\boldsymbol{s}_4, \boldsymbol{s}_5$
Monotonicity Relation 4	(z_c, z_8)	t_m	$\mathbf{1}[T_\omega(z_c) = t_m]$	$\geq$	$\mathbf{1}[T_\omega(z_8) = t_m]$	$\boldsymbol{s}_4, \boldsymbol{s}_7$
Monotonicity Relation 5	(z_8, z_e)	t_m	$\mathbf{1}[T_\omega(z_8) = t_m]$	$\leq$	$\mathbf{1}[T_\omega(z_e) = t_m]$	$\boldsymbol{s}_4, \boldsymbol{s}_6, \boldsymbol{s}_7$
Monotonicity Relation 6	(z_e, z_c)	t_m	$\mathbf{1}[T_\omega(z_e) = t_m]$	$\geq$	$\mathbf{1}[T_\omega(z_c) = t_m]$	$\boldsymbol{s}_6$
Monotonicity Relation 7	(z_c, z_8)	t_l	$\mathbf{1}[T_\omega(z_c) = t_l]$	$\geq$	$\mathbf{1}[T_\omega(z_8) = t_l]$	$\boldsymbol{s}_5$
Monotonicity Relation 8	(z_8, z_e)	t_l	$\mathbf{1}[T_\omega(z_8) = t_l]$	$\geq$	$\mathbf{1}[T_\omega(z_e) = t_l]$	$\boldsymbol{s}_4, \boldsymbol{s}_6$
Monotonicity Relation 9	(z_e, z_c)	t_l	$\mathbf{1}[T_\omega(z_e) = t_l]$	$\leq$	$\mathbf{1}[T_\omega(z_c) = t_l]$	$\boldsymbol{s}_4, \boldsymbol{s}_5, \boldsymbol{s}_6$

□

Table A.3: Elimination of MTO Response-types Due to Monotonicity Relations Regarding Unordered Monotonicity

Panel A	**All 27 Possible Response-types**																										
Counterfactual Choices	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
$T_\omega(z_c)$	t_h	t_h	t_h	t_h	t_h	t_h	t_h	t_h	t_h	t_m	t_m	t_m	t_m	t_m	t_m	t_m	t_m	t_m	t_l	t_l	t_l	t_l	t_l	t_l	t_l	t_l	t_l
$T_\omega(z_8)$	t_h	t_h	t_h	t_m	t_m	t_m	t_l	t_l	t_l	t_h	t_h	t_h	t_m	t_m	t_m	t_l	t_l	t_l	t_h	t_h	t_h	t_m	t_m	t_m	t_l	t_l	t_l
$T_\omega(z_e)$	t_h	t_m	t_l	t_h	t_m	t_l	t_h	t_m	t_l	t_h	t_m	t_l	t_h	t_m	t_l	t_h	t_m	t_l	t_h	t_m	t_l	t_h	t_m	t_l	t_h	t_m	t_l
Panel B																											
Monotonicity 1	✓	✓	✓	✓	✓	✓	✓	✓	✓	✗	✗	✗	✓	✓	✓	✓	✓	✓	✗	✗	✗	✓	✓	✓	✓	✓	✓
Monotonicity 2	✓	✗	✗	✓	✓	✓	✓	✓	✓	✓	✗	✗	✓	✓	✓	✓	✓	✓	✓	✗	✗	✓	✓	✓	✓	✓	✓
Monotonicity 3	✓	✓	✓	✓	✓	✓	✓	✓	✓	✗	✓	✓	✗	✓	✓	✗	✓	✓	✗	✓	✓	✗	✓	✓	✗	✓	✓
Monotonicity 4	✓	✓	✓	✓	✓	✓	✓	✓	✓	✗	✗	✗	✓	✓	✓	✗	✗	✗	✓	✓	✓	✓	✓	✓	✓	✓	✓
Monotonicity 5	✓	✗	✓	✓	✓	✓	✓	✗	✓	✓	✗	✓	✓	✓	✓	✓	✗	✓	✓	✗	✓	✓	✓	✓	✓	✗	✓
Monotonicity 6	✓	✗	✓	✓	✗	✓	✓	✗	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✗	✓	✓	✗	✓	✓	✗	✓
Monotonicity 7	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✗	✗	✗	✗	✗	✗	✓	✓	✓
Monotonicity 8	✓	✓	✓	✓	✓	✓	✗	✗	✓	✓	✓	✓	✓	✓	✓	✗	✗	✓	✓	✓	✓	✓	✓	✓	✗	✗	✓
Monotonicity 9	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✗	✗	✓	✗	✗	✓	✗	✗	✓
Not Eliminated	1			4		6			9					14	15												27

Panel A lists all the 27 possible response-types that the response variable $S_\omega = [T_\omega(z_c)\,,\, T_\omega(z_8)\,,\, T_\omega(z_e)]$ can take. Rows present the counterfactual neighborhood choices that would arise if a family were assigned to control group, Section 8 and experimental group, that is $T_\omega(z_c), T_\omega(z_8)$ and $T_\omega(z_e)$ respectively. Columns present all the values of response-type as choices range over $\text{supp}(T) = \{t_h, t_m, t_l\}$. Panel B describes an elimination process based on the nine monotonicity relations of the monotonicity property **P-1**. Those monotonicity relations are listed below for convenience:

	Z-pairs	T	Monotonicity Relations		
Monotonicity Relation 1	(z_c, z_8)	t_h	$\mathbf{1}[T_\omega(z_c) = t_h]$	$\geq$	$\mathbf{1}[T_\omega(z_8) = t_h]$
Monotonicity Relation 2	(z_8, z_e)	t_h	$\mathbf{1}[T_\omega(z_8) = t_h]$	$\leq$	$\mathbf{1}[T_\omega(z_e) = t_h]$
Monotonicity Relation 3	(z_e, z_c)	t_h	$\mathbf{1}[T_\omega(z_e) = t_h]$	$\leq$	$\mathbf{1}[T_\omega(z_c) = t_h]$
Monotonicity Relation 4	(z_c, z_8)	t_m	$\mathbf{1}[T_\omega(z_c) = t_m]$	$\leq$	$\mathbf{1}[T_\omega(z_8) = t_m]$
Monotonicity Relation 5	(z_8, z_e)	t_m	$\mathbf{1}[T_\omega(z_8) = t_m]$	$\geq$	$\mathbf{1}[T_\omega(z_e) = t_m]$
Monotonicity Relation 6	(z_e, z_c)	t_m	$\mathbf{1}[T_\omega(z_e) = t_m]$	$\leq$	$\mathbf{1}[T_\omega(z_c) = t_m]$
Monotonicity Relation 7	(z_c, z_8)	t_l	$\mathbf{1}[T_\omega(z_c) = t_l]$	$\leq$	$\mathbf{1}[T_\omega(z_8) = t_l]$
Monotonicity Relation 8	(z_8, z_e)	t_l	$\mathbf{1}[T_\omega(z_8) = t_l]$	$\leq$	$\mathbf{1}[T_\omega(z_e) = t_l]$
Monotonicity Relation 9	(z_e, z_c)	t_l	$\mathbf{1}[T_\omega(z_e) = t_l]$	$\geq$	$\mathbf{1}[T_\omega(z_c) = t_l]$

Check mark ✓ indicates that the response-type displayed by the top column of the table does not violate the choice restriction denoted by the panel row. Cross sign ✗indicates that the response-type violates the choice restriction and should be eliminated. The last row in each panel presents the response-types that survive the elimination process. These surviving response-types generate the MTO response matrix of Lemma **L-3**.

A.6 Proof of Property P-2:

Proof. Suppose the property does not hold. Therefore there must exist $\boldsymbol{s}, \boldsymbol{s'} \in \text{supp}(\boldsymbol{S})$ and instrumental values $z, z' \in \text{supp}(Z)$ such that

$$\boldsymbol{s} \in \Sigma_t(z)\ ,\ \boldsymbol{s'} \in \Sigma_t(z') \text{ such that } \boldsymbol{s'} \notin \Sigma_t(z)\ ,\ \boldsymbol{s} \notin \Sigma_t(z'). \tag{99}$$

The definition of $\Sigma_t(z)$ is given by $\Sigma_t(z) \equiv \{\boldsymbol{s} \in \text{supp}(\boldsymbol{S}); \boldsymbol{B}_t[z, \boldsymbol{s}] = 1\}$ and $\boldsymbol{B}_t[z, \boldsymbol{s}] = 1$ implies that $\mathbf{1}[T(z) = t | \boldsymbol{S} = \boldsymbol{s}] = 1$. Therefore it must also be the case that

$$\mathbf{1}[T(z) = t | \boldsymbol{S} = \boldsymbol{s}] = \mathbf{1}[T_\omega(z') = t | \boldsymbol{S} = \boldsymbol{s'}] = 1 \tag{100}$$

$$\text{and also } \mathbf{1}[T(z) = t | \boldsymbol{S} = \boldsymbol{s'}] = \mathbf{1}[T_\omega(z') = t | \boldsymbol{S} = \boldsymbol{s}] = 0. \tag{101}$$

Now consider two agents $\omega, \omega' \in \Omega$ associated with each of the response-types, that is $\boldsymbol{S}_\omega = \boldsymbol{s}$ and $\boldsymbol{S}_{\omega'} = \boldsymbol{s}$. Equations (100)–(101) implies that:

$$T_\omega(z) = t\ ,\ T_{\omega'}(z') = t \text{ and also } T_{\omega'}(z) \neq t\ ,\ T_\omega(z') \neq t. \tag{102}$$

Reordering the terms in equation (102) we arise with the following strictly inequalities:

$$\mathbf{1}[T_\omega(z) = t] > \mathbf{1}[T_\omega(z') = t] \text{ and also } \mathbf{1}[T_{\omega'}(z) = t] < \mathbf{1}[T_{\omega'}(z') = t], \tag{103}$$

which violates unordered monotonicity (13). □

A.7 Proof of Theorem T-2:

For sake of notational simplicity, this proof focus on the identification of the expected value of counterfactual outcomes $E(Y(t)|\boldsymbol{S})$. Nevertheless, the proof also applies to any real-valued function $\kappa : \mathbb{R} \to \mathbb{R}$ of the outcome of interest Y, that is, $E(\kappa(Y(t))|\boldsymbol{S})$. For clarity, the proof is constructed based on the application of two Lemmas L-8–L-9 described below.

1. Lemma L-8 provides a general solution to systems of linear equations. The solution is well-known in the literature of linear algebra. The result is stated only for sake of completeness. Fo classical works on this topic, see Magnus and Neudecker (1999) and Barnett (1990).

2. Lemma L-9 exploits the decomposition of the indicator matrix $\boldsymbol{B}_t = \mathbf{1}[\boldsymbol{R} = t]$ into $\boldsymbol{B}_t = \boldsymbol{C}_t\boldsymbol{A}_t$. The lemma shows that the Moore-Penrose inverse of $\boldsymbol{B}_t$, namely $\boldsymbol{B}_t^+$, can be expressed as $\boldsymbol{B}_t^+ = \boldsymbol{A}_t'\big(\boldsymbol{A}_t\boldsymbol{A}_t'\big)^{-1}(\boldsymbol{C}_t'\boldsymbol{C}_t)^{-1}\boldsymbol{C}_t'$.

Lemma L-8. The general solution for $\boldsymbol{x}$ in the system of linear equations $\boldsymbol{b} = \boldsymbol{B}\boldsymbol{x}$ is given by:

$$\boldsymbol{b} = \boldsymbol{B}\boldsymbol{x} \quad \Rightarrow \quad \boldsymbol{x} = \boldsymbol{B}^+\boldsymbol{b} + (\boldsymbol{I} - \boldsymbol{B}^+\boldsymbol{B})\boldsymbol{\lambda} \text{ for any } \boldsymbol{\lambda} \in \mathbb{R}^{|\boldsymbol{b}|}. \tag{104}$$

The term $\boldsymbol{\lambda}$ in (104) denotes an arbitrary real-valued vector whose dimension is the same as the vector $\boldsymbol{b}$, $\boldsymbol{I}$ stands for the identity matrix and $\boldsymbol{B}^+$ is the Moore-Penrose Pseudoinverse of matrix $\boldsymbol{B}$.

Proof. 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 $\boldsymbol{b} = \boldsymbol{B}\boldsymbol{x}$, then it must be that:

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

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

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

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

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

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

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

□

Lemma L-9. Let $\boldsymbol{B}_t$ be the binary matrix associated with a response matrix $\boldsymbol{R}$ for which unordered monotonicity 13 holds. Then the Moore-Penrose pseudoinverse $\boldsymbol{B}_t^+$ is given by:

$$\boldsymbol{B}_t^+ = \boldsymbol{A}_t'\big(\boldsymbol{A}_t\boldsymbol{A}_t'\big)^{-1}(\boldsymbol{C}_t'\boldsymbol{C}_t)^{-1}\boldsymbol{C}_t'.$$

Proof. As mentioned, matrix $\boldsymbol{B}^+$ is defined by four properties: (1) $\boldsymbol{B}\boldsymbol{B}^+\boldsymbol{B} = \boldsymbol{B}$; (2) $\boldsymbol{B}^+\boldsymbol{B}\boldsymbol{B}^+ = \boldsymbol{B}^+$; (3) $\boldsymbol{B}^+\boldsymbol{B}$ is symmetric and (4) $\boldsymbol{B}\boldsymbol{B}^+$ is symmetric. Matrix $\boldsymbol{B}^+$ is unique and always exists for any real-valued matrix $\boldsymbol{B}$ (Magnus and Neudecker, 1999). Thus it suffices to show that $\boldsymbol{B}_t^+ = \boldsymbol{A}_t'\big(\boldsymbol{A}_t\boldsymbol{A}_t'\big)^{-1}(\boldsymbol{C}_t'\boldsymbol{C}_t)^{-1}\boldsymbol{C}_t'$ satisfies the properties (1) to (4) listed above.

$$\begin{aligned}
&(1)\ \boldsymbol{B}_t\cdot\boldsymbol{B}_t^+\cdot\boldsymbol{B}_t && = \boldsymbol{C}_t\boldsymbol{A}_t\cdot\boldsymbol{A}_t'\big(\boldsymbol{A}_t\boldsymbol{A}_t'\big)^{-1}(\boldsymbol{C}_t'\boldsymbol{C}_t)^{-1}\boldsymbol{C}_t'\cdot\boldsymbol{C}_t\boldsymbol{A}_t \\
& && = \boldsymbol{C}_t\boldsymbol{A}_t = \boldsymbol{B}_t \\
&(2)\ \boldsymbol{B}_t^+\cdot\boldsymbol{B}_t\cdot\boldsymbol{B}_t^+ && = \boldsymbol{A}_t'\big(\boldsymbol{A}_t\boldsymbol{A}_t'\big)^{-1}(\boldsymbol{C}_t'\boldsymbol{C}_t)^{-1}\boldsymbol{C}_t'\cdot\boldsymbol{C}_t\boldsymbol{A}_t\cdot\boldsymbol{A}_t'\big(\boldsymbol{A}_t\boldsymbol{A}_t'\big)^{-1}(\boldsymbol{C}_t'\boldsymbol{C}_t)^{-1}\boldsymbol{C}_t' \\
& && = \boldsymbol{A}_t'\big(\boldsymbol{A}_t\boldsymbol{A}_t'\big)^{-1}(\boldsymbol{C}_t'\boldsymbol{C}_t)^{-1}\boldsymbol{C}_t' = \boldsymbol{B}_t^+ \\
&(3)\ \boldsymbol{B}_t^+\cdot\boldsymbol{B}_t && = \boldsymbol{A}_t'\big(\boldsymbol{A}_t\boldsymbol{A}_t'\big)^{-1}(\boldsymbol{C}_t'\boldsymbol{C}_t)^{-1}\boldsymbol{C}_t'\cdot\boldsymbol{C}_t\boldsymbol{A}_t \\
& && = \boldsymbol{A}_t'\big(\boldsymbol{A}_t\boldsymbol{A}_t'\big)^{-1}\boldsymbol{A}_t \text{ which is symmetric} \\
&(4)\ \boldsymbol{B}_t\cdot\boldsymbol{B}_t^+ && = \boldsymbol{C}_t\boldsymbol{A}_t\cdot\boldsymbol{A}_t'\big(\boldsymbol{A}_t\boldsymbol{A}_t'\big)^{-1}(\boldsymbol{C}_t'\boldsymbol{C}_t)^{-1}\boldsymbol{C}_t \\
& && = \boldsymbol{C}_t(\boldsymbol{C}_t'\boldsymbol{C}_t)^{-1}\boldsymbol{C}_t \text{ which is symmetric.}
\end{aligned}$$

□

The proof of Theorem **T-2** is written below and it utilizes the two lemmas just stated.

Proof. Equation (28) in Section 7 connects observed data with unobserved causal parameters. The equation generates two system of linear equations for each $t \in \text{supp}(T)$ that govern the identification of counterfactual outcomes. Those are the equations (35)–(36) in Section 7.1 which are listed below for convenience:

$$\text{Outcome Equation: } \boldsymbol{Q}_Z(t) = \boldsymbol{B}_t\cdot\boldsymbol{Q}_S(t) \tag{105}$$

$$\text{Propensity Score Equation: } \boldsymbol{P}_Z(t) = \boldsymbol{B}_t\cdot\boldsymbol{P}_S \tag{106}$$

Recall that $\boldsymbol{Q}_S(t)$ stands for the vector $E(Y(t)|\boldsymbol{S}=\boldsymbol{S})\,P(\boldsymbol{S}=\boldsymbol{s})$ as $\boldsymbol{s}$ ranges in $\text{supp}(\boldsymbol{S})$ and $\boldsymbol{P}_S$ denotes the vector of response-type probabilities. Thus the identification of counterfactual outcomes hinges on the solution of the system linear equations (105)–(106).

According to Lemma (**L-8**), the solution to the system of linear equation in (105) is give by:

$$\boldsymbol{Q}_Z(t) = \boldsymbol{B}_t\cdot\boldsymbol{Q}_S(t) \Rightarrow \boldsymbol{\lambda}'\boldsymbol{Q}_S(t) = \boldsymbol{\lambda}\boldsymbol{B}_t^+\boldsymbol{Q}_Z(t) \text{ for any } \boldsymbol{\lambda} \text{ such that } \boldsymbol{\lambda}'(\boldsymbol{I}-\boldsymbol{B}_t\boldsymbol{B}_t^+) = \boldsymbol{0}. \tag{107}$$

Thus, we first prove that $\boldsymbol{A}_t(\boldsymbol{I} - \boldsymbol{B}^+\boldsymbol{B}) = \boldsymbol{0}$:

$$\boldsymbol{A}_t\boldsymbol{B}_t^+\boldsymbol{B}_t = \boldsymbol{A}_t\boldsymbol{A}_t'\big(\boldsymbol{A}_t\boldsymbol{A}_t'\big)^{-1}(\boldsymbol{C}_t'\boldsymbol{C}_t)^{-1}\boldsymbol{C}_t'\boldsymbol{B}_t \tag{108}$$

$$= (\boldsymbol{C}_t'\boldsymbol{C}_t)^{-1}\boldsymbol{C}_t'\boldsymbol{B} \tag{109}$$

$$= (\boldsymbol{C}_t'\boldsymbol{C}_t)^{-1}\boldsymbol{C}_t'\boldsymbol{C}_t\boldsymbol{A}_t \tag{110}$$

$$= \boldsymbol{A}_t \tag{111}$$

$$\therefore \boldsymbol{A}_t\boldsymbol{B}_t^+\boldsymbol{B}_t = \boldsymbol{A}_t \tag{112}$$

$$\Rightarrow \boldsymbol{A}_t(\boldsymbol{I} - \boldsymbol{B}^+\boldsymbol{B}) = \boldsymbol{0}, \tag{113}$$

where the first equality applies Lemma **L-9**, the second cancels matrix $(\boldsymbol{A}_t\boldsymbol{A}_t')$ with its inverse, the third uses decomposition $\boldsymbol{B}_t = \boldsymbol{C}_t\boldsymbol{A}_t$, and the fourth cancels matrix $(\boldsymbol{C}_t'\boldsymbol{C}_t)$ with its inverse.

Now if $\boldsymbol{A}_t(\boldsymbol{I} - \boldsymbol{B}^+\boldsymbol{B}) = \boldsymbol{0}$, then, by (107), we have that:

The solution to $\boldsymbol{A}_t\boldsymbol{Q}_Z(t) = \boldsymbol{A}_t\boldsymbol{B}_t \cdot \boldsymbol{Q}_S(t)$ is unique and given by $\boldsymbol{A}_t\boldsymbol{Q}_S(t) = \boldsymbol{A}_t\boldsymbol{B}_t^+\boldsymbol{Q}_Z(t)$. (114)

Therefore $\boldsymbol{A}_t\boldsymbol{Q}_S(t)$ is identified. Moreover, we have that:

$$\boldsymbol{A}_t\boldsymbol{B}_t^+ = \boldsymbol{A}_t\boldsymbol{A}_t'\big(\boldsymbol{A}_t\boldsymbol{A}_t'\big)^{-1}(\boldsymbol{C}_t'\boldsymbol{C}_t)^{-1}\boldsymbol{C}_t' = (\boldsymbol{C}_t'\boldsymbol{C}_t)^{-1}\boldsymbol{C}_t'. \tag{115}$$

Therefore the theorem is proved by the following equalities:

$$\boldsymbol{A}_t\boldsymbol{Q}_S(t) = \boldsymbol{A}_t\boldsymbol{B}_t^+\boldsymbol{Q}_Z(t) + \boldsymbol{A}_t(\boldsymbol{I} - \boldsymbol{B}^+\boldsymbol{B}), \tag{116}$$

$$= \boldsymbol{A}_t\boldsymbol{B}_t^+\boldsymbol{Q}_Z(t), \tag{117}$$

$$= (\boldsymbol{C}_t'\boldsymbol{C}_t)^{-1}\boldsymbol{C}_t'\boldsymbol{Q}_Z(t). \tag{118}$$

$$\tag{119}$$

The first equality applies the solution of linear system in Lemma 104 to equation (105), that is, $\boldsymbol{Q}_Z(t) = \boldsymbol{B}_t \cdot \boldsymbol{Q}_S(t)$. The second equality uses the fact that $\boldsymbol{A}_t(\boldsymbol{I} - \boldsymbol{B}^+\boldsymbol{B}) = \boldsymbol{0}$, in equation (112). The third equality utilizes the result stated in equation (115). As a conclusion, we have that the equality $\boldsymbol{A}_t\boldsymbol{Q}_S(t) = (\boldsymbol{C}_t'\boldsymbol{C}_t)^{-1}\boldsymbol{C}_t'\boldsymbol{Q}_Z(t)$ holds. By setting the outcome to one, we have that $\boldsymbol{A}_t\boldsymbol{P}_S(t) = (\boldsymbol{C}_t'\boldsymbol{C}_t)^{-1}\boldsymbol{C}_t'\boldsymbol{P}_Z(t)$ also holds. Indeed, the same steps that transform the equation $\boldsymbol{Q}_Z(t) = \boldsymbol{B}_t \cdot \boldsymbol{Q}_S(t)$ of (105) into equality in (118) can be applied to equation $\boldsymbol{P}_Z(t) = \boldsymbol{B}_t \cdot \boldsymbol{P}_S(t)$ in (106) to generate $\boldsymbol{A}_t\boldsymbol{P}_S(t) = (\boldsymbol{C}_t'\boldsymbol{C}_t)^{-1}\boldsymbol{C}_t'\boldsymbol{P}_Z(t)$. Note also that the rank of $\boldsymbol{A}_t$ is equal to the number of its rows in $\boldsymbol{A}_t$, which is also the same as the number of rows in $\boldsymbol{Q}_S(t)$. This means that we are exhausting the information content of $\boldsymbol{Q}_S(t)$ and any other causal parameter regarding the counterfactual outcome means $Y(t)$ that is identified must be a linear combination of the elements in $\boldsymbol{A}_t\boldsymbol{Q}_S(t)$.

□

A.8 Proof of Theorem T-3:

Proof. 1. Our goal is to identify $\boldsymbol{P}_S$, but according to Equation (36), that is, $\boldsymbol{P}_Z = \boldsymbol{B}_P \boldsymbol{P}_S$, where $\boldsymbol{B}_P$ is given by:

$$\boldsymbol{R} = \begin{bmatrix} 1 & 2 & 3 & 1 & 1 & 3 & 1 \\ 1 & 2 & 3 & 2 & 2 & 2 & 1 \\ 1 & 2 & 3 & 3 & 2 & 3 & 3 \end{bmatrix} \quad \Rightarrow \quad \boldsymbol{B}_P = \begin{bmatrix} 1 & 0 & 0 & 1 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 1 & 1 \end{bmatrix} \quad \therefore \quad \text{rank}(\boldsymbol{B}_P) = 7.$$

Thus, according to Lemma 104 , $\boldsymbol{P}_S$ is identified through $\boldsymbol{P}_S = \boldsymbol{B}_P^+ \boldsymbol{P}_Z$ as $\left(\boldsymbol{I}_9 - \boldsymbol{B}_P^+ \boldsymbol{B}_P\right) = \boldsymbol{0}$. The Moore-Penrose inverse of matrix $\boldsymbol{B}_P$ above is given by:

$$\boldsymbol{B}_P^+ = \frac{1}{9} \cdot \begin{bmatrix} 1 & 1 & 7 & 1 & 1 & -2 & 1 & 1 & -2 \\ -2 & 1 & 1 & 7 & 1 & 1 & -2 & 1 & 1 \\ 1 & -2 & 1 & 1 & -2 & 1 & 1 & 7 & 1 \\ 3 & -6 & 3 & 3 & 3 & -6 & -6 & 3 & 3 \\ 3 & 0 & -3 & -6 & -0 & 6 & 3 & -0 & -3 \\ -3 & 3 & -0 & -3 & 3 & 0 & 6 & -6 & -0 \\ -0 & 6 & -6 & 0 & -3 & 3 & 0 & -3 & 3 \end{bmatrix} \quad \text{and } \boldsymbol{P}_S = \boldsymbol{B}_P^+ \boldsymbol{P}_Z. \tag{120}$$

Remark 1.1. Note that $\boldsymbol{P}_Z$ has dimension 9×1 while $\boldsymbol{P}_S$ has dimension 7×1. It is useful to examine the system of linear equations defined by $\boldsymbol{P}_Z = \boldsymbol{B}_P \boldsymbol{P}_S$ as a linear transformation (linear map) $\mathcal{L} : \mathcal{V} \to \mathcal{W}$ where the transformation $\mathcal{L}$ is characterized by matrix $\boldsymbol{B}_P$ from the domain $\mathcal{V} \equiv \mathbb{R}^7$ to a linear subspace $\mathcal{W} \subset \mathbb{R}^9$. The solution to the system is unique, which means that the kernel of this transformation has dimension zero, that is $\ker(\mathcal{L}) = \{\boldsymbol{0}\}$. The ranknullity theorem states that the dimension of the kernel of a linear transformation plus the dimensions of its image is equal to the dimension of its domain, that is:

$$\dim(\ker(\mathcal{L})) + \dim(\text{im}(\mathcal{L})) = \dim(\mathcal{V}).$$

If we consider the dimension of the domain as 7, then the dimension of the image must be also 7. This means that if $\boldsymbol{P}_S$ were to take values in $\mathbb{R}^7$, then not all vectors in $\mathbb{R}^9$ would qualify as image $\boldsymbol{P}_Z$ of the system of linear equations $\boldsymbol{P}_Z = \boldsymbol{B}_P \boldsymbol{P}_S$. Indeed, if we sum the rows $1, 4, 7$ of $\boldsymbol{B}_P$ we would obtain a row of elements ones. We also obtain a row of elements one if we sum the rows 2,5,8 of $\boldsymbol{B}_P$. Finally, if we sum the rows $3, 6, 9$, of $\boldsymbol{B}_P$ we would also obtain a row of elements one. This fact implies that any vector $\boldsymbol{P}_Z$ in the image of $\mathcal{L}$ must be such that the sum of the elements in the rows $1, 4, 7$ must be equal to the sum of the rows $2, 5, 8$ which is also equal to the sum of the elements in the rows $3, 6, 9$. This criteria reduces the dimension of the image of the linear transformation $\mathcal{L}$ from 9 to 7, as expected. The vector of propensity scores $\boldsymbol{P}_Z$ complies with this criteria. Indeed, the sums are equal to one and stem from the fact that $P(T = t_h|Z = z) + P(T = t_m|Z = z) + P(T = t_h|Z = z) = 1$ for each $z \in \{z_c, z_8, z_e\}$.

2. The identified causal parameters is a direct consequence of Theorem T-2 applied to the matrix decomposition listed in equations (10)–(12).

3. According to equation (29) in Section 7, we have that:

$$E(X_\omega \cdot \mathbf{1}[T_\omega = t]|Z_\omega) = \sum_{s \in \text{supp}(S)} \mathbf{1}[T_\omega = t|S_\omega = s, Z_\omega]\, E(X_\omega \cdot \mathbf{1}[S_\omega = s]),$$

which is equivalent to the equation for propensity scores (30) if $\mathbf{1}[T_\omega = t]$ were replaced by

$X_\omega \cdot \mathbf{1}[T_\omega = t]$ and if $\mathbf{1}[S_\omega = s]$ were replaced by $X_\omega \cdot \mathbf{1}[S_\omega = s]$. Thereby we can express equation (29) in matrix notation by $\boldsymbol{Q}_Z = \boldsymbol{B}_P \boldsymbol{Q}_S$ (instead of $\boldsymbol{Q}_Z = \boldsymbol{B}_Q \boldsymbol{Q}_S$) when X_ω is the targeted variable of $\boldsymbol{Q}_Z$ and $\boldsymbol{Q}_S$. Thus, by the rationale of item (1), $E(X_\omega \cdot \mathbf{1}[S_\omega = s])$ is identified for all $s \in \text{supp}(S)$. The proof is completed by the fact that probabilities $P(S_\omega = s)$ are identified for all $s \in \text{supp}(S)$.

□

A.9 Proof of Lemma L-5:

Proof. The definition of the **TOT** parameters is given below:

$$\mathbf{TOT}(z_e, z_c) \equiv \frac{E(Y|Z=z_e) - E(Y|Z=z_c)}{P(T=t_l|Z=z_e)}, \tag{121}$$

$$\mathbf{TOT}(z_8, z_c) \equiv \frac{E(Y|Z=z_8) - E(Y|Z=z_c)}{P(T\in\{t_m,t_l\}|Z=z_8)}. \tag{122}$$

This proof focus on expressing the parameter $\mathbf{TOT}(z_e, z_c)$ in terms of neighborhood choices and response-types. We can use the law of iterated expectations to rewrite the expectation $E(Y|Z=z_e)$ as:

$$E(Y|Z=z_e) = \sum_{t\in\{t_h,t_m,t_l\}} E(Y|T=t, Z=z_e)\, P(T=t|Z=z_e), \tag{123}$$

$$\equiv \sum_{t\in\{t_h,t_m,t_l\}} E(Y\cdot\mathbf{1}[T=t]|Z=z_e), \tag{124}$$

$$= \sum_{t\in\{t_h,t_m,t_l\}} [0,0,1]\cdot\boldsymbol{Q}_Z(t), \tag{125}$$

where the last equality uses the matrix notation and the definition of $\boldsymbol{Q}_Z(t)$ in (31) of Section 7.1. On the same token, we can express the expectation $E(Y|Z=z_c)$ as:

$$E(Y|Z=z_c) = \sum_{t\in\{t_h,t_m,t_l\}} [1,0,0]\cdot\boldsymbol{Q}_Z(t). \tag{126}$$

We can use equations (126) and (125) to express the difference in expectations of the parameter $\mathbf{TOT}(z_e, z_c)$ as:

$$E(Y|Z=z_e) - E(Y|Z=z_c) = \sum_{t\in\{t_h,t_m,t_l\}} \big([0,0,1]-[1,0,0]\big)\cdot\boldsymbol{Q}_Z(t),$$

$$= \sum_{t\in\{t_h,t_m,t_l\}} [-1,0,1]\cdot\boldsymbol{Q}_Z(t).$$

Thereby:

$$\mathbf{TOT}(z_e, z_c) \equiv \frac{\sum_{t\in\{t_h,t_m,t_l\}}[-1,0,1]\cdot\boldsymbol{Q}_Z(t)}{P(T=t_l|Z=z_e)} \tag{127}$$

$$\equiv \left(\frac{\sum_{t\in\{t_h,t_m,t_l\}}[-1,0,1]\cdot\boldsymbol{Q}_Z(t)}{\sum_{t\in\{t_h,t_m,t_l\}}[-1,0,1]\cdot\boldsymbol{P}_Z(t)}\right)\cdot\left(\frac{\sum_{t\in\{t_h,t_m,t_l\}}[-1,0,1]\cdot\boldsymbol{P}_Z(t)}{P(T=t_l|Z=z_e)}\right). \tag{128}$$

However, equations (35)–(36) in Section 7 provide the following equalities:

$$\text{Outcome Equation: } \boldsymbol{Q}_Z(t) = \boldsymbol{B}_t\cdot\boldsymbol{Q}_S(t); t\in\{t_h,t_m,t_l\} \tag{129}$$

$$\text{Propensity Score Equation: } \boldsymbol{P}_Z(t) = \boldsymbol{B}_t\cdot\boldsymbol{P}_S; t\in\{t_h,t_m,t_l\} \tag{130}$$

According to equality (129), the summation $\sum_{t\in\{t_h,t_m,t_l\}}[-1,0,1]\cdot\boldsymbol{Q}_Z(t)$ in (128) is given by:

$$\sum_{t\in\{t_h,t_m,t_l\}}[-1,0,1]\cdot\boldsymbol{Q}_Z(t) \tag{131}$$
$$=\sum_{t\in\{t_h,t_m,t_l\}}[-1,0,1]\cdot\boldsymbol{B}_t\cdot\boldsymbol{Q}_S(t) \tag{132}$$
$$=[0,0,0,-1,-1,0,0]\boldsymbol{Q}_S(t_h)+[0,0,0,0,0,-1,0]\boldsymbol{Q}_S(t_m)+[0,0,0,1,1,1,0]\boldsymbol{Q}_S(t_l) \tag{133}$$
$$=E(Y(t_l)-Y(t_h)|\boldsymbol{S}=\boldsymbol{s}_4)P(\boldsymbol{S}=\boldsymbol{s}_4)+E(Y(t_l)-Y(t_h)|\boldsymbol{S}=\boldsymbol{s}_5)P(\boldsymbol{S}=\boldsymbol{s}_5)+ \tag{134}$$
$$+E(Y(t_l)-Y(t_l)|\boldsymbol{S}=\boldsymbol{s}_6)P(\boldsymbol{S}=\boldsymbol{s}_6) \tag{135}$$
$$=E(Y(t_l)-Y(t_h)|\boldsymbol{S}\in\{\boldsymbol{s}_4,\boldsymbol{s}_5\})P(\boldsymbol{S}\in\{\boldsymbol{s}_4,\boldsymbol{s}_5\})+E(Y(t_l)-Y(t_l)|\boldsymbol{S}=\boldsymbol{s}_6)P(\boldsymbol{S}=\boldsymbol{s}_6) \tag{136}$$

According to equality (130), the summation $\sum_{t\in\{t_h,t_m,t_l\}}[-1,0,1]\cdot\boldsymbol{P}_Z(t)$ in (128) is given by:

$$\sum_{t\in\{t_h,t_m,t_l\}}[-1,0,1]\cdot\boldsymbol{P}_Z(t) \tag{137}$$
$$=\sum_{t\in\{t_h,t_m,t_l\}}[-1,0,1]\cdot\boldsymbol{B}_t\cdot\boldsymbol{P}_S \tag{138}$$
$$=[0,0,0,-1,-1,0,0]\boldsymbol{P}_S+[0,0,0,0,0,-1,0]\boldsymbol{P}_S+[0,0,0,1,1,1,0]\boldsymbol{P}_S \tag{139}$$
$$=P(\boldsymbol{S}=\boldsymbol{s}_4)+P(\boldsymbol{S}=\boldsymbol{s}_5)+P(\boldsymbol{S}=\boldsymbol{s}_6) \tag{140}$$
$$=P(\boldsymbol{S}\in\{\boldsymbol{s}_4,\boldsymbol{s}_5,,\boldsymbol{s}_6\}) \tag{141}$$

Moreover, we can use equality (130) to express the probability $P(T=t_l|Z=z_e)$ as:

$$P(T=t_l|Z=z_e)=[0,0,1]\boldsymbol{B}_{t_l}\cdot\boldsymbol{P}_S \tag{142}$$
$$=P(\boldsymbol{S}=\boldsymbol{s}_3)+P(\boldsymbol{S}=\boldsymbol{s}_4)+P(\boldsymbol{S}=\boldsymbol{s}_5)+P(\boldsymbol{S}=\boldsymbol{s}_6) \tag{143}$$
$$=P(\boldsymbol{S}\in\{\boldsymbol{s}_3,\boldsymbol{s}_4,\boldsymbol{s}_5,\boldsymbol{s}_6\}) \tag{144}$$

We can use equations (136), (141) and (144) to express the parameter $\mathbf{TOT}(z_e,z_c)$ in equation (128) as:

$$\mathbf{TOT}(z_e,z_c)=\left(\frac{E(Y(t_l)-Y(t_h)|\boldsymbol{S}\in\{\boldsymbol{s}_4,\boldsymbol{s}_5\})P(\boldsymbol{S}\in\{\boldsymbol{s}_4,\boldsymbol{s}_5\})+E(Y(t_l)-Y(t_l)|\boldsymbol{S}=\boldsymbol{s}_6)P(\boldsymbol{S}=\boldsymbol{s}_6)}{P(\boldsymbol{S}\in\{\boldsymbol{s}_4,\boldsymbol{s}_5,,\boldsymbol{s}_6\})}\right)\cdot$$
$$\cdot\left(\frac{P(\boldsymbol{S}\in\{\boldsymbol{s}_4,\boldsymbol{s}_5,\boldsymbol{s}_6\})}{P(\boldsymbol{S}\in\{\boldsymbol{s}_3,\boldsymbol{s}_4,\boldsymbol{s}_5,,\boldsymbol{s}_6\})}\right) \tag{145}$$

The ratio of probabilities in (145) can be rewritten as $1-P(\boldsymbol{s}_3|\boldsymbol{S}\in\{\boldsymbol{s}_3,\boldsymbol{s}_4,\boldsymbol{s}_5,\boldsymbol{s}_6\})$ and therefore:

$$\mathbf{TOT}(z_e,z_c)=\left(\frac{E(Y(t_l)-Y(t_h)|\boldsymbol{S}\in\{\boldsymbol{s}_4,\boldsymbol{s}_5\})P(\boldsymbol{S}\in\{\boldsymbol{s}_4,\boldsymbol{s}_5\})+E(Y(t_l)-Y(t_l)|\boldsymbol{S}=\boldsymbol{s}_6)P(\boldsymbol{S}=\boldsymbol{s}_6)}{P(\boldsymbol{S}\in\{\boldsymbol{s}_4,\boldsymbol{s}_5,,\boldsymbol{s}_6\})}\right)\cdot$$
$$\cdot\left(1-P(\boldsymbol{s}_3|\boldsymbol{S}\in\{\boldsymbol{s}_3,\boldsymbol{s}_4,\boldsymbol{s}_5,\boldsymbol{s}_6\})\right).$$

The expression for $\mathbf{TOT}(z_8,z_c)$ is obtained by the same steps applied to $\mathbf{TOT}(z_e,z_c)$. □

A.10 Proof of Lemma L-6:

Proof. Let $\boldsymbol{B}_t[z,\cdot]$ denotes the row in the binary matrix $\boldsymbol{B}_t$ associated with the value z of the instrument Z. But $\Sigma_t(z) \equiv \{\boldsymbol{s} \in \text{supp}(\boldsymbol{S}); \boldsymbol{B}_t[z,\boldsymbol{s}] = 1\}$, thus the row $\boldsymbol{B}_t[z,\cdot]$ indicates the response-types that belong to set $\Sigma_t(z)$. The nested choices property **P-2** states that for any values $z, z' \in \text{supp}(Z)$ and any $t \in \text{supp}(T)$, we have that $\Sigma_t(z) \subset \Sigma_t(z')$ or $\Sigma_t(z') \subset \Sigma_t(z)$. Otherwise stated, **P-2** implies that for any values $z, z' \in \text{supp}(Z)$ and any $t \in \text{supp}(T)$, we have that:

$$\boldsymbol{B}_t[z,\cdot] \geq \boldsymbol{B}_t[z',\cdot] \text{ or } \boldsymbol{B}_t[z,\cdot] \leq \boldsymbol{B}_t[z',\cdot]. \tag{146}$$

The propensity score equality $\boldsymbol{P}_Z(t) = \boldsymbol{B}_t \cdot \boldsymbol{P}_S$ in (36), enables to express the $E(\mathbf{1}[T=t]|Z=z)$ as:

$$P(T=t|Z=z) \equiv E(\mathbf{1}[T=t]|Z=z) = \boldsymbol{B}_t[z,\cdot] \cdot \boldsymbol{P}_S. \tag{147}$$

Let $P(T=t|Z=z) > P(T=t|Z=z')$ for some values $z, z' \in \text{supp}(Z)$ and $t \in \text{supp}(T)$, Thus we have that:

$$P(T=t|Z=z) = \boldsymbol{B}_t[z,\cdot] \cdot \boldsymbol{P}_S > \boldsymbol{B}_t[z',\cdot] \cdot \boldsymbol{P}_S = P(T=t|Z=z'). \tag{148}$$

Moreover according to equation (146), it must be the case that $\boldsymbol{B}_t[z,\cdot] \geq \boldsymbol{B}_t[z',\cdot]$ In summary we have that:

$$P(T=t|Z=z) > P(T=t|Z=z') \Rightarrow \boldsymbol{B}_t[z,\cdot] \geq \boldsymbol{B}_t[z',\cdot]. \tag{149}$$

It must also be the case that $\boldsymbol{B}_t[z,\cdot] - \boldsymbol{B}_t[z',\cdot]$ is a row that has only elements 0 or 1. Indeed, the row $\boldsymbol{B}_t[z,\cdot] - \boldsymbol{B}_t[z',\cdot]$ simply indicates the response-types that belong to $\Sigma_t(z)$ and do not belong to $\Sigma_t(z')$. Let $\Delta_t(z,z')$ represent this set of response-types, that is:

$$\Delta_t(z,z') \equiv \{\boldsymbol{s}; \boldsymbol{B}_t[z,\boldsymbol{s}] - \boldsymbol{B}_t[z',\boldsymbol{s}] = 1\} \text{ for } t,z,z' \text{ such that } P(T=t|Z=z) > P(T=t|Z=z'). \tag{150}$$

Set $\Delta_t(z,z')$ can be also written as $\Delta_t(z,z') = \Sigma_t(z) \setminus \Sigma_t(z')$. According to (147), we can express the difference in expectation $E(\mathbf{1}[T=t]|Z=z) - E(\mathbf{1}[T=t]|Z=z')$ for $P(T=t|Z=z) > P(T=t|Z=z')$ can be expressed as:

$$\begin{aligned}
E(\mathbf{1}[T=t]|Z=z) - E(\mathbf{1}[T=t]|Z=z') &= \Big(\boldsymbol{B}_t[z,\cdot] - \boldsymbol{B}_t[z',\cdot]\Big) \cdot \boldsymbol{P}_S && (151)\\
&= \sum_{\boldsymbol{s} \in \sum S} \mathbf{1}[\boldsymbol{B}_t[z,\boldsymbol{s}] = 1]\mathbf{1}[\boldsymbol{B}_t[z',\boldsymbol{s}] = 0]P(\boldsymbol{S}=\boldsymbol{s}) && (152)\\
&= \sum_{\boldsymbol{s} \in \sum S} \mathbf{1}[\boldsymbol{B}_t[z,\boldsymbol{s}] - \boldsymbol{B}_t[z',\boldsymbol{s}] = 1]P(\boldsymbol{S}=\boldsymbol{s}) && (153)\\
&= P(\boldsymbol{s} \in \Delta_t(z,z')) && (154)\\
&= P(\boldsymbol{s} \in \Sigma_t(z) \setminus \Sigma_t(z')) && (155)\\
&= P(\boldsymbol{s} \in \Sigma_t(z) \oplus \Sigma_t(z')), && (156)
\end{aligned}$$

where $\oplus$ denotes symmetric difference.

The expected difference $E(Y \cdot \mathbf{1}[T=t]|Z=z) - E(Y \cdot \mathbf{1}[T=t_h]|Z=z')$ is examined following

the same utilized to examine the difference $E(\mathbf{1}[T=t]|Z=z) - E(\mathbf{1}[T=t_h]|Z=z')$. Namely:

$$E(Y\mathbf{1}[T=t]|Z=z) - E(Y\mathbf{1}[T=t]|Z=z') = \Big(\boldsymbol{B}_t[z,\cdot] - \boldsymbol{B}_t[z',\cdot]\Big) \cdot \boldsymbol{Q}_S(t) \tag{157}$$

$$= \sum_{\boldsymbol{s}\in\sum S} \big(\mathbf{1}[\boldsymbol{B}_t[z,\boldsymbol{s}]=1]\mathbf{1}[\boldsymbol{B}_t[z',\boldsymbol{s}]=0]\big) E(Y(t)|\boldsymbol{S}=\boldsymbol{s})P(\boldsymbol{S}=\boldsymbol{s}) \tag{158}$$

$$= \sum_{\boldsymbol{s}\in\sum S} \mathbf{1}[\boldsymbol{B}_t[z,\boldsymbol{s}] - \boldsymbol{B}_t[z',\boldsymbol{s}]=1] E(Y(t)|\boldsymbol{S}=\boldsymbol{s})P(\boldsymbol{S}=\boldsymbol{s}) \tag{159}$$

$$= \sum_{\boldsymbol{s}\in\Delta_t(z,z')} E(Y(t)|\boldsymbol{S}=\boldsymbol{s})P(\boldsymbol{S}=\boldsymbol{s}) \tag{160}$$

$$= E(Y(t)|\boldsymbol{s}\in\Delta_t(z,z')) \cdot P(\boldsymbol{S}\in\Delta_t(z,z')) \tag{161}$$

$$= E(Y(t)|\boldsymbol{s}\in\Sigma_t(z)\setminus\Sigma_t(z')) \cdot P(\boldsymbol{S}\in\Sigma_t(z)\setminus\Sigma_t(z')) \tag{162}$$

$$= E(Y(t)|\boldsymbol{s}\in\Sigma_t(z)\oplus\Sigma_t(z')) \cdot P(\boldsymbol{S}\in\Sigma_t(z)\oplus\Sigma_t(z')). \tag{163}$$

The ratio of the expected differences can be obtained by combining equations (156) and (163):

$$\frac{E(Y\mathbf{1}[T=t]|Z=z) - E(Y\mathbf{1}[T=t]|Z=z')}{E(\mathbf{1}[T=t]|Z=z) - E(\mathbf{1}[T=t]|Z=z')} = \frac{E(Y(t)|\boldsymbol{s}\in\Sigma_t(z)\oplus\Sigma_t(z')) \cdot P(\boldsymbol{S}\in\Sigma_t(z)\oplus\Sigma_t(z'))}{P(\boldsymbol{s}\in\Sigma_t(z)\oplus\Sigma_t(z'))} \tag{164}$$

$$= E(Y(t)|\boldsymbol{s}\in\Sigma_t(z)\oplus\Sigma_t(z')). \tag{165}$$

□

A.11 Proof of Theorem T-4:

The proof of Theorem **T-4** consists of combining Lemma **L-10** stated below with Lemma **L-6** of Section 8.1.

Lemma L-10. Consider a 2SLS regression where the instrumental variables consist of two binary indicators Z_1, Z_2, where the choice indicator D that plays the role of the endogenous variable and the second stage that uses Y as dependent variable such that $E(|Y|) < \infty$:

$$\text{First Stage } D_\omega = \gamma_1 \cdot Z_{\omega,1} + \gamma_2 \cdot Z_{\omega,2} + \epsilon_{\omega,D} \tag{166}$$

$$\text{Second Stage } Y_\omega = \kappa + \beta \cdot D_{t,\omega} + \epsilon_{\omega,Y}. \tag{167}$$

Let $\mathbb{Z}_1, \mathbb{Z}_2, \mathbb{D}, \mathbb{Y}$ denote the vectors of observed data associated with random variables Z_1, Z_2, D and Y respectively. If the binary variables $\mathbb{Z}_1, \mathbb{Z}_2$, are orthogonal, that is, $\mathbb{Z}_1'\mathbb{Z}_2 = 0$, then the 2SLS estimator for β converges in probability to the following ratio:

$$\beta \xrightarrow{p} \frac{E(Y|Z_1 = 1) - E(Y|Z_2 = 1)}{E(D|Z_1 = 1) - E(D|Z_2 = 1)} \tag{168}$$

Proof. Let N be the sample size. Let the $N \times 2$ matrix whose first column denotes a vector of element ones and the second vector is the observed data on the choice variable D be given by $\mathbb{X} = [\mathbf{1}, \mathbb{D}]$. Let $\mathbb{Z} = [\mathbb{Z}_1, \mathbb{Z}_2]$ denotes the data matrix of instrumental values whose dimension is $N \times 2$. Let the projection into the linear space generated by the columns in $\mathbb{Z}$ be:

$$\mathbb{P}_\mathbb{Z} = \mathbb{Z}\big(\mathbb{Z}'\mathbb{Z}\big)^{-1}\mathbb{Z}'.$$

In this notation, the parameters κ, β of equation (167) are estimated as:

$$\begin{pmatrix} \kappa \\ \beta \end{pmatrix} = \Big(\mathbb{X}'\mathbb{P}_\mathbb{Z}\mathbb{X}'\Big)^{-1}\mathbb{X}'\mathbb{Y}. \tag{169}$$

Let N_1, N_2 denote the sum of the binary vectors $\mathbb{Z}_1, \mathbb{Z}_2$ respectively. If $\mathbb{Z}_1'\mathbb{Z}_2 = 0$, then the expression in (169) is equal to:

$$\begin{pmatrix} \kappa \\ \beta \end{pmatrix} = \begin{pmatrix} N_1 + N_2 & \big(\mathbb{Z}_1 + \mathbb{Z}_2\big) \\ \mathbb{D}'\big(\mathbb{Z}_1 + \mathbb{Z}_2\big) & \frac{(\mathbb{D}'\mathbb{Z}_1)^2}{N_1} + \frac{(\mathbb{D}'\mathbb{Z}_2)^2}{N_2} \end{pmatrix}^{-1} \begin{pmatrix} \mathbb{Y}'\big(\mathbb{Z}_1 + \mathbb{Z}_2\big) \\ \frac{(\mathbb{D}'\mathbb{Z}_1)(\mathbb{Y}'\mathbb{Z}_1)}{N_1} + \frac{\big(\mathbb{D}'\mathbb{Z}_2\big)\cdot\big(\mathbb{Y}'\mathbb{Z}_2\big)}{N_2} \end{pmatrix} \tag{170}$$

After performing the matrix multiplications, the estimator β takes the following expression:

$$\beta = \frac{\left(\frac{\big(\mathbb{D}'\mathbb{Z}_1\big)\big(\mathbb{Y}'\mathbb{Z}_1\big)}{N_1} + \frac{\big(\mathbb{D}'\mathbb{Z}_2\big)\big(\mathbb{Y}'\mathbb{Z}_2\big)}{N_2}\right) - \left(\frac{\big(\mathbb{D}'(\mathbb{Z}_1+\mathbb{Z}_2)\big)\big(\mathbb{Y}'(\mathbb{Z}_1+\mathbb{Z}_2)\big)}{N_1+N_2}\right)}{\left(\frac{\big(\mathbb{D}'\mathbb{Z}_1\big)^2}{N_1} + \frac{\big(\mathbb{D}'\mathbb{Z}_2\big)^2}{N_2}\right) - \left(\frac{\big(\mathbb{D}'(\mathbb{Z}_1+\mathbb{Z}_2)\big)^2}{N_1+N_2}\right)} \tag{171}$$

After some algebraic manipulations, the expression in (171) can be rewritten as:

$$\beta = \frac{\left(\frac{\mathbb{Y}'\mathbb{Z}_1}{N_1} - \frac{\mathbb{Y}'\mathbb{Z}_2}{N_2}\right)\left(\frac{\mathbb{D}'\mathbb{Z}_1}{N_1}\frac{N_2}{N_1+N_2} - \frac{\mathbb{D}'\mathbb{Z}_2}{N_1}\frac{N_2}{N_1+N_2}\right)}{\left(\frac{\mathbb{D}'\mathbb{Z}_1}{N_1} - \frac{\mathbb{D}'\mathbb{Z}_2}{N_2}\right)\left(\frac{\mathbb{D}'\mathbb{Z}_1}{N_1}\frac{N_2}{N_1+N_2} - \frac{\mathbb{D}'\mathbb{Z}_2}{N_1}\frac{N_2}{N_1+N_2}\right)} \tag{172}$$

$$= \frac{\left(\frac{\mathbb{Y}'\mathbb{Z}_1}{N_1} - \frac{\mathbb{Y}'\mathbb{Z}_2}{N_2}\right)}{\left(\frac{\mathbb{D}'\mathbb{Z}_1}{N_1} - \frac{\mathbb{D}'\mathbb{Z}_2}{N_2}\right)} \tag{173}$$

$$\tag{174}$$

And by the law of large numbers we have that

$$\frac{\mathbb{Y}'\mathbb{Z}_1}{N_1} \xrightarrow{p} E(Y|Z_1 = 1) \tag{175}$$

$$\frac{\mathbb{Y}'\mathbb{Z}_2}{N_2} \xrightarrow{p} E(Y|Z_2 = 1) \tag{176}$$

$$\frac{\mathbb{D}'\mathbb{Z}_1}{N_1} \xrightarrow{p} E(D|Z_1 = 1) \tag{177}$$

$$\frac{\mathbb{D}'\mathbb{Z}_2}{N_2} \xrightarrow{p} E(D|Z_2 = 1) \tag{178}$$

$$\tag{179}$$

as desired. □

As mentioned, the proof of Theorem **T-4** consists of combining Lemma **L-10** just stated and Lemma **L-6** of Section 8.1

Proof. The 2SLS regression stated in Theorem **T-4** is represented by the one in Lemma **L-10** under a suitable transformation in variables. Indeed variables Z_1, Z_2, D, Y in Lemma **L-10** are to be replaced by $\mathbf{1}[Z = z], \mathbf{1}[Z = z'], \mathbf{1}[T = t], Y \cdot \mathbf{1}[T = t]$. Note that the IV indicators $\mathbf{1}[Z = z], \mathbf{1}[Z = z']$ are orthogonal, thus the result of Lemma **L-10** applies. Namely, the estimator β of the 2SLS in Theorem **T-4** converges in probability to the following ratio:

$$\beta \xrightarrow{p} \frac{E(Y \cdot \mathbf{1}[T = t]|Z = z) - E(Y \cdot \mathbf{1}[T = t]|Z = z')}{E(\mathbf{1}[T = t]|Z = z) - E(\mathbf{1}[T = t]|Z = z')}. \tag{180}$$

According to Lemma **L-6**, the expression in equation (180) identifies a counterfactual outcome mean given by: $E(Y(t)|\boldsymbol{S} \in \Sigma_t(z) \setminus \Sigma_t(z'))$. □

A.12 Proof of Lemma L-7:

The proof is based on two results on least squares regression for categorical independent variable take repetitive values. The results are presented in Lemmas **L-11**–**L-12**. Both lemmas are based on the notation described below.

Consider a linear regression $Y_\omega = X_\omega \boldsymbol{\beta} + \epsilon_\omega$. Let N be the sample size. The observed data is represented by:

1. Let $\mathbb{X}$ be the $N \times K$ matrix of observed independent variables.
2. Let $\mathbb{Y}$ be the $N \times 1$ vector of observed outcomes.
3. Let $\mathbb{W}$ be the $N \times N$ diagonal matrix whose diagonal elements are set to be the weight W_ω associated with element ω of the data.

Thus, the Least Squares (LS) estimator is given by:

$$\hat{\boldsymbol{\beta}}^{LS}_{K,1} = \left(\mathbb{X}'_{N,K}\mathbb{X}_{N,K}\right)^{-1}\mathbb{X}'_{N,K}\mathbb{Y}_{N,1}, \tag{181}$$

and the Weighted Least Squares (WLS) estimator is given by:

$$\hat{\boldsymbol{\beta}}^{WLS}_{K,1} = \left(\mathbb{X}'_{N,K}\mathbb{W}_{N,N}\mathbb{X}_{N,K}\right)^{-1}\mathbb{X}'_{N,K}\mathbb{W}_{N,N}\mathbb{Y}_{N,1}. \tag{182}$$

Now suppose that the rows of $\mathbb{X}$ take only $J > K$ distinct values $\mathbb{X}_1, ..., \mathbb{X}_J$. Let the set of the indexes ω associated with the observed data be $\Omega \equiv \{1, ..., N\}$ and let the data be partitioned according to the distinct rows of $\mathbb{X}$, that is,

$$\Omega = \cup_{j=1}^{J}\Omega_j \text{ such that } \Omega_j \equiv \{\omega \in \Omega; \boldsymbol{X}_\omega = \mathbb{X}_j\}. \tag{183}$$

Let $N_1, \ldots, N_J$ be the sample sizes associated with each partition set $\Omega_j; j \in \{1, ..., J\}$, associated with the possible rows $\mathbb{X}_1, ..., \mathbb{X}_J$ in $\mathbb{X}$. Thus we have that the total sample size is given by $N = N_1 + ... + N_J$. In this notation, consider the following matrices of observed data:

1. Let $\bar{\mathbb{X}}$ be the $J \times K$ matrix that stacks the distinct values $\mathbb{X}_1, ..., \mathbb{X}_J$ that the rows of $\mathbb{X}$ can take.
2. Let $\bar{\mathbb{Y}}$ be the $J \times 1$ matrix of the sample means of the observed outcome $\mathbb{Y}$ for each of the row values $\mathbb{X}_1, ..., \mathbb{X}_J$, that is:
$$\bar{\mathbb{Y}} = [\bar{\mathbb{Y}}_1, ..., \bar{\mathbb{Y}}_J]'; \; \bar{\mathbb{Y}}_j = \frac{\sum_{\omega \in \Omega_j} Y_\omega}{N_j}.$$
3. Let $\bar{\mathbb{W}}$ be the $J \times J$ diagonal matrix whose j-th diagonal element is given by $\bar{\mathbb{W}}_{j,j}$.

Our goal is to relate the least squares estimates that use the full data $\mathbb{X}_{N,K}, \mathbb{Y}_{N,1}$ with the linear regression that uses the partitioned data $\bar{\mathbb{X}}_{J,K}, \bar{\mathbb{Y}}_{J,1}$.

Lemma L-11. If we set the weights W_ω to the inverse of the sample size $1/N_j$ (or N/N_j) such that $X_\omega = \mathbb{X}_j$, we have that the weighted least square regression that uses the full data set is numerically identical to the linear regression that uses the partitioned data set. Specifically, if

$W_\omega = 1/N_j; X_\omega = \mathbb{X}_j$ we have that:

$$\hat{\boldsymbol{\beta}}^{WLS}_{K,1} = \underbrace{\left(\mathbb{X}'_{N,K}\mathbb{W}_{N,N}\mathbb{X}_{N,K}\right)^{-1}\mathbb{X}'_{N,K}\mathbb{W}_{N,N}\mathbb{Y}_{N,1}}_{\text{Full Data Set}} = \underbrace{\left(\bar{\mathbb{X}}'_{J,K}\bar{\mathbb{X}}_{J,K}\right)^{-1}\bar{\mathbb{X}}'_{J,K}\bar{\mathbb{Y}}_{J,1}}_{\text{Partitioned Data Set}}. \tag{184}$$

Proof. It suffices to show that $\mathbb{X}'\mathbb{W}\mathbb{X} = \bar{\mathbb{X}}'\bar{\mathbb{X}}$ and $\mathbb{X}'\mathbb{W}\mathbb{Y} = \bar{\mathbb{X}}'\bar{\mathbb{Y}}$.

$$\mathbb{X}'\mathbb{W}\mathbb{X} = \sum_{\omega\in\Omega}(\boldsymbol{X}'_\omega\boldsymbol{X}_\omega)\cdot W_\omega = \sum_{j=1}^{J}\sum_{\omega\in\Omega_j}\frac{\boldsymbol{X}'_\omega\boldsymbol{X}_\omega}{N_j} = \sum_{j=1}^{J}\frac{1}{N_j}\sum_{\omega\in\Omega_j}\boldsymbol{X}'_\omega\boldsymbol{X}_\omega = \sum_{j=1}^{J}\frac{1}{N_j}N_j\mathbb{X}'_j\mathbb{X}_j = \bar{\mathbb{X}}'\bar{\mathbb{X}}. \tag{185}$$

We also have that:

$$\mathbb{X}'\mathbb{W}\mathbb{Y} = \sum_{\omega\in\Omega}(\boldsymbol{X}'_\omega\boldsymbol{Y}_\omega)\cdot W_\omega = \sum_{j=1}^{J}\sum_{\omega\in\Omega_j}\frac{\boldsymbol{X}'_\omega\boldsymbol{Y}_\omega}{N_j} = \sum_{j=1}^{J}\mathbb{X}'_j\sum_{\omega\in\Omega_j}\frac{\boldsymbol{Y}_\omega}{N_j} = \sum_{j=1}^{J}\mathbb{X}'_j\bar{\mathbb{Y}}_j = \bar{\mathbb{X}}'\bar{\mathbb{Y}}. \tag{186}$$

□

On the other hand, we can relate the least squares estimator of the full data set to a weighted least squares estimator of the partitioned data set:

Lemma L-12. If the diagonal element of matrix Let $\bar{\mathbb{W}}$ is given by $\bar{\mathbb{W}}_{j,j} = N_j$ (or $\bar{\mathbb{W}}_{j,j} = N_j/N$) then the weighted least square regression that uses the partitioned data set is numerically identical to the linear regression that uses the full data set. Specifically:

$$\hat{\boldsymbol{\beta}}^{LS}_{K,1} = \underbrace{\left(\mathbb{X}'_{N,K}\mathbb{X}_{N,K}\right)^{-1}\mathbb{X}'_{N,K}\mathbb{Y}_{N,1}}_{\text{Full Data Set}} = \underbrace{\left(\bar{\mathbb{X}}'_{J,K}\bar{\mathbb{W}}_{J,J}\bar{\mathbb{X}}_{J,K}\right)^{-1}\bar{\mathbb{X}}'_{J,K}\bar{\mathbb{W}}_{J,J}\bar{\mathbb{Y}}_{J,1}}_{\text{Partitioned Data Set}}. \tag{187}$$

Proof. It suffices to show that $\mathbb{X}'\mathbb{X} = \bar{\mathbb{X}}'\bar{\mathbb{W}}\bar{\mathbb{X}}$ and $\mathbb{X}'\mathbb{Y} = \bar{\mathbb{X}}'\bar{\mathbb{W}}\bar{\mathbb{Y}}$.

$$\mathbb{X}'\mathbb{X} = \sum_{\omega\in\Omega}(\boldsymbol{X}'_\omega\boldsymbol{X}_\omega) = \sum_{j=1}^{J}N_j\mathbb{X}'_j\mathbb{X}_j = \sum_{j=1}^{J}\left(\mathbb{X}'_j\mathbb{X}_j\right)\bar{\mathbb{W}}_{j,j} = \bar{\mathbb{X}}'\bar{\mathbb{W}}\bar{\mathbb{X}}. \tag{188}$$

We also have that:

$$\mathbb{X}'\mathbb{Y} = \sum_{\omega\in\Omega}(\boldsymbol{X}'_\omega\boldsymbol{Y}_\omega) = \sum_{j=1}^{J}\sum_{\omega\in\Omega_j}\boldsymbol{X}'_\omega\boldsymbol{Y}_\omega = \sum_{j=1}^{J}N_j\cdot\mathbb{X}'_j\sum_{\omega\in\Omega_j}\frac{\boldsymbol{Y}_\omega}{N_j} = \sum_{j=1}^{J}\bar{\mathbb{W}}_{j,j}\left(\mathbb{X}'_j\bar{\mathbb{Y}}_j\right) = \bar{\mathbb{X}}'\bar{\mathbb{W}}\bar{\mathbb{Y}}. \tag{189}$$

□

We are now equipped to examine the estimator in Lemma **L-7**.

Proof. The estimator of the lemma stacks the observed data across the possible choice values. Some notation is necessary to investigate the estimator. Let N be the sample size and for notation simplicity, let the observed variables be indexed by $\omega \in \Omega \equiv \{1, ..., N\}$. Now consider the following notation:

1. Let $\mathbf{D}_{t,\omega} \equiv \mathbf{1}[T_\omega = t]$ be the indicator whether family ω chooses choice $t \in \{t_h, t_m, t_l\}$.

2. Let $\mathbb{D}_t$ be the N-dimensional vector of observed indicators $D_{t,\omega}$ across data indexed by $\omega \in \Omega$.

3. Let $\mathbb{D}$ be the $3N$-dimensional vector of observed data that stacks the observed indicators $\mathbb{D}_t$ across neighborhood choices t_h, t_m, t_l. Namely, $\mathbb{D} \equiv [\mathbb{D}'_{t_h},\ \mathbb{D}'_{t_m},\ \mathbb{D}'_{t_l}]'$.

In the same fashion,

1. Let $\boldsymbol{B}_{t,\omega} \equiv \boldsymbol{B}_t[Z_\omega, \cdot]$ denotes the row in the binary matrix $\boldsymbol{B}_t$ associated with the instrumental value Z_ω assigned to family ω.

2. Let the matrix $\mathbb{B}_t$ be the $N \times 7$ matrix that stacks $B_{t,\omega}$ across all families ω in the sample.

3. Let $\mathbb{B}$ be the $3N \times 7$ matrix that stacks matrices $\mathbb{B}_t$ across neighborhood choices t_h, t_m, t_l. Namely, $\mathbb{B} = [\mathbb{B}'_{t_h},\ \mathbb{B}'_{t_m},\ \mathbb{B}'_{t_l}]'$

The identification of the response-type probabilities $\boldsymbol{P}_S$ is governed by the Propensity Score Equation (36) of Section 7.1. The equation is displayed below:

$$\boldsymbol{P}_Z(t) = \boldsymbol{B}_t \cdot \boldsymbol{P}_S; t \in \{t_h, t_m, t_l\} \tag{190}$$

$$\Rightarrow \underbrace{\begin{bmatrix} \boldsymbol{P}_Z(t_h) \\ \boldsymbol{P}_Z(t_m) \\ \boldsymbol{P}_Z(t_l) \end{bmatrix}}_{\boldsymbol{P}_Z} = \underbrace{\begin{bmatrix} \boldsymbol{B}_{t_h} \\ \boldsymbol{B}_{t_m} \\ \boldsymbol{B}_{t_l} \end{bmatrix}}_{\boldsymbol{B}_P} \cdot \underbrace{\begin{bmatrix} P(\boldsymbol{S} = \boldsymbol{s}_1) \\ P(\boldsymbol{S} = \boldsymbol{s}_2) \\ P(\boldsymbol{S} = \boldsymbol{s}_3) \\ P(\boldsymbol{S} = \boldsymbol{s}_4) \\ P(\boldsymbol{S} = \boldsymbol{s}_5) \\ P(\boldsymbol{S} = \boldsymbol{s}_6) \\ P(\boldsymbol{S} = \boldsymbol{s}_7) \end{bmatrix}}_{\boldsymbol{P}_S} \tag{191}$$

$$\therefore \boldsymbol{P}_Z = \boldsymbol{B}_P \cdot \boldsymbol{P}_S. \tag{192}$$

Matrix $\boldsymbol{B}_P$ has full column rank, thus and its Moore-Penrose pseudo inverse is given by $\boldsymbol{B}_P^+ = \left(\boldsymbol{B}'_P \boldsymbol{B}_P\right)^{-1} \boldsymbol{B}'_P$. A solution to the system of linear equations in (192) is given by

$$\boldsymbol{P}_S = \left(\boldsymbol{B}'_P \boldsymbol{B}_P\right)^{-1} \boldsymbol{B}'_P \boldsymbol{P}_Z. \tag{193}$$

Note that $\boldsymbol{P}_Z$ has dimension 9×1 while $\boldsymbol{P}_S$ has dimension 7×1. Nevertheless, the solution to the system of linear equations $\boldsymbol{P}_Z = \boldsymbol{B}_P \cdot \boldsymbol{P}_S$ exists and is unique (see Remark *1.1* in Appendix A.8 for explanation).

Lemma **L-11** can be used to estimate the response-type probabilities as a weighted least squares regression. On the other hand, Lemma **L-12** can be used to estimate these response-type probabilities as a standard least squares regression. Most interesting, both approaches generate the same numerical estimates.

Consider the least square regression that uses $\mathbb{B}$ as the independent variable and $\mathbb{D}$ as the dependent variable. The values that the rows of each matrix $\mathbb{B}_t; t \in \{t_h, t_m, t_l\}$ take are the same given the voucher assignment $Z_\omega \in \{z_c, z_8, z_e\}$. Let $N_z = \sum_{\omega \in \Omega} \mathbf{1}[Z_\omega = z]$ be the sum of families assigned to voucher $z \in \{z_c, z_8, z_e\}$. Thus $\hat{P}_z = N_z/N$ be an estimate for $P(Z = z); z \in \{z_c, z_8, z_e$. Following Lemma **L-11**, consider a inverse probability weighting scheme that assigns the inverse

of voucher assignment probability to each agent ω of the sample. Specifically, let the weight associated with agent ω be $W_\omega = \hat{P}_z^{-1} = N/N_z$ such that $Z_\omega = z \in \{z_c, z_8, z_e\}$. Let $\mathbb{W}$ be the $N \times N$ diagonal matrix whose diagonal elements $\mathbb{W}[\omega,\omega] = \hat{P}_z^{-1}$ such that $Z_\omega = z \in \{z_c, z_8, z_e\}$. Under this weighting scheme, the weighted least squares regression of $\mathbb{D}$ on $\mathbb{B}$ is given by:

$$\widehat{\boldsymbol{P}}_S \;=\; \left(\mathbb{B}'\left(\mathbf{I}_{3,3}\otimes\mathbb{W}\right)\mathbb{B}\right)^{-1}\mathbb{B}'\left(\mathbf{I}_{3,3}\otimes\mathbb{W}\right)\mathbb{D}, \tag{194}$$

where $\mathbf{I}_{3,3}$ stands for the identity matrix of dimension 3×3, $\otimes$ is the kronecker product and $\widehat{\boldsymbol{P}}_S$ is a 7×1 vector of estimated response-type probabilities:

$$\widehat{\boldsymbol{P}}_S = [\hat{P}_{\boldsymbol{s}_1}, \hat{P}_{\boldsymbol{s}_2}, \hat{P}_{\boldsymbol{s}_3}, \hat{P}_{\boldsymbol{s}_4}, \hat{P}_{\boldsymbol{s}_5}, \hat{P}_{\boldsymbol{s}_6}, \hat{P}_{\boldsymbol{s}_7}]'.$$

According to equation (184) in Lemma **L-11**, this estimator is numerically equivalent to:

$$\widehat{\boldsymbol{P}}_S = \left(\boldsymbol{B}_P'\boldsymbol{B}_P\right)^{-1}\boldsymbol{B}_P'\widehat{\boldsymbol{P}}_Z, \tag{195}$$

$$\widehat{\boldsymbol{P}}_S = \left(\begin{bmatrix}\boldsymbol{B}_{t_h}\\ \boldsymbol{B}_{t_m}\\ \boldsymbol{B}_{t_l}\end{bmatrix}'\cdot\begin{bmatrix}\boldsymbol{B}_{t_h}\\ \boldsymbol{B}_{t_m}\\ \boldsymbol{B}_{t_l}\end{bmatrix}\right)^{-1}\begin{bmatrix}\boldsymbol{B}_{t_h}\\ \boldsymbol{B}_{t_m}\\ \boldsymbol{B}_{t_l}\end{bmatrix}'\widehat{\boldsymbol{P}}_Z, \tag{196}$$

where $\widehat{\boldsymbol{P}}_Z$ is the propensity score estimate generated by sample means, that is:

$$\widehat{\boldsymbol{P}}_Z = \left[\widehat{\boldsymbol{P}}_Z(t_h), \widehat{\boldsymbol{P}}_Z(t_m), \widehat{\boldsymbol{P}}_Z(t_l)\right]', \tag{197}$$

$$\text{such that } \widehat{\boldsymbol{P}}_Z(t) = \left[\widehat{\boldsymbol{P}}(T=t|Z=z_c), \widehat{\boldsymbol{P}}(T=t|Z=z_8), \widehat{\boldsymbol{P}}(T=t|Z=z_e)\right]', \tag{198}$$

$$\text{where } \widehat{\boldsymbol{P}}(T=t|Z=z) = \frac{\sum_{\omega\in\Omega}\mathbf{1}[T_\omega=t]\mathbf{1}[Z_\omega=z]}{\sum_{\omega\in\Omega}\mathbf{1}[Z_\omega=z]}. \tag{199}$$

By the large law of numbers, $\widehat{\boldsymbol{P}}_Z$ converges in probability to $\boldsymbol{P}_Z$ and thereby we have that:

$$\widehat{\boldsymbol{P}}_S \xrightarrow{p} \left(\boldsymbol{B}_P'\boldsymbol{B}_P\right)^{-1}\boldsymbol{B}_P\boldsymbol{P}_Z = \boldsymbol{B}_P^{+}\boldsymbol{P}_Z, \text{ as desired.} \tag{200}$$

We now consider the estimate of response-type probabilities using the standard least square estimator. Our goal is to show that the estimates obtained by the estimator displayed in equation (204) generate a result that is numerically equivalent to the estimator based on the weighted linear regression in (194). To do so, it is useful to state a property of the system of linear equations expressed by $\boldsymbol{P}_Z = \boldsymbol{B}_P\cdot\boldsymbol{P}_S$ in (192).

As mentioned, the solution to $\boldsymbol{P}_Z = \boldsymbol{B}_P\cdot\boldsymbol{P}_S$ exists, it is unique and is given by $\boldsymbol{P}_S = \left(\boldsymbol{B}_P'\boldsymbol{B}_P\right)^{-1}\boldsymbol{B}_P'\boldsymbol{P}_Z$ in (193) (see Appendix A.8). Now let $\boldsymbol{K}$ be a squared 9×9 invertible matrix, that is $\det(\boldsymbol{K})\neq 0$. Then it must be the case that:

$$\text{if } \boldsymbol{P}_S = \left(\boldsymbol{B}_P'\boldsymbol{B}_P\right)^{-1}\boldsymbol{B}_P'\boldsymbol{P}_Z \text{ solves } \boldsymbol{P}_Z = \boldsymbol{B}_P\boldsymbol{P}_S, \tag{201}$$

$$\text{then } \boldsymbol{P}_S = \left(\boldsymbol{B}_P'\boldsymbol{K}'\boldsymbol{K}\boldsymbol{B}_P\right)^{-1}\boldsymbol{B}_P'\boldsymbol{K}'\boldsymbol{K}\boldsymbol{P}_Z \text{ also solves } \left(\boldsymbol{K}\boldsymbol{P}_Z\right) = \left(\boldsymbol{K}\boldsymbol{B}_P\right)\cdot\boldsymbol{P}_S. \tag{202}$$

Moreover, suppose that $\boldsymbol{K}$ is a positive-definite symmetric 9×9 invertible matrix. Thus we have that $\boldsymbol{K}^{1/2}\boldsymbol{K}^{1/2} = \boldsymbol{K}$ and:

$$\boldsymbol{P}_S = \left(\boldsymbol{B}_P'\boldsymbol{K}\boldsymbol{B}_P\right)^{-1}\boldsymbol{B}_P'\boldsymbol{K}\boldsymbol{P}_Z \text{ solves } \left(\boldsymbol{K}^{1/2}\boldsymbol{P}_Z\right) = \left(\boldsymbol{K}^{1/2}\boldsymbol{B}_P\right)\cdot\boldsymbol{P}_S. \tag{203}$$

If the solution for (201) exists and is unique, then the solutions in (202) and (203) exist and are also unique. These results will be used next.

Consider the standard least square regression defined by the estimator displayed below:

$$\widetilde{\boldsymbol{P}}_S = \left(\mathbb{B}'\mathbb{B}\right)^{-1}\mathbb{B}'\mathbb{D}. \tag{204}$$

Let $\widetilde{\boldsymbol{W}}$ be a 9×9 matrix defined as:

$$\widetilde{\boldsymbol{W}} = \left(\mathbf{I}_{3,3} \otimes \widehat{\boldsymbol{W}}\right) \tag{205}$$

$$\text{where } \widehat{\boldsymbol{W}} = \begin{pmatrix} \hat{P}(Z = z_c) & 0 & 0 \\ 0 & \hat{P}(Z = z_8) & 0 \\ 0 & 0 & \hat{P}(Z = z_e) \end{pmatrix}, \text{ such that} \tag{206}$$

$$\hat{P}(Z = z) = \frac{\sum_{\omega \in \Omega} \mathbf{1}[Z_\omega = z_c]}{N} \text{ for } z \in \{z_c, z_8, z_e\}. \tag{207}$$

According to equation (187) in Lemma **L-12**, this estimator (204) is numerically equivalent to:

$$\widetilde{\boldsymbol{P}}_S = \left(\boldsymbol{B}_P'\widetilde{\boldsymbol{W}}\mathbb{B}\right)^{-1}\boldsymbol{B}_P'\widetilde{\boldsymbol{W}}\widehat{\boldsymbol{P}}_Z \tag{208}$$

$$\widetilde{\boldsymbol{P}}_S = \left(\boldsymbol{B}_P'\left(\mathbf{I}_{3,3} \otimes \widehat{\boldsymbol{W}}\right)\mathbb{B}\right)^{-1}\boldsymbol{B}_P'\left(\mathbf{I}_{3,3} \otimes \widehat{\boldsymbol{W}}\right)\widehat{\boldsymbol{P}}_Z \tag{209}$$

$$= \left(\begin{bmatrix} \boldsymbol{B}_{t_h} \\ \boldsymbol{B}_{t_m} \\ \boldsymbol{B}_{t_l} \end{bmatrix}' \begin{bmatrix} \widehat{\boldsymbol{W}} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \widehat{\boldsymbol{W}} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \widehat{\boldsymbol{W}} \end{bmatrix} \begin{bmatrix} \boldsymbol{B}_{t_h} \\ \boldsymbol{B}_{t_m} \\ \boldsymbol{B}_{t_l} \end{bmatrix}\right)^{-1} \cdot \begin{bmatrix} \boldsymbol{B}_{t_h} \\ \boldsymbol{B}_{t_m} \\ \boldsymbol{B}_{t_l} \end{bmatrix}' \begin{bmatrix} \widehat{\boldsymbol{W}} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \widehat{\boldsymbol{W}} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \widehat{\boldsymbol{W}} \end{bmatrix} \widehat{\boldsymbol{P}}_Z. \tag{210}$$

Observe that it is always the case that the matrix $\widetilde{\boldsymbol{W}}$ is an invertible symmetric positive-definite matrix. Thus, according to (203), we have that:

$$\widetilde{\boldsymbol{P}}_S = \left(\boldsymbol{B}_P'\widetilde{\boldsymbol{W}}\mathbb{B}\right)^{-1}\boldsymbol{B}_P'\widetilde{\boldsymbol{W}}\widehat{\boldsymbol{P}}_Z \text{ solves } \left(\widetilde{\boldsymbol{W}}^{1/2}\widehat{\boldsymbol{P}}_Z\right) = \left(\widetilde{\boldsymbol{W}}^{1/2}\boldsymbol{B}_P\right)\boldsymbol{P}_S. \tag{211}$$

Again, according to (203), $\widetilde{\boldsymbol{P}}_S$ is also the solution for:

$$\widetilde{\boldsymbol{P}}_S \text{is also a solution for } \widehat{\boldsymbol{P}}_Z = \boldsymbol{B}_P\boldsymbol{P}_S. \tag{212}$$

But this solution is unique and always exists, thereby we have that:

$$\widetilde{\boldsymbol{P}}_S = \widehat{\boldsymbol{P}}_S \text{ defined in (195)}. \tag{213}$$

Equation (213) implies that the weighted least squares estimators in (194) and the standard least square estimator in (204) generate the same numerical output. This only happens due to the way that matrices *mathbbB* and vector $\mathbb{D}$ are generated. In general, these estimators produce similar but not identical values.

Finally, by the large of large numbers we have that:

$$\widetilde{\boldsymbol{P}}_S \xrightarrow{p} \left(\boldsymbol{B}_P'\left(\mathbf{I}_{3,3} \otimes \mathbf{W}\right)\boldsymbol{B}_P\right)^{-1}\boldsymbol{B}_P\left(\mathbf{I}_{3,3} \otimes \mathbf{W}\right)\boldsymbol{P}_Z, \tag{214}$$

where matrix $\mathbf{W}$ denotes a 3×3 diagonal matrix given by:

$$\mathbf{W} = \begin{bmatrix} P(Z = z_c) & 0 & 0 \\ 0 & P(Z = z_8) & 0 \\ 0 & 0 & P(Z = z_e) \end{bmatrix} \tag{215}$$

We can evoke equations (201) and (203) once more to generate the equality:

$$\Big(\boldsymbol{B}'_P \left(\mathbf{I}_{3,3} \otimes \mathbf{W}\right) \boldsymbol{B}_P\Big)^{-1} \boldsymbol{B}_P \left(\mathbf{I}_{3,3} \otimes \mathbf{W}\right) \boldsymbol{P}_Z = \Big(\boldsymbol{B}'_P \boldsymbol{B}_P\Big)^{-1} \boldsymbol{B}_P \boldsymbol{P}_Z. \tag{216}$$

□

A.13 Estimation of Expected Value of Baseline Variables by Response-types:

This section focus on the estimation of the expected value of pre-program variables X conditioned on the response-types, that is, $E(X|\boldsymbol{S}=\boldsymbol{s}); \boldsymbol{s}\in\{\boldsymbol{s}_1,...,\boldsymbol{s}_7\}$. The procedure also holds for the estimation of $E(g(X)|\boldsymbol{S}=\boldsymbol{s})$ for any real-valued function $g:\mathbb{R}\to\mathbb{R}$.

According to equation (29) in Section 7, we have that:

$$E(X_\omega\cdot\mathbf{1}[T_\omega=t]|Z_\omega)=\sum_{s\in\text{supp}(S)}\mathbf{1}[T_\omega=t|S_\omega=s,Z_\omega]\,E(X_\omega\cdot\mathbf{1}[S_\omega=s]). \tag{217}$$

It is useful to express equation (217) in matrix form. To do so, let

$$\boldsymbol{Q}_Z^X(t)=\begin{bmatrix}E(X\cdot\mathbf{1}[T=t]|Z=z_c)\\E(X\cdot\mathbf{1}[T=t]|Z=z_8)\\E(X\cdot\mathbf{1}[T=t]|Z=z_e)\end{bmatrix}=\begin{bmatrix}E(X|T=t,Z=z_c)\,P(T=t|Z=z_c)\\E(X|T=t,Z=z_8)\,P(T=t|Z=z_8)\\E(X|T=t,Z=z_e)\,P(T=t|Z=z_e)\end{bmatrix}\text{ and }\boldsymbol{Q}_Z^X=\begin{bmatrix}\boldsymbol{Q}_Z(t_h)\\\boldsymbol{Q}_Z(t_m)\\\boldsymbol{Q}_Z(t_l)\end{bmatrix}. \tag{218}$$

Consider also the following notation:

$$\boldsymbol{Q}_S^X=\begin{bmatrix}E(X\cdot\mathbf{1}[\boldsymbol{S}=\boldsymbol{s}_1])\\\vdots\\E(X\cdot\mathbf{1}[\boldsymbol{S}=\boldsymbol{s}_7])\end{bmatrix}=\begin{bmatrix}E(X|\boldsymbol{S}=\boldsymbol{s}_1)\,P(\boldsymbol{S}=\boldsymbol{s}_1)\\\vdots\\E(X|\boldsymbol{S}=\boldsymbol{s}_7)\,P(\boldsymbol{S}=\boldsymbol{s}_7)\end{bmatrix}. \tag{219}$$

In this notation, equation (217) can be expressed in matrix form by:

$$\boldsymbol{Q}_Z^X(t)=\boldsymbol{B}_t\boldsymbol{Q}_S^X\text{ for each }t\in\{t_h,t_m,t_l\}. \tag{220}$$

$$\therefore\begin{bmatrix}\boldsymbol{Q}_Z^X(t_h)\\\boldsymbol{Q}_Z^X(t_m)\\\boldsymbol{Q}_Z^X(t_l)\end{bmatrix}=\begin{bmatrix}\boldsymbol{B}_{t_h}\\\boldsymbol{B}_{t_m}\\\boldsymbol{B}_{t_l}\end{bmatrix}\cdot\boldsymbol{Q}_S^X \tag{221}$$

$$\Rightarrow\boldsymbol{Q}_Z^X=\boldsymbol{B}_P\boldsymbol{Q}_S^X. \tag{222}$$

Equation (222) is symmetric to the propensity score equality $\boldsymbol{P}_Z(t)=\boldsymbol{B}_t\cdot\boldsymbol{P}_S$. Indeed, if we set X to 1, we would have the propensity score equation. Following Lemma **L-7**, let $D_{t,\omega}\equiv\mathbf{1}[T_\omega=t];t\in\{t_h,t_m,t_l\}$ is the binary indicator whether family ω chooses neighborhood t, and $\boldsymbol{B}_{t,\omega}\equiv\boldsymbol{B}_t[Z_\omega,\cdot]$ denotes the row of matrix $\boldsymbol{B}_t$ associated with the instrumental value $Z_\omega\in\{z_c,z_8,z_e\}$ assigned to family ω. Then Lemma **L-7** explains that the parameter $\boldsymbol{\beta}$ of the Least Squares Regression:

$$D_{t,\omega}=\boldsymbol{B}_{t,\omega}\boldsymbol{\beta}+\epsilon_\omega\text{ across all }t\in\text{supp}(T), \tag{223}$$

evaluates the response-type probabilities, that is,

$$\beta\xrightarrow{p}\boldsymbol{P}_S\equiv\begin{bmatrix}P(\boldsymbol{S}=\boldsymbol{s}_1)\\\vdots\\P(\boldsymbol{S}=\boldsymbol{s}_7)\end{bmatrix} \tag{224}$$

We can reframe the result using indicators. That is to say that the parameter $\boldsymbol{\beta}$ in the regres-

sion (225) below:

$$\mathbf{1}[T_\omega = t] = \boldsymbol{B}_{t,\omega}\boldsymbol{\beta} + \epsilon_\omega \text{ across all } t \in \text{supp}(T), \tag{225}$$

evaluates $\boldsymbol{P}_S$ based on the formula $\boldsymbol{P}_S = \left(\boldsymbol{B}_P'\boldsymbol{B}_P\right)^{-1}\boldsymbol{B}_P\boldsymbol{P}_Z$. In summary we have that:

$$\text{Regression } \mathbf{1}[T_\omega = t] = \boldsymbol{B}_{t,\omega}\boldsymbol{\beta} + \epsilon_\omega \text{ across all } t \in \text{supp}(T), \tag{226}$$

$$\Rightarrow \beta \xrightarrow{p} \boldsymbol{P}_S = \left(\boldsymbol{B}_P'\boldsymbol{B}_P\right)^{-1}\boldsymbol{B}_P \cdot \boldsymbol{P}_Z \tag{227}$$

$$\therefore \beta \xrightarrow{p} \begin{bmatrix} E(\mathbf{1}[\boldsymbol{S} = \boldsymbol{s}_1]) \\ \vdots \\ E(\mathbf{1}[\boldsymbol{S} = \boldsymbol{s}_7]) \end{bmatrix} = \left(\boldsymbol{B}_P'\boldsymbol{B}_P\right)^{-1}\boldsymbol{B}_P \cdot \begin{bmatrix} E(\mathbf{1}[T = t_h]|Z = z_c) \\ E(\mathbf{1}[T = t_h]|Z = z_m) \\ E(\mathbf{1}[T = t_h]|Z = z_l) \\ E(\mathbf{1}[T = t_m]|Z = z_c) \\ E(\mathbf{1}[T = t_m]|Z = z_m) \\ E(\mathbf{1}[T = t_m]|Z = z_l) \\ E(\mathbf{1}[T = t_l]|Z = z_c) \\ E(\mathbf{1}[T = t_l]|Z = z_m) \\ E(\mathbf{1}[T = t_l]|Z = z_l) \end{bmatrix}. \tag{228}$$

If $\mathbf{1}[T_\omega = t]$ in (226) were replaced by $X_\omega\mathbf{1}[T_\omega = t]$, then equation (228) would become:

$$\text{Regression } X_\omega\mathbf{1}[T_\omega = t] = \boldsymbol{B}_{t,\omega}\boldsymbol{\beta}_X + \epsilon_\omega \text{ across all } t \in \text{supp}(T), \tag{229}$$

$$\therefore \boldsymbol{\beta}_X \xrightarrow{p} \begin{bmatrix} E(X_\omega \cdot \mathbf{1}[\boldsymbol{S} = \boldsymbol{s}_1]) \\ \vdots \\ E(X_\omega \cdot \mathbf{1}[\boldsymbol{S} = \boldsymbol{s}_7]) \end{bmatrix} = \left(\boldsymbol{B}_P'\boldsymbol{B}_P\right)^{-1}\boldsymbol{B}_P \cdot \begin{bmatrix} E(X_\omega \cdot \mathbf{1}[T = t_h]|Z = z_c) \\ E(X_\omega \cdot \mathbf{1}[T = t_h]|Z = z_m) \\ E(X_\omega \cdot \mathbf{1}[T = t_h]|Z = z_l) \\ E(X_\omega \cdot \mathbf{1}[T = t_m]|Z = z_c) \\ E(X_\omega \cdot \mathbf{1}[T = t_m]|Z = z_m) \\ E(X_\omega \cdot \mathbf{1}[T = t_m]|Z = z_l) \\ E(X_\omega \cdot \mathbf{1}[T = t_l]|Z = z_c) \\ E(X_\omega \cdot \mathbf{1}[T = t_l]|Z = z_m) \\ E(X_\omega \cdot \mathbf{1}[T = t_l]|Z = z_l) \end{bmatrix}. \tag{230}$$

Otherwise stated, if If $\mathbf{1}[T_\omega = t]$ in (226) were replaced by $X_\omega\mathbf{1}[T_\omega = t]$, then the parameter $\boldsymbol{\beta}_X$ in the regression (231) would estimate:

$$\text{Regression } X_\omega\mathbf{1}[T_\omega = t] = \boldsymbol{B}_{t,\omega}\boldsymbol{\beta}_X + \epsilon_\omega \text{ across all } t \in \text{supp}(T), \tag{231}$$

$$\therefore \boldsymbol{\beta}_X \xrightarrow{p} \boldsymbol{Q}_S^X = \left(\boldsymbol{B}_P'\boldsymbol{B}_P\right)^{-1}\boldsymbol{B}_P \cdot \boldsymbol{Q}_Z^X. \tag{232}$$

Now recall that $\boldsymbol{Q}_S^X$ in (219) is the vector of the expectations $E(X|\boldsymbol{S} = \boldsymbol{s})P(\boldsymbol{S} = \boldsymbol{s})$, that is the expectation of X conditioned on $\boldsymbol{S}$ *multiplied* by the response-type probabilities $P(\boldsymbol{S} = \boldsymbol{s})$. We could divide each element in the estimate $\boldsymbol{\beta}_X$ of (231) by the respective response-type probability to obtain the estimates of $E(X|\boldsymbol{S} = \boldsymbol{s}_j); j = 1, ..., 7$. Another approach is to multiply each column of the 1×7 row $\boldsymbol{B}_{t,\omega}$ in regression (231) by the estimated response-type probabilities. This way

new estimated parameters would deliver the expectations $E(X|\boldsymbol{S}=\boldsymbol{s}_j); j=1,...,7$. Namely:

$$\text{Regression } X_\omega \mathbf{1}[T_\omega = t] = \left(\boldsymbol{B}_{t,\omega} \odot \hat{P}'_S\right)\boldsymbol{\beta}_X + \epsilon_\omega \text{ across all } t \in \text{supp}(T), \tag{233}$$

$$\therefore \boldsymbol{\beta}_X \xrightarrow{p} \left(\left(\boldsymbol{B}'_P \boldsymbol{B}_P\right)^{-1} \boldsymbol{B}_P \cdot \boldsymbol{Q}_Z^X)\right) \div \boldsymbol{P}_S = \begin{bmatrix} E(X_\omega|\boldsymbol{S}=\boldsymbol{s}_1) \\ \vdots \\ E(X_\omega|\boldsymbol{S}=\boldsymbol{s}_7) \end{bmatrix} \tag{234}$$

where $\div$ means element-wise division.

A.14 Proof of Theorem T-5:

Proof. Our goal is to show that the conditional expectation of counterfactual outcome $E(Y(t)|U_t = u)$ is identified. Our starting point is the expectation $E(Y \cdot \mathbf{1}[T=t]|P_t(Z)=p)$ in (235) which can be evaluated through observed data. Based on equation (62), we can rewrite the expectation as:

$$E\Big(Y \cdot \mathbf{1}[T=t]|P_t(Z)=p\Big) = E\Big(Y(t) \cdot \mathbf{1}[P(Z) \geq U_t]|P(Z)=p\Big), \tag{235}$$

$$= E\Big(Y(t) \cdot \mathbf{1}[p \geq U_t]\Big) = \int_0^p E(Y(t)|U_t=u)du. \tag{236}$$

Equality (235) applies the separability condition in (62) that stems form unordered monotonicity. Equality (236) is due to the independence relation $(U_t, Y(t)) \perp\!\!\!\perp P(Z)$ which is a consequence of the IV exclusion restriction as stated in Section 9.1. In summary, equations (235)–(236) show that this expectation is equivalent to integrating $E(Y(t)|U_t=u)$ over $[0,p]$. If U_t is absolutely continuous, then Lebesgue differentiation theorem holds. Thus the counterfactual expectation conditional on $U_t = p; p \in [0,1]$ can be evaluated by the partial derivative of $E(Y \cdot \mathbf{1}[T=t]|P(Z)=p)$ with respect to the propensity score $P(Z)$ at value p. This is stated in equation (237) below:

$$E\Big(Y \cdot \mathbf{1}[T=t]|P_t(Z)=p\Big) = \int_0^p E(Y(t)|U_t=u)du \Rightarrow \frac{\partial E(Y \cdot \mathbf{1}[T=t]|P(Z)=p)}{\partial p} = E(Y(t)|U_t=p). \tag{237}$$

Equations (238)–(240) below show that $E(Y \cdot \mathbf{1}[T=t]|P(Z)=p)$ can be equivalently expressed as $E(Y|T=t, P_t(Z)=p)p$:

$$E(Y \cdot \mathbf{1}[T=t]|P_T(Z)=p) = E(Y|T=t, P_t(Z)=p)E(\mathbf{1}[T=t]|P(Z)=p) \tag{238}$$

$$= E(Y|T=t, P_T(Z)=p)p \tag{239}$$

$$\therefore \frac{\partial E(Y|T=t, P(Z)=p)p}{\partial p} = E(Y(t)|U_t=p) \tag{240}$$

Equation (239) can be combined with equation (237) to state an equivalent identification approach in (240):

$$\therefore \frac{\partial E(Y|T=t, P(Z)=p)p}{\partial p} = E(Y(t)|U_t=p) \tag{241}$$

□

A.15 Proof of Theorem T-6:

The proof is based on standard properties of the least squares regression discussed below. Consider a linear regression $Y_\omega = X_\omega \boldsymbol{\beta} + \epsilon_\omega$. Let N be the sample size. The observed data is represented by:

1. Let $\mathbb{X}$ be the $N \times K$ matrix of observed independent variables.
2. Let $\mathbb{Y}$ be the $N \times 1$ vector of observed outcomes.

Thus, the Least Squares (LS) estimator is given by:

$$\hat{\boldsymbol{\beta}}^{LS}_{K,1} = \left(\mathbb{X}'_{N,K}\mathbb{X}_{N,K}\right)^{-1}\mathbb{X}'_{N,K}\mathbb{Y}_{N,1}. \tag{242}$$

Let $\widehat{\mathbb{Y}}$ be the $N \times 1$ vector of fitted outcomes, which is expressed by (243) and can be understood as the projection onto the space generated by the columns of $\mathbb{X}$.

$$= \mathbf{P}_{\mathbb{X}}\mathbb{Y}_{N,1}, \text{ such that } \mathbf{P}_{\mathbb{X}} \equiv \mathbb{X}_{N,K}\left(\mathbb{X}'_{N,K}\mathbb{X}_{N,K}\right)^{-1}\mathbb{X}'_{N,K}\mathbb{Y}_{N,1}. \tag{243}$$

A standard result in linear algebra is that any linearly independent combination of the columns in $\mathbb{X}$ generate the same projection of $\mathbf{P}_{\mathbb{X}}$. For instance, let $\mathbb{A}$ be a $K \times K$ invertible matrix and let $\mathbb{X}\mathbb{A}$ represent the linearly independent combination of the columns in $\mathbb{X}$, therefore we have that:

$$\mathbf{P}_{\mathbb{X}\mathbb{A}} \equiv \mathbb{X}\mathbb{A}\left(\mathbb{A}'\mathbb{X}'\mathbb{X}\mathbb{A}\right)^{-1}\mathbb{A}'\mathbb{X}' = \mathbb{X}\mathbb{A}\mathbb{A}^{-1}\left(\mathbb{X}'\mathbb{X}\right)^{-1}(\mathbb{A}')^{-1}\mathbb{A}'\mathbb{X}' = \mathbb{X}\left(\mathbb{X}'\mathbb{X}\right)^{-1}\mathbb{X}' = \mathbf{P}_{\mathbb{X}}. \tag{244}$$

Now suppose that the rows of $\mathbb{X}$ take only $J > K$ distinct values $\mathbb{X}_1, ..., \mathbb{X}_J$. Let the set of the indexes ω associated with the observed data be $\Omega \equiv \{1, ..., N\}$ and let the data be partitioned according to the distinct rows of $\mathbb{X}$, that is,

$$\Omega = \cup_{j=1}^{J}\Omega_j \text{ such that } \Omega_j \equiv \{\omega \in \Omega; \boldsymbol{X}_\omega = \mathbb{X}_j\}. \tag{245}$$

Let $N_1, \ldots, N_J$ be the sample sizes associated with each partition set $\Omega_j; j \in \{1, ..., J\}$, associated with the possible rows $\mathbb{X}_1, ..., \mathbb{X}_J$ in $\mathbb{X}$. Thus we have that the total sample size is given by $N = N_1 + ... + N_J$. In this notation, consider the following matrices of observed data:

1. Let $\bar{\mathbb{X}}$ be the $J \times K$ matrix that stacks the distinct values $\mathbb{X}_1, ..., \mathbb{X}_J$ that the rows of $\mathbb{X}$ can take.
2. Let $\mathbb{I}_j$ be the $N \times 1$ vector that indicates if X_ω takes values in $\mathbb{X}_j$, that is $\mathbb{I}_j \equiv [\mathbf{1}[X_\omega = \mathbb{X}_j; \omega \in \Omega]$.
3. Let $\mathbb{I}$ be the $N \times J$ matrix generated by the rectangular array of vectors $\mathbb{I}_j; j \in \{1, ..., J\}$, namely, $\mathbb{I} = [\mathbb{I}_1, ..., \mathbb{I}_J]$.
4. Let $\bar{\mathbb{Y}}$ be the $J \times 1$ matrix of the sample means of the observed outcome $\mathbb{Y}$ for each of the row values $\mathbb{X}_1, ..., \mathbb{X}_J$, that is:

$$\bar{\mathbb{Y}} = [\bar{\mathbb{Y}}_1, ..., \bar{\mathbb{Y}}_J]'; \; \bar{\mathbb{Y}}_j = \frac{\mathbb{Y}'\mathbb{I}_j}{\sum_{\omega\in\Omega}\mathbb{I}_j}.$$

Under this notation, we can state the following lemma:

Lemma L-13. If $\bar{\mathbb{X}}$ is an invertible matrix, then the projection that uses observed independent variables $\mathbb{X}$, that is $\mathbf{P}_{\mathbb{X}}$, is equal to the projection that uses the indicator matrix $\mathbb{I}$, that is $\mathbf{P}_{\mathbb{X}} = \mathbf{P}_{\mathbb{I}}$.

Proof. The proof is given by equation (246) that uses the fact that $\mathbb{X} = \mathbb{X}\bar{\mathbb{X}}$ and that $\bar{\mathbb{X}}$ is invertible.

$$\mathbf{P}_{\mathbb{X}} = \mathbf{P}_{\mathbb{I}\bar{\mathbb{X}}} = \mathbb{I}\bar{\mathbb{X}}\Big(\bar{\mathbb{X}}'\mathbb{I}'\mathbb{I}\bar{\mathbb{X}}\Big)^{-1}\bar{\mathbb{X}}'\mathbb{I}' = \mathbb{X}\mathbb{A}\mathbb{A}^{-1}\Big(\mathbb{I}'\mathbb{I}\Big)^{-1}(\bar{\mathbb{X}}')^{-1}\bar{\mathbb{X}}'\mathbb{I}'\mathbb{X}' = \mathbb{I}\Big(\mathbb{I}'\mathbb{I}\Big)^{-1}\mathbb{I}' = \mathbf{P}_{\mathbb{I}}. \tag{246}$$

□

The proof of the theorem is based on Lemmas **L-12**–**L-13** and is described below.

Proof. Let the sample size be N. For sake of notation simplicity, let the observed data be indexed by the set $\Omega = \{1, ..., N\}$ and the support of the instrumental variable Z be given by $\text{supp}(Z) = \{z_1, ..., z_{N_Z}\}$. Now consider the following notation for each $t \in \text{supp}(T)$:

1. Let $H_t(z_i) \equiv h_i; i = 1, ..., N_Z$ be the values that function $H_t(z)$ takes across the values $z \in \text{supp}(Z)$.

2. Let $\boldsymbol{\lambda}(h_i) = [\lambda_1(h_i), ..., \lambda_{N_Z}(h_i)]'; i = 1, ..., N_Z$ be the $N_Z \times 1$ vector that the vector-valued function $\boldsymbol{\lambda}(h_i)$ takes for each value $h_i; i = 1, ..., N_Z$.

3. Let $\boldsymbol{M} = [\boldsymbol{\lambda}_t(h_1), ..., \boldsymbol{\lambda}_t(h_{N_Z})]'$ be the $N_Z \times N_Z$ matrix generated by the rectangular array of vectors $\boldsymbol{\lambda}(h_i); i = 1, ..., N_Z$. The i-th row in $\boldsymbol{M}$ stands for the transpose of the vector $\boldsymbol{\lambda}(h_i)$.

4. Let $\mathbb{I}_i; i = 1, ..., N_Z$ be the $N \times 1$ vector that indicates if Z_ω takes values in $\text{supp}(Z)$ that is $\mathbb{I}_i \equiv [\mathbf{1}[Z_\omega = z_i; \omega \in \Omega]$.

5. Let $\mathbb{I}$ be the $N \times N_Z$ matrix generated by the rectangular array of vectors $\mathbb{I}_i; i \in \{1, ..., N_Z\}$, namely, $\mathbb{I} = [\mathbb{I}_1, ..., \mathbb{I}_{N_Z}]$.

6. Let $\mathbb{Y}$ be the $N \times 1$ vector of observed outcomes, that is $\mathbb{Y} = [Y_\omega; \omega \in \Omega]$.

7. Let $\mathbb{D}_t$ be the $N \times 1$ vector that indicates if the treatment choice across agents ω is equal to $t \in \text{supp}(T)$, that is $\mathbb{D}_t = [\mathbf{T}_\omega = \mathbf{t}; \omega \in \Omega]$.

8. Let $\widehat{\boldsymbol{P}}(t)$ denotes the $N_Z \times 1$ vector of estimated propensity score probabilities:

$$\widehat{\boldsymbol{P}}(t) = [\hat{P}_t(z_1), ..., \hat{P}_t(z_{N_Z})]', \text{ such that} \tag{247}$$

$$\hat{P}_t(z_{N_Z}) = \frac{\sum_{\omega=1}^{N} \mathbf{1}[T_\omega = t]\mathbf{1}[Z_\omega = z]}{\sum_{\omega=1}^{N} \mathbf{1}[Z_\omega = z]}. \tag{248}$$

9. Let $\overline{\mathbb{Y}\mathbb{D}}_t$ denotes the $N_Z \times 1$ vector of the sample estimates for $E(Y \cdot D_t | Z = z)$ across the values $z \in \text{supp}(Z)$:

$$\overline{\mathbb{Y}\mathbb{D}}_t = [\hat{Y}_t(z_1), ..., \hat{Y}_t(z_{N_Z})]', \text{ such that} \tag{249}$$

$$\hat{Y}_t(z) = \frac{\sum_{\omega=1}^{N} Y_\omega \cdot \mathbf{1}[T_\omega = t]\mathbf{1}[Z_\omega = z]}{\sum_{\omega=1}^{N} \mathbf{1}[Z_\omega = z]}. \tag{250}$$

10. Let $\bar{\mathbb{W}}$ be the $N_Z \times N_Z$ diagonal matrix whose i-th diagonal element is given by the sample estimates for the probabilities $P(Z = z_i); i =, ..., N_Z$, that is, $\bar{\mathbb{W}}[i, i] = \frac{\sum_{\omega=1}^{N} \mathbf{1}[Z_\omega = z]}{N}$.

The vector $\boldsymbol{\lambda}_{t,\omega}$ is defined by $\boldsymbol{\lambda}_t(H_t(Z_\omega))$ and stands for the vector $\boldsymbol{\lambda}_t(z)$ associated with the instrumental value $Z_\omega = z \in \text{supp}(Z)$. Thus the values that $\boldsymbol{\lambda}_{t,\omega}$ takes are repeated and only depend on the values of the instrumental variable.

To prove Theorem **T-6**, it suffices to focus on the behavior of the fitted values of two regressions:

$$Y_\omega D_{t,\omega} = \boldsymbol{\lambda}_{t,\omega}\boldsymbol{\beta}_t + \epsilon_{\omega,D}, \tag{251}$$

$$D_{t,\omega} = \boldsymbol{\lambda}_{t,\omega}\boldsymbol{\theta}_t + \epsilon_{\omega,D}. \tag{252}$$

The estimator of the linear regression (251) is given by:

$$\widehat{\boldsymbol{\beta}}_t = \Big((\mathbb{I}\boldsymbol{M})'(\mathbb{I}\boldsymbol{M})\Big)^{-1}(\mathbb{I}\boldsymbol{M})'(\mathbb{Y}\odot\mathbb{D}_t). \tag{253}$$

According to Lemma **L-13**, the fitted values of regression estimator (251) are given by:

$$\widehat{(\mathbb{Y}\odot\mathbb{D}_t)}_t = \mathbf{P}_{\mathbb{I}}(\mathbb{Y}\odot\mathbb{D}_t) = \mathbb{I}\Big(\mathbb{I}'\mathbb{I}\Big)^{-1}\mathbb{I}'(\mathbb{Y}\odot\mathbb{D}_t). \tag{254}$$

A consequence of equation (254) is that the fitted values of regression (251) would remain the same if its covariates were to be replaced by the indicators associated with the indicator matrix $\mathbb{I}$. In this case, the new linear regression is given by:

$$Y_\omega D_{t,\omega} = \sum_{i=1}^{N_Z}\mathbf{1}[Z_\omega = z_i]\gamma_{i,t} + \epsilon_{\omega,D}, \tag{255}$$

and the new estimator is given by:

$$\widetilde{\boldsymbol{\gamma}}_t = \Big((\mathbb{I})'(\mathbb{I})\Big)^{-1}(\mathbb{I})'(\mathbb{Y}\odot\mathbb{D}_t). \tag{256}$$

Now, according to Lemma **L-12** of Appendix A.12, this estimator can be equivalently written as:

$$\widetilde{\boldsymbol{\gamma}}_t = \Big((\boldsymbol{I}_{N_Z,N_Z})'\bar{\mathbb{W}}(\boldsymbol{I}_{N_Z,N_Z})\Big)^{-1}(\boldsymbol{I}_{N_Z,N_Z})'\bar{\mathbb{W}}(\overline{\mathbb{Y}\mathbb{D}}_t) \tag{257}$$

$$= \bar{\mathbb{W}}^{-1}\bar{\mathbb{W}}(\overline{\mathbb{Y}\mathbb{D}}_t) \tag{258}$$

$$= \overline{\mathbb{Y}\mathbb{D}}_t \tag{259}$$

$$\xrightarrow{p} [E(Y\cdot D_t|Z=z_1),...,E(Y\cdot D_t|Z=z_{N_Z})]', \tag{260}$$

where $\boldsymbol{I}_{N_Z,N_Z}$ stands for the N_Z-dimensional identity matrix and the second equality comes from the fact that $\bar{\mathbb{W}}$ is an $N_Z \times N_Z$ invertible matrix. We also have that the fitted values of the linear regression (255) are given by:

$$\sum_{i=1}^{N_Z}\mathbf{1}[Z_\omega = z_i]\gamma_{i,t} = \gamma_{i,t} = \overline{\mathbb{Y}\mathbb{D}}_t \xrightarrow{p} E(Y\cdot D_t|Z=z_i) \text{ for all } i = 1,...,N_Z. \tag{261}$$

Moreover, Lemma **L-12** of Appendix A.12 explains that the fitted values of the linear regression (255) are identical to the fitted values of the linear regression (251), therefore:

$$\boldsymbol{\lambda}(H_t(z_i))'\widehat{\boldsymbol{\beta}}_t = \gamma_{i,t} \xrightarrow{p} E(Y\cdot D_t|Z=z_i) \text{ for all } i = 1,...,N_Z. \tag{262}$$

If we outcome Y_ω were set to one, then we would have the linear regression (252). In this case, equation (262) above translates to:

$$\boldsymbol{\lambda}(H_t(z_i))'\widehat{\boldsymbol{\theta}}_t \xrightarrow{p} E(D_t|Z=z_i) = P(T=t|Z=z_i) \text{ for all } i = 1,...,N_Z. \tag{263}$$

Finally we can combine equations (262) and (263) to investigate the ration of expected differences

of Theorem **T-6**. Specifically:

$$\widehat{\Lambda}_t(z,z') = \frac{\Big(\boldsymbol{\lambda}(h_t) - \boldsymbol{\lambda}(h_t')\Big)'\widehat{\boldsymbol{\beta}}_t}{\Big(\boldsymbol{\lambda}(h_t) - \boldsymbol{\lambda}(h_t')\Big)'\widehat{\boldsymbol{\theta}}_t}, \tag{264}$$

$$= \frac{\Big(\boldsymbol{\lambda}(H_t(z)) - \boldsymbol{\lambda}(H_t(z'))\Big)'\widehat{\boldsymbol{\beta}}_t}{\Big(\boldsymbol{\lambda}(H_t(z')) - \boldsymbol{\lambda}(H_t(z'))\Big)'\widehat{\boldsymbol{\theta}}_t}, \tag{265}$$

$$= \frac{\Big(\boldsymbol{\lambda}(H_t(z))'\widehat{\boldsymbol{\beta}}_t - \boldsymbol{\lambda}(H_t(z'))'\widehat{\boldsymbol{\beta}}_t\Big)}{\Big(\boldsymbol{\lambda}(H_t(z))'\widehat{\boldsymbol{\theta}}_t - \boldsymbol{\lambda}(H_t(z'))'\widehat{\boldsymbol{\theta}}_t\Big)}, \tag{266}$$

$$\xrightarrow{p} \frac{E(Y \cdot D_t | Z = z) - E(Y \cdot D_t | Z = z')}{P(T = t | Z = z) - P(T = t | Z = z')} \tag{267}$$

$$= E(Y(t) | \boldsymbol{S} \in \Sigma_t(z) \oplus \Sigma_t(z')), \tag{268}$$

where convergency (267) is due to (262)-(263) the last equality (268) comes from Lemma **L-6**. □

B Binary Choice Model with Binary Instrumental Variable

The ITT effect consist of the causal effect of being offered a voucher. Kling et al. (2005) explain that the TOT is a Bloom (1984) estimator that evaluates the causal effect of being offered a voucher for the families that relocate using the voucher, that is, the voucher compliers. Both voucher effects, ITT and TOT are well-suited to evaluate the consequences of the housing policies that offer rent subsidizing vouchers. The binary choice model with binary instrumental variable is useful to illustrate these two effects.

1. Instrumental variable $Z_\omega \in \{0,1\}$ denotes a voucher assignment for family ω such that $Z_\omega = 1$ if family ω is a voucher recipient and $Z_\omega = 0$ if family ω receives no voucher.

2. The relocation decision T_ω for family ω such that $T_\omega = 0$ if family ω does not relocate and $T_\omega = 1$ if family relocates.

3. Counterfactual relocation decision $T_\omega(z)$ stands for the relocation decision that family ω would choose if it had been assigned to voucher $z \in \{0,1\}$.

4. Counterfactual outcomes $(Y_\omega(0), Y_\omega(1))$ denote the potential outcomes when relocation choice T_ω is *fixed* at values 0 and 1.

5. The observed outcome for family ω is given by $Y_\omega = Y_\omega(0)(1 - T_\omega) + Y_\omega(1)T_\omega$.

6. The response-type variable $\boldsymbol{S}_\omega$ that is defined by the unobserved vector of potential relocation decisions that a family ω would choose if voucher assignment were set to zero and one, i.e., $\boldsymbol{S}_\omega = [T_\omega(0), T_\omega(1)]'$.

Table A.4 describes the four vectors of potential response-types that $\boldsymbol{S}_\omega$ can take. The model is completed by the standard assumption that the instrumental variable Z_ω is independent of counterfactual variables:

$$(Y_\omega(0), Y_\omega(1), T_\omega(0), T_\omega(1)) \perp\!\!\!\perp Z_\omega. \tag{269}$$

The following equation comes as a direct consequence of Equation (269) and the definition of $\boldsymbol{S}_\omega$:

$$(Y_\omega(0), Y_\omega(1)) \perp\!\!\!\perp Z_\omega | \boldsymbol{S}_\omega. \tag{270}$$

In this notation, the relocation decision T_ω can be expressed in terms of response-type $\boldsymbol{S}_\omega$ as:

$$\begin{aligned} T_\omega &= (1 - Z_\omega)T_\omega(0) + Z_\omega T_\omega(0) && (271)\\ &= [\mathbf{1}(Z_\omega = 0), \mathbf{1}(Z_\omega = 1)] \cdot [T_\omega(0), T_\omega(1)]' && (272)\\ &= [\mathbf{1}(Z_\omega = 0), \mathbf{1}(Z_\omega = 1)] \cdot \boldsymbol{S}_\omega, && (273) \end{aligned}$$

where Equation (271) comes from the definition of $T_\omega(z); z \in \{0,1\}$, and Equation (273) comes from the definition of $\boldsymbol{S}_\omega$. A consequence of Equation (273) is that T_ω is deterministic conditioned on Z_ω and $\boldsymbol{S}_\omega$.

Table A.4: Possible Response-types for the Binary Relocation Choice with Binary Voucher

Voucher Types	Voucher Assignment	Relocation Countefactuals	**Response-types** Never Takers	Compliers	Always Takers	Defiers
No Voucher	$Z_\omega = 0$	$T_\omega(0)$	0	0	1	1
Voucher Recipient	$Z_\omega = 1$	$T_\omega(1)$	0	1	1	0

The expected value of observed outcomes conditioned on voucher assignment in this model is given by:

$$\begin{aligned} E(Y_\omega|Z_\omega = 1) &= E(Y_\omega|Z_\omega = 1, \boldsymbol{S}_\omega = [0,0]')\, P(\boldsymbol{S}_\omega = [0,0]') + E(Y_\omega|Z_\omega = 1, \boldsymbol{S}_\omega = [0,1]')\, P(\boldsymbol{S}_\omega = [0,1]') \\ &\quad + E(Y_\omega|Z_\omega = 1, \boldsymbol{S}_\omega = [1,1]')\, P(\boldsymbol{S}_\omega = [1,1]') + E(Y_\omega|Z_\omega = 1, \boldsymbol{S}_\omega = [1,0]')\, P(\boldsymbol{S}_\omega = [1,0]') \qquad (274) \\ &= E(Y_\omega(0)|\boldsymbol{S}_\omega = [0,0]')\, P(\boldsymbol{S}_\omega = [0,0]') + E(Y_\omega(1)|\boldsymbol{S}_\omega = [0,1]')\, P(\boldsymbol{S}_\omega = [0,1]') \\ &\quad + E(Y_\omega(1)|\boldsymbol{S}_\omega = [1,1]')\, P(\boldsymbol{S}_\omega = [1,1]') + E(Y_\omega(0)|\boldsymbol{S}_\omega = [1,0]')\, P(\boldsymbol{S}_\omega = [1,0]'), \qquad (275) \end{aligned}$$

where Equation (274) comes from the law of iterated expectations. Equation (275) comes the equation for observed outcome $Y_\omega = Y_\omega(0)(1 - T_\omega) + Y_\omega(1)T_\omega$, the fact that T_{omega} is deterministic conditioned on $\boldsymbol{S}_\omega$ and Z_ω and the independence relation $(Y_\omega(0), Y_\omega(1)) \perp\!\!\!\perp Z_\omega|\boldsymbol{S}_\omega$ of Equations 270. In the same fashion, we can express $E(Y_\omega|Z_\omega = 1)$ by:

$$\begin{aligned} E(Y_\omega|Z_\omega = 0) &= E(Y_\omega(0)|\boldsymbol{S}_\omega = [0,0]')\, P(\boldsymbol{S}_\omega = [0,0]') + E(Y_\omega(0)|\boldsymbol{S}_\omega = [0,1]')\, P(\boldsymbol{S}_\omega = [0,1]') \\ &\quad + E(Y_\omega(1)|\boldsymbol{S}_\omega = [1,1]')\, P(\boldsymbol{S}_\omega = [1,1]') + E(Y_\omega(1)|\boldsymbol{S}_\omega = [1,0]')\, P(\boldsymbol{S}_\omega = [1,0]'). \qquad (276) \end{aligned}$$

The Intention-to-treat effect ITT is defined by $E(Y_\omega|Z_\omega = 1) - E(Y_\omega|Z_\omega = 0)$ and refers to the causal effect of the vouchers Z_ω on outcome Y_ω. According to Equations (275)–(276), the ITT can be expressed in terms of response-types as:

$$\begin{aligned} ITT &= E(Y_\omega|Z_\omega = 1) - E(Y_\omega|Z_\omega = 0) \\ &= E(Y_\omega(1) - Y_\omega(0)|\boldsymbol{S}_\omega = [0,1]')\, P(\boldsymbol{S}_\omega = [0,1]') + E(Y_\omega(0) - Y_\omega(1)|\boldsymbol{S}_\omega = [1,0]')\, P(\boldsymbol{S}_\omega = [1,0]'). \qquad (277) \end{aligned}$$

Equation (277) states that the ITT is a mixture between the contradicting effects. By contradicting I mean the causal effect relocating compared to not relocation for the compliers ($\boldsymbol{S}_\omega = [0,1]'$) and the causal effect of not relocatong compared to relocating for the definers.

The probability of relocation conditioned on receiving the voucher is expressed in terms of response-types by:

$$\begin{aligned} P(T_\omega|Z_\omega = 1) &= E(\mathbf{1}[T_\omega = 1]|Z_\omega = 1, \boldsymbol{S}_\omega = [0,0]')\, P(\boldsymbol{S}_\omega = [0,0]') + E(\mathbf{1}[T_\omega = 1]|Z_\omega = 1, \boldsymbol{S}_\omega = [0,1]')\, P(\boldsymbol{S}_\omega = [0,1]') \\ &\quad + E(\mathbf{1}[T_\omega = 1]|Z_\omega = 1, \boldsymbol{S}_\omega = [1,1]')\, P(\boldsymbol{S}_\omega = [1,1]') + E(\mathbf{1}[T_\omega = 1]|Z_\omega = 1, \boldsymbol{S}_\omega = [1,0]')\, P(\boldsymbol{S}_\omega = [1,0]') \\ &= P(\boldsymbol{S}_\omega = [0,1]') + P(\boldsymbol{S}_\omega = [1,1]'), \qquad (278) \end{aligned}$$

where Equation (278) comes from the fact that T_ω is deterministic conditioned on $\boldsymbol{S}_\omega$ and Z_ω. Using the same reasoning, the probability of relocation conditioned on not receiving the voucher is

expressed in terms of response-types by:

$$P(T_\omega|Z_\omega = 0) = P(\boldsymbol{S}_\omega = [1,0]') + P(\boldsymbol{S}_\omega = [1,1]'). \tag{279}$$

Thus the difference in propensity of relocation across voucher assignments is given by:

$$P(T_\omega|Z_\omega = 1) - P(T_\omega|Z_\omega = 0) = P(\boldsymbol{S}_\omega = [0,1]') - P(\boldsymbol{S}_\omega = [1,0]'); \tag{280}$$

C Additional Information on the MTO Intervention

This section presents the statistical description of selected variables of the MTO intervention regarding neighborhood choice, poverty levels and voucher compliance. A baseline survey was conducted at the onset of the intervention, after which families were re-contacted in 1997 and 2000. As mentioned in

As mentioned in Section 3, MTO families were randomly allocated into three groups: 28% to the Section 8 group (z_8), 41% to the experimental group (z_e) and 31% to control (z_c).

Section 8 families were offered a rent-subsidizing voucher that could be used if a family agreed to relocate from the original housing projects to eligible private-market dwellings. The vouchers consisted of a tenant-based subsidy in which the rent of an eligible dwelling was paid directly to the landlord. All families were required to pay 30% of the household's monthly adjusted gross income for rent and utilities. The subsidy amounts were calculated based on the Applicable Payment Standard (APS) criteria set by the Housing and Urban Development (HUD). The rental subsidy differed depending on the number of bedrooms in the dwelling and the family's size. The eligible units comprised all of the houses and apartments available for rent that complied with the APS criteria. The landlords of an eligible dwelling could not discriminate against a voucher recipient who met the same requirements as a renter without a voucher. The lease was renewed automatically unless the owner (or the voucher recipient) stated otherwise in a written notice.

Experimental families were offered a voucher similar to the Section 8 voucher but could be used only in low-poverty neighborhoods. – those whose fraction of poor households was below 10% according to the 1990 US Census. Families that used the voucher were required to live for a year in low-poverty neighborhoods. After this period, families could use the voucher as a regular Section 8 voucher without geographical constraints. Less than two percent of the families that move using the vouchers returned to their original neighborhood. Experimental families also received some counseling from local nonprofit organizations to search for houses. Control families were offered no voucher.

Neighborhood choices are defined in accordance to the MTO design. This enables to clearly determine the incentives generated by each voucher and to clearly exploit the exogenous variation of voucher assignments. As mentioned, families decide among three neighborhood options: (1) high-poverty t_h; (2) medium-poverty t_m; or (3) low-poverty t_l. High-poverty neighborhoods consist of the high-poverty housing projects targeted at the intervention onset. The choice of high-poverty neighborhood is equivalent to not relocating. Low-poverty neighborhoods comprise the ones targeted by the experimental voucher, those whose fraction of poor residents is below 10% according to the 1990 US Census. Medium-poverty neighborhoods are defined by exclusion, the ones other than the housing projects targeted by MTO and whose fraction of poor residents is above 10% according to the 1990 US Census.

A sizeable share of families did not use the voucher to relocate. MTO noncompliance was substantial, but not unusual. Table A.5 classifies voucher recipients into three categories: (1)

compliers – families that used the vouchers; (2) self-movers – families that had moved without the voucher at the time of the impact interim evaluation in 2002; (3) stayers – families that had not moved since intervention onset until the interim evaluation in 2002. Half of the experimental families and 40% of the Section 8 families did not use the voucher to relocate. Self-movers totals 36% of experimental families and 24% of Section 8 families. Around 20% of families that receive vouchers and 30% of control families had not move from the targeted housing projects by the time of the interim evaluation.

Table A.5: Sample Sizes and Compliance Rates by Site

Voucher Assignment	All Sites N	All Sites %	Relocation Decision	All Sites N	All Sites %	Baltimore N	Baltimore %	Boston N	Boston %	Chicago N	Chicago %	Los Angeles N	Los Angeles %	New York N	New York %
Experimental	1729	41%	Compliers	818	47%	146	58%	168	46%	155	34%	167	67%	182	45%
			Self-movers	618	36%	97	38%	149	41%	234	51%	53	21%	85	21%
			Stayers	293	17%	9	4%	49	13%	71	15%	30	12%	134	33%
Section 8	1209	28%	Compliers	716	59%	135	72%	129	48%	134	66%	130	77%	188	49%
			Self-movers	276	23%	45	24%	86	32%	55	27%	25	15%	65	17%
			Stayers	217	18%	7	4%	52	19%	13	6%	13	8%	132	34%
Control	1310	31%	Self-movers	917	70%	174	88%	240	74%	189	81%	172	66%	142	48%
			Stayers	393	30%	23	12%	86	26%	43	19%	88	34%	153	52%
Total	4248														

This tables describe the sample sizes of families by voucher assignment and voucher compliance. MTO final sample consists of 4,248 families, 41% of those were assigned to the experimental group, 28% to Section 8 and 31% to control. Voucher recipients are classified into three groups: (1) compliers – families that used the vouchers to relocate; (2) self-movers – families that had moved without the voucher at the time of the impact interim evaluation in 2002 (four to seven years after enrollment); (3) stayers – families that had not moved since intervention onset in 1994–1998 until the interim evaluation in 2002.

An impact interim evaluation conducted in 2002 (four to seven years after enrollment) assessed six study domains: (1) mobility, housing, and neighborhood; (2) physical and mental health; (3) child educational achievement; (4) youth delinquency; (5) employment and earnings; and (6) household income and public assistance.[38] The MTO Long Term Evaluation consists of data collected between 2008 and 2010.

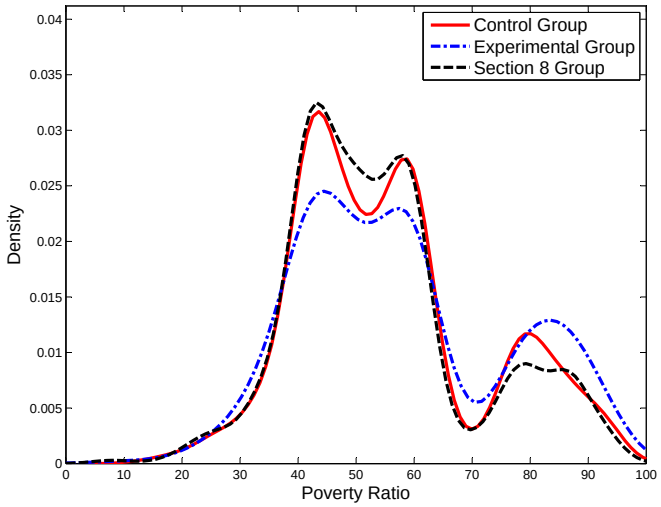
C.1 Distribution of Neighborhood Poverty By Voucher Assignment and Voucher Compliance

This section described the distribution of neighborhood poverty of MTO participating families by voucher assignment and relocation decision. Figure 11 shows the probability density estimation of baseline neighborhood poverty by voucher assignment. As expected, poverty distributions conditional on voucher assignments are very similar due to the randomized assignment of vouchers.

Figure 12 presents baseline neighborhood poverty for the Experimental group by neighborhood relocation, i.e., moving with voucher, moving without voucher and not moving. Families that did

[38] See Gennetian et al. (2012); Orr et al. (2003) for detailed descriptions of the intervention and the available data.

Figure 11: **Density Estimation of Baseline Neighborhood Poverty (1990 Census) by Voucher Assignment**



This figure presents the density estimation of baseline neighborhood poverty levels by voucher assignment, i.e., Control, Experimental and Section 8 groups. Poverty levels are computed according to the US 1990 Census data as the fraction of households whose income falls below the national poverty threshold for each 1990 census tract. Estimates are based on the normal kernel with optimal normal bandwidth. See columns 2–6 of Table 6 for inference on the average level of neighborhood poverty by voucher assignment.

not move had lived in slightly lower poverty level neighborhoods when compared to families that moved. Figure 12 also shows the poverty density of the neighborhood chosen by families that relocated using the Experimental voucher. The poverty levels of relocation neighborhoods are substantially lower than those of baseline neighborhoods as expected.

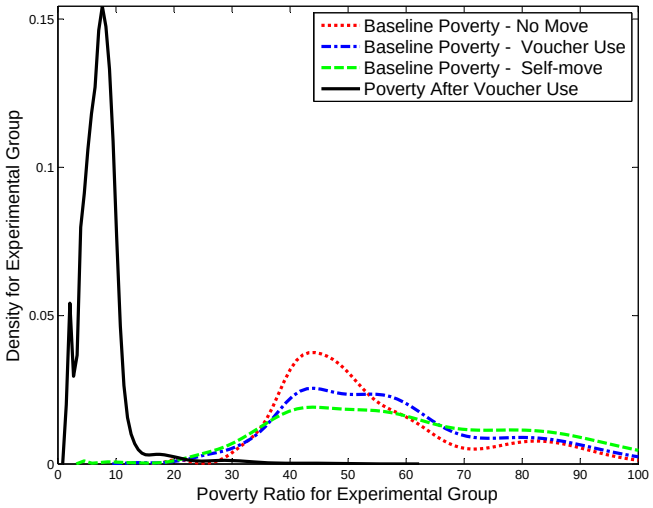
Figure 13 examines neighborhood poverty of families assigned to the Section 8 voucher. It shows a similar pattern as that observed in Figure 12. The poverty levels of Section 8 relocation neighborhoods are lower than those of baseline neighborhoods. However poverty levels of Section 8 relocation neighborhoods are higher than those faced by the families that relocated using the Experimental voucher in Figure 12.

C.2 The Neighborhood Choice

The paper exploits the exogenous variation of voucher assignments to identify the causal effect of neighborhood relocation. To do so, it uses a stylized version of the MTO intervention in which families are faced with three neighborhood choices:

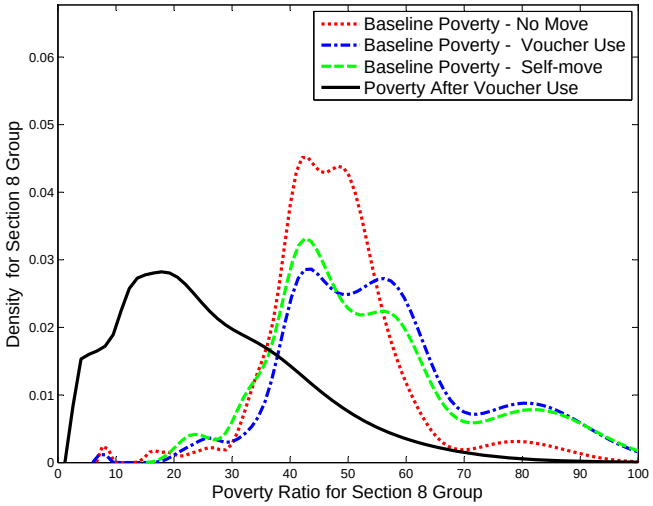
1. High-poverty neighborhood (t_h) which is equivalent to the choice of not relocating;
2. Medium-poverty neighborhood (t_m);
3. Low-poverty neighborhood (t_l);

Figure 12: **Density Estimation of Baseline Neighborhood Poverty (1990 Census) of the Experimental Group by Voucher Compliance**



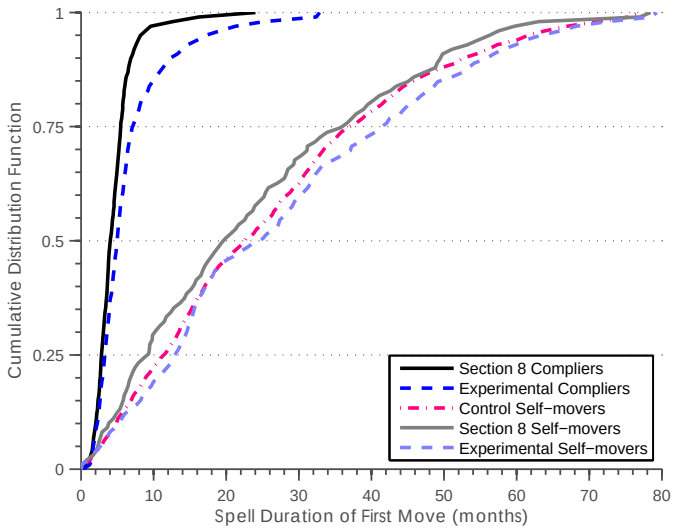
This figure presents the density estimation of baseline neighborhood poverty for the Experimental group conditional on relocation choice, i.e., (1) do not relocate, (2) relocate using the voucher and (3) relocate without using the Experimental voucher. See columns 7–11 of Table 6 for inference on the average level of neighborhood poverty by voucher assignment and compliance. The graph also presents the neighborhood poverty density of the families that use the Experimental voucher after relocation. Estimates are based on the normal kernel with optimal normal bandwidth.

Figure 13: **Density Estimation of Baseline Neighborhood Poverty of the Section 8 Group by Voucher Compliance**



This figure presents the density estimation of baseline neighborhood poverty for the Section 8 group conditional on relocation choice, i.e., (1) do not relocate, (2) relocate using the voucher and (3) relocate without using the Experimental voucher. See columns 12–16 of Table 6 for inference on the average level of neighborhood poverty by voucher assignment and compliance. The graph also presents the neighborhood poverty density of the families that use the Section 8 voucher after relocation. Estimates are based on the normal kernel with optimal normal bandwidth.

Figure 14: **Duration of Spells From Intervention Onset to the First Relocation**



This figure shows the cumulative distribution of month spells from the intervention onset until first move. It shows five curves that can be divided into two groups. The first group focus on experimental and Section 8 families that use the voucher. The second group presents the control, experimental and Section 8 families classified as self-movers at the time of the interim evaluation.

Neighborhood choices are retrieved as following. Experimental families that use the voucher chose low-poverty neighborhoods (t_l). Section 8 families that use the voucher may chose low (t_l) or medium-poverty (t_m) neighborhoods. I use poverty levels of the 1990 U.S. Census to distinguish between these choices. Families that did not move according to the interim evaluation in 2002 chose high-poverty neighborhoods (t_h). It remains to assign the neighborhood choice for families classified as self-movers at the time of the interim evaluation. To do so, I explore the available information on the time spells from the onset of the intervention until first relocation. Table 14 presents the cumulative distribution of months from intervention onset to first relocation.

Voucher compliers were supposed to move within six months of the assignment. However, this rule was not strictly enforced. Nearly all section 8 families and more than 90% of experimental families that comply with the voucher moved within a year. I use a simple approach to assign the neighborhood choice for families classified as self-movers. I set a threshold for the length of self-movers spells such that the distribution of spells of self-movers match the spell distribution of families that chose low or medium-poverty neighborhoods, that is, the voucher compliers. Specifically, I chose a threshold that minimizes the Pearson's chi-squared statistic for the difference in spell distributions. Self-movers families that relocate before this threshold are chose either low or medium-poverty neighborhood according to the poverty level of the chosen neighborhood measured at 1990 US Census.

A counterfactual outcome is defined as the outcome that would occur by setting the family choice at the onset of the intervention to one of the three neighborhood options. Families may

Table A.6: Distribution of Spell Duration of First Relocation by Voucher and Compliance

	Voucher Compliers		No Voucher Use		
Spell Duration	Experimental	Section 8	Control	Experimental	Section 8
50	0.04	0.06	0.03	0.03	0.03
70	0.11	0.14	0.04	0.04	0.05
90	0.20	0.27	0.05	0.06	0.08
110	0.30	0.39	0.07	0.07	0.09
130	0.40	0.54	0.08	0.08	0.11
150	0.49	0.64	0.11	0.10	0.13
170	0.59	0.76	0.12	0.11	0.14
200	0.71	0.88	0.14	0.13	0.18
250	0.80	0.95	0.18	0.15	0.24
375	0.90	0.98	0.27	0.23	0.34
525	0.95	0.99	0.41	0.40	0.45

This table presents the quantiles associated with the the spell duration in days since randomization until the first relocation for MTO participating families. The first column gives values of spell duration. Next two columns provide the quantiles for the families that comply with the Experimental and Section 8 vouchers. The next column provides the quantile for the spell duration of the first move for families assigned to the control group that relocate. The remaining two columns provide quantiles for the spell duration of families assigned to Experimental and Section 8 groups that relocate without using the vouchers.

further relocate after the neighborhood choice. Counterfactual outcomes subsume subsequent relocations induced by the initial neighborhood decision. Neighborhood causal effects are obtained by comparing counterfactual outcomes.

MTO vouchers play the role of instrumental variables because of their impact on neighborhood choice. That is, vouchers impact family outcomes by affecting the family's choice of neighborhood relocation. Voucher assignment is assumed to be independent of the counterfactual outcomes generated by fixing the relocation decisions, even though voucher assignments are not independent of observed outcomes conditioned on relocation choice. Thus, voucher income effects cannot explain the difference in the outcome distribution of the families who relocate to a low poverty neighborhood whether they use their vouchers or not. This difference is explained by the confounding effects of the unobserved family variables that affect both the choice of neighborhood relocation.

D Choice Restrictions and Response Matrix under WARP

Table A.7: Choice Restrictions from WARP

	Counterfactual Choice		Choice Restrictions		
	Voucher	Neighborhood			
1	z_c	t_h	$T_\omega(z_c) = t_h$	$\Rightarrow$	$T_\omega(z_8) \in \{t_h, t_m, t_l\}$ and $T_\omega(z_e) \in \{t_h, t_l\}$
2	z_c	t_m	$T_\omega(z_c) = t_m$	$\Rightarrow$	$T_\omega(z_8) \in \{t_m, t_l\}$ and $T_\omega(z_e) \in \{t_m, t_l\}$
3	z_c	t_l	$T_\omega(z_c) = t_l$	$\Rightarrow$	$T_\omega(z_8) \in \{t_m, t_l\}$ and $T_\omega(z_e) = t_l$
4	z_8	t_h	$T_\omega(z_8) = t_h$	$\Rightarrow$	$T_\omega(z_c) = t_h$ and $T_\omega(z_e) = t_h$
5	z_8	t_m	$T_\omega(z_8) = t_m$	$\Rightarrow$	$T_\omega(z_c) \in \{t_h, t_m, t_l\}$ and $T_\omega(z_e) \in \{t_h, t_m, t_l\}$
6	z_8	t_l	$T_\omega(z_8) = t_l$	$\Rightarrow$	$T_\omega(z_c) \in \{t_h, t_m, t_l\}$ and $T_\omega(z_e) = t_l$
7	z_e	t_h	$T_\omega(z_e) = t_h$	$\Rightarrow$	$T_\omega(z_c) = t_h$ and $T_\omega(z_8) \in \{t_h, t_m\}$
8	z_e	t_m	$T_\omega(z_e) = t_m$	$\Rightarrow$	$T_\omega(z_c) = t_m$ and $T_\omega(z_8) = t_m$
9	z_e	t_l	$T_\omega(z_e) = t_l$	$\Rightarrow$	$T_\omega(z_c) \in \{t_h, t_m, t_l\}$ and $T_\omega(z_8) \in \{t_m, t_l\}$

Suppose a family ω chooses a high-poverty neighborhood under control, i.e. $T_\omega(z_c) = t_h$, then restriction 1 of Table A.7 states that this family chooses either high or low-poverty neighborhoods under the experimental voucher ($T_\omega(z_e) \in \{t_h, t_l\}$). Suppose this family opts for low-poverty, that is $T_\omega(z_e) = t_l$, then restriction 9 states that this family chooses either low or medium-poverty neighborhoods under section 8 ($T_\omega(z_8) \in \{t_m, t_l\}$). If this family opts for low-poverty $T_\omega(z_8) = t_l$, then Restriction 6 states that this family must choose a low-poverty neighborhood under experimental voucher, which complies with the family previous choice. Therefore the response-type $S_\omega \equiv [T_\omega(z_c), T_\omega(z_8), T_\omega(z_e)]' = [t_h, t_l, t_l]'$ complies with WARP. If instead the family had chosen high-poverty neighborhood under Section 8, that is $T_\omega(z_8) = t_h$ then Restriction 4 states that the neighborhood choice under the experimental should be high-poverty neighborhood. Thereby the response-type $S_\omega = [t_h, t_l, t_h]'$ does not comply with WARP.

See Table A.9 of Mathematical Appendix A for the full elimination of response-types of the MTO intervention under WARP. Choice Restriction 8 dominates Choice Restriction 1, that is to say that all response-types eliminated by Choice Restriction 1 are also eliminated by Choice Restriction 8. Choice Restriction 9 dominates Choice Restriction 4. Table A.8 presents the Response Matrix generated by the elimination process of Table A.9.

Table A.8: Response Matrix Generated by WARP

	Counterfactual Choice	Response-types								
		s_1	s_2	s_3	s_4	s_5	s_6	s_7	s_8	s_9
Control	$T_\omega(z_c)$	t_h	t_h	t_h	t_h	t_m	t_m	t_m	t_l	t_l
Section 8	$T_\omega(z_8)$	t_h	t_m	t_m	t_l	t_m	t_m	t_l	t_m	t_l
Experimental	$T_\omega(z_e)$	t_h	t_h	t_l	t_l	t_m	t_l	t_l	t_l	t_l

This table presents the values that the response variable $S_\omega = [T_\omega(z_c)\ ,\ T_\omega(z_8)\ ,\ T_\omega(z_e)]$ can take when instrumental variable Z ranges over, $\text{supp}(Z) = \{z_c, z_8, z_e\}$ and treatment status T ranges over $\text{supp}(T) = \{t_h, t_m, t_l\}$. The first column gives the values of the instrumental variable. The remaining columns enumerate the 9 possible response-types.

Table A.9: Elimination of MTO Response-types for Choice Restrictions from WARP

	Counterfactual Choices			**Choice Restriction Elimination**								
	$T_\omega(z_c)$	$T_\omega(z_8)$	$T_\omega(z_e)$	Res. 1	Res. 2	Res. 3	Res. 4	Res. 5	Res. 6	Res. 7	Res. 8	Res. 9
1	t_h	t_h	t_h	✓	✓	✓	✓	✓	✓	✓	✓	✓
2	t_h	t_h	t_m	✗	✓	✓	✗	✓	✓	✓	✗	✓
3	t_h	t_h	t_l	✓	✓	✓	✗	✓	✓	✓	✓	✗
4	t_h	t_m	t_h	✓	✓	✓	✓	✓	✓	✓	✓	✓
5	t_h	t_m	t_m	✗	✓	✓	✓	✓	✓	✓	✗	✓
6	t_h	t_m	t_l	✓	✓	✓	✓	✓	✓	✓	✓	✓
7	t_h	t_l	t_h	✓	✓	✓	✓	✓	✗	✗	✓	✓
8	t_h	t_l	t_m	✗	✓	✓	✓	✓	✗	✓	✗	✓
9	t_h	t_l	t_l	✓	✓	✓	✓	✓	✓	✓	✓	✓
10	t_m	t_h	t_h	✓	✗	✓	✗	✓	✓	✗	✓	✓
11	t_m	t_h	t_m	✓	✗	✓	✗	✓	✓	✓	✗	✓
12	t_m	t_h	t_l	✓	✗	✓	✗	✓	✓	✓	✓	✗
13	t_m	t_m	t_h	✓	✗	✓	✓	✓	✓	✗	✓	✓
14	t_m	t_m	t_m	✓	✓	✓	✓	✓	✓	✓	✓	✓
15	t_m	t_m	t_l	✓	✓	✓	✓	✓	✓	✓	✓	✓
16	t_m	t_l	t_h	✓	✗	✓	✓	✓	✗	✗	✓	✓
17	t_m	t_l	t_m	✓	✓	✓	✓	✓	✗	✓	✗	✓
18	t_m	t_l	t_l	✓	✓	✓	✓	✓	✓	✓	✓	✓
19	t_l	t_h	t_h	✓	✓	✗	✗	✓	✓	✗	✓	✓
20	t_l	t_h	t_m	✓	✓	✗	✗	✓	✓	✓	✗	✓
21	t_l	t_h	t_l	✓	✓	✗	✗	✓	✓	✓	✓	✗
22	t_l	t_m	t_h	✓	✓	✗	✓	✓	✓	✗	✓	✓
23	t_l	t_m	t_m	✓	✓	✗	✓	✓	✓	✓	✗	✓
24	t_l	t_m	t_l	✓	✓	✓	✓	✓	✓	✓	✓	✓
25	t_l	t_l	t_h	✓	✓	✗	✓	✓	✗	✗	✓	✓
26	t_l	t_l	t_m	✓	✓	✗	✓	✓	✗	✓	✗	✓
27	t_l	t_l	t_l	✓	✓	✓	✓	✓	✓	✓	✓	✓

This table presents all possible values that the response variable S can possibly take when instrumental variable Z ranges over, $\text{supp}(Z) = \{z_c, z_8, z_e\}$ and treatment status T ranges over $\text{supp}(T) = \{t_h, t_m, t_l\}$. The first column enumerates the 27 possible response-types. Columns 2 to 4 presents the response-types according to the vector of the values that $[T_\omega(z_c)\,,\, T_\omega(z_8)\,,\, T_\omega(z_e)]$ takes. Columns 5 to 13 indicate whether the response-type violates any of the following choice restrictions:

Choice Restriction 1	$T_\omega(z_c) = t_h$	$\Rightarrow$	$T_\omega(z_8) \in \{t_h, t_m, t_l\}$ and $T_\omega(z_e) \in \{t_h, t_l\}$
Choice Restriction 2	$T_\omega(z_c) = t_m$	$\Rightarrow$	$T_\omega(z_8) \in \{t_m, t_l\}$ and $T_\omega(z_e) \in \{t_m, t_l\}$
Choice Restriction 3	$T_\omega(z_c) = t_l$	$\Rightarrow$	$T_\omega(z_8) \in \{t_m, t_l\}$ and $T_\omega(z_e) = t_l$
Choice Restriction 4	$T_\omega(z_8) = t_h$	$\Rightarrow$	$T_\omega(z_c) = t_h$ and $T_\omega(z_e) = t_h$
Choice Restriction 5	$T_\omega(z_8) = t_m$	$\Rightarrow$	$T_\omega(z_c) \in \{t_h, t_m, t_l\}$ and $T_\omega(z_e) \in \{t_h, t_m, t_l\}$
Choice Restriction 6	$T_\omega(z_8) = t_l$	$\Rightarrow$	$T_\omega(z_c) \in \{t_h, t_m, t_l\}$ and $T_\omega(z_e) = t_l$
Choice Restriction 7	$T_\omega(z_e) = t_h$	$\Rightarrow$	$T_\omega(z_c) = t_h$ and $T_\omega(z_8) \in \{t_h, t_m\}$
Choice Restriction 8	$T_\omega(z_e) = t_m$	$\Rightarrow$	$T_\omega(z_c) = t_m$ and $T_\omega(z_8) = t_m$
Choice Restriction 9	$T_\omega(z_e) = t_l$	$\Rightarrow$	$T_\omega(z_c) \in \{t_h, t_m, t_l\}$ and $T_\omega(z_8) \in \{t_m, t_l\}$

A check mark sign indicates that the associated response-type does not violates the relation. A cross sign indicates that the associate response-type violates the choice restriction.

E Connection with the Random Utility Model

The identification analysis of Section 4 is the result of the combination of three strategies. The first strategy is to use of MTO vouchers as instrumental variables for neighborhood relocation. The second strategy is o use a causal framework that allows to summarizes the identification problem of neighborhood effects into binary properties of the response matrix. The third one is to rely on economics, i.e. the Weak Axiom of Reveled Preferences (WARP), to reduce the column-dimension of the response matrix and thereby rendering identification results. In the case of MTO, evoking the Strong Axiom of Revealed Preferences (SARP) instead of WARP does not imply in further elimination of response-types.

A related literature in economics studies the effect of individual rationality on aggregate data. A substantial economic literature uses Random Utility Models (RUM) to examine if observed empirical data on prices and consumed goods is consistent an underlying framework where agents maximize utility representing rational preferences (McFadden, 2005). The term random in RUM refers to unobserved heterogeneity across agents. This literature does not uses SARP to identify causal effects, but rather explore how SARP impacts statistical quantities of observed data. McFadden and Richter (1991) coined the term Axiom of Revealed Stochastic Preference (ARSP) for the collection of inequalities that must hold on aggregate data of prices and consumption when heterogeneous individuals are rational. Blundell et al. (2003, 2008) examines the consequences of revealed preferences on the quantiles of Engel curves. They develop a nonparametric estimation of the demand function for consumption goods. Blundell et al. (2014) uses inequality restrictions generated by revealed preferences to investigate the estimation of consumer demand.

A recent paper of Kitamura et al. (2014) implements a nonparametric test that verifies if empirical data comply with the inequalities generated by ARSP. Kitamura et al. (2014) major insight is to form a coarse partition of each budget set W_i such that no other budget set, say W_j, intersect the interior of the partition subsets associated with W_i. This insight allows to transform a continuous utility maximization problem into a discrete problem were the agent selects a consumption bundle that belongs to a finite list of possible choices. They generate a test that explore the choice restrictions generated by SARP. It is useful to clarify Kitamura et al. (2014) approach using a setup that features the MTO experiment. Our goal is to show that the example also generates the same response matrix generated by WARP displayed in Table A.8 of Appendix D.

Let $u_\omega : \text{supp}(K_E) \times \text{supp}(K_S) \times \text{supp}(K_X) \to \mathbb{R}^+$ represent a non-satiable rational preferences for agent ω over the consumption bundle consisting of three goods K_L, K_H and K_X. Let K_X denotes a divisible good in $\mathbb{R}^+$ and K_L, K_H denote indivisible goods whose support is the natural numbers. Let $K = [K_L\,,\, K_H\,,\, K_X]$ to represent a vector of consumption goods associate with the price vector $\boldsymbol{p} = [p_H\,,\, p_L\,,\, p_X] > 0$, such that $\text{supp}(K) = \mathbb{N} \times \mathbb{N} \times \mathbb{R}^+$. Also let the wealth of each agent ω be standardized to 1. Under this setup, the budget plane of any agent ω only depends on price p and is given by $W(\boldsymbol{p}) = \{K \in \mathbb{N} \times \mathbb{N} \times \mathbb{R}^+; pK = 1\}$. The consumption choice for agent ω facing prices

Table A.10: Consumption Bundles

Voucher	Prices	**Possible Consumption Bundles**								
Control	$\boldsymbol{p}^C$	$[0,0,1]$	$[1,0,0]$	$[0,1,0]$	$[0,0,1]$	$[0,0,1]$	$[0,1,0]$	$[0,0,1]$	$[1,0,0]$	$[0,1,0]$
Experimental	$\boldsymbol{p}^E$	$[0,0,1]$	$[1,0,.4]$	$[0,1,0]$	$[1,0,.4]$	$[1,0,.4]$	$[1,0,.4]$	$[0,0,1]$	$[1,0,.4]$	$[1,0,.4]$
Section 8	$\boldsymbol{p}^S$	$[0,0,1]$	$[1,0,.4]$	$[0,1,0.4]$	$[0,1,.4]$	$[1,0,.4]$	$[0,1,.4]$	$[0,1,.4]$	$[0,1,.4]$	$[1,0,.4]$
Voucher	Z Assignment	s_1	s_2	s_3	s_4	s_5	s_6	s_7	s_8	s_9
Control	$Z = z_c$	t_h	t_l	t_m	t_h	t_h	t_m	t_h	t_l	t_m
Section 8	$Z = z_8$	t_h	t_l	t_m	t_m	t_l	t_m	t_m	t_m	t_l
Experimental	$Z = z_e$	t_h	t_l	t_m	t_l	t_l	t_l	t_h	t_l	t_l

This tables presents the combinations of possible consumption bundles that survive SARP according to prices $\boldsymbol{p}^C$, $\boldsymbol{p}^E$ and $\boldsymbol{p}^S$. The table also maps these bundles into the neighborhood choice and voucher assignments. This generates nine response-types.

p_ω is given by:

$$K_\omega(\boldsymbol{p}_\omega) = \underset{k \in W(\boldsymbol{p}_\omega)}{\arg\max}\, u_\omega(k).$$

The econometrician only observes the random sample of $(K_\omega(\boldsymbol{p}_\omega), \boldsymbol{p}_\omega)$.

Now suppose prices can only take three values $\boldsymbol{p}^C = [1,1,1]$, $\mathbf{p}^E = [0.6,1,1]$ and $\boldsymbol{p}^S = [0.6,0.6,1]$. Under discrete goods and non-satiable preferences, the possible consumption bundles are given by: $K_\omega(\boldsymbol{p}^C) \in \{[1,0,0],[0,1,0],[0,0,1]\}$, $K_\omega(\boldsymbol{p}^E) \in \{[1,0,0.4],[0,1,0],[0,0,1]\}$, and $K_\omega(\boldsymbol{p}^S) \in \{[1,0,0.4],[0,1,0.4],[0,0,1]\}$. We are now able to link this consumer model to the MTO experiment. Good K_L indicates the choice of relocating to a low-poverty neighborhood, K_H indicates the choice of relocating to a high-poverty neighborhood and $[K_L, K_H] = [0,0]$ denotes no relocation. The price values represent MTO voucher assignments. Baseline price $\boldsymbol{p}^C$ stands for no voucher. Price $\boldsymbol{p}^E$ stands for the experimental voucher, which subsidizes the relocation to low-poverty neighborhood and price $\boldsymbol{p}^S$ stands for Section 8 voucher, which subsidizes the relocation to either low or high-neighborhood relocation.

For each price there are 3 possible consumption bundles. There are also 3 price vector, which totals 27 possible combinations of consumption bundles across price vectors. The same number of possible response-types. We test which of those combinations satisfy SARP, i.e., if the transitive closure of directly revealed preferences is acyclical. The combinations that do not violate SARP are described in Table A.10. There are a total of nine response-type. The two last response-types of Table A.10 are purged due to Assumption **A-2**.

Section 4 models neighborhood choice using a more natural approach than the one described above. It does not defines relocation decisions as a goods nor assigns prices to neighborhood choices. Instead, the model of Section 4 explores the relation of budget sets generated by voucher assignments and relocation choices. This is a simpler approach as no budget set hyperplane intersects. In other words, symmetric difference of any two budget sets is empty. Budget sets are either identical, disjoint or proper subsets. SARP restrictions are applied directly to choice rules based on the budget sets relations. Kitamura et al. (2014) tests if there is a distribution across potential rational

agent types that would generate the observed distribution of prices and consumption goods. They explain that the agent types distribution is commonly non-identified. In the case of MTO, I was able to identify this distribution, that is, the response-types probabilities of **T-3**.

F Monotone Incentives and Unordered Monotonicity in MTO

Each choice restriction in **L-2** of Section 4 consists of a statement on how family choices change as the instrument varies. A natural question is whether these choice restrictions, generated by revealed preference analysis, could also generated by an approach hat exploits the concept of monotonicity. This question is addressed by Heckman and Pinto (2018) who use a set of indicator inequalities to capture the notion that a shift in instrumental values induce choice changes toward the same direction for all agents. As mention, their unordered monotonicity criteria is defined as:

Unordered Monotonicity. *For each pair of values* $(z, z') \in \text{supp}(Z) \times \text{supp}(Z); z \neq z'$, *and for each treatment* $t \in \text{supp}(T)$, *one of the following inequalities holds for all agents* ω*:*

$$\mathbf{1}[T_\omega(z) = t] \geq \mathbf{1}[T_\omega(z') = t] \quad \text{or} \quad \mathbf{1}[T_\omega(z) = t] \leq \mathbf{1}[T_\omega(z') = t]. \tag{281}$$

Angrist and Imbens (1995) termed monotonicity for the assumption that the choice indicator increases (or decreases) as the instrument varies for all agents. A vast literature that relies on their monotonicity criteria to identify and interpret causal effects. Vytlacil (2004) explains that Angrist and Imbens (1995) monotonicity implies and is implied by ordered choice models. Thereby it does not apply to the unordered case of MTO. In contrast, unordered monotonicity (281) applies to unordered choice models with multiple treatments and categorical instruments. Unordered monotonicity (281) does not imply or is implied by the ordered monotonicity of Angrist and Imbens (1995). Nor it requires that the values of the choice indicator or the instrumental variable be ordered.[39] See Heckman and Pinto (2018) for general properties and applications of unordered monotonicity.

Pinto (2016) shows that, under WARP **L-2** and Normal Choice **A-2**, monotonic incentives (Remark *4.1*) imply unordered monotonicity. In this section I show that the reverse is not true. By reverse I mean that it is not the case that all response matrix where unordered monotonicity holds are generated by an associated incentive matrix with monotonic incentives.

In MTO, unordered monotonicity consists of nine inequalities generated by the combination of three choice values $\big(t_l, t_m, t_h\big)$ and three pairs of instrumental values $\big(\,(z_c, z_8),\ (z_c, z_e),\ (z_e, z_8)\big)$. Consider neighborhood choice t_h and instrument values (z_8, z_e). Under unordered monotonicity, either $\mathbf{1}[T_\omega(z_8) = t_h] \leq \mathbf{1}[T_\omega(z_e) = t_h]$ or $\mathbf{1}[T_\omega(z_8) = t_h] \geq \mathbf{1}[T_\omega(z_e) = t_h]$ holds. One method o infer the direction of the inequality is by comparing the probabilities of the neighborhood choice conditioned on instrumental values, also termed propensity score. For instance, if $P(T = t_h|Z = z_8) < P(T = t_h|Z = z_e)$ then $\mathbf{1}[T_\omega(z_8) = t_h] \leq \mathbf{1}[T_\omega(z_e) = t_h]$ must hold instead of $\mathbf{1}[T_\omega(z_8) =$

[39] In contrast, Angrist and Imbens (1995) monotonicity focuses on ordered values of the choice and the instrumental variable. Let $\text{supp}(T) = \{t_1, t_2, \ldots, t_N\}$ such that $t_{j+1} > t_j; j \in 1, \ldots, N-1$ and the ordered values $\text{supp}(Z) = \{z_1, z_2, \ldots, z_K\}$. Then we can define Angrist and Imbens (1995) monotonicity by:

$$T_\omega(z_j) \geq T_\omega(z_{j+1}) \quad \text{or} \quad T_\omega(z_j) \leq T_\omega(z_{j+1}) \quad \text{for all agents } \omega.$$

$t_h] \geq \mathbf{1}[T_\omega(z_e) = t_h]$.

Table A.11 presents the nine monotonicity relations that generate the response matrix $\boldsymbol{R}$ in **L-3**. Each inequality has a clear justifications. Relation 1 states that a family is more likely to remain in a high-poverty neighborhood under control assignment (no rent-subsidizing voucher) than under Section 8. Relation 2 state that families are more likely to choose a high-poverty neighborhood (no relocation) under experimental than Section 8 voucher. Section 8 voucher subsidises a greater number of dwellings than the experimental voucher, thus it offers more incentives for families to relocate. Relations 3 and 6 state that families are more likely to choose high or medium-poverty neighborhoods under control than under experimental assignment. Indeed, the experimental voucher incentivizes the relocation to a low-poverty neighborhood. Thus some families that choose high or medium-poverty neighborhoods under no incentives (control assignment) would change their choice to low-poverty neighborhoods if a rent subsidy were available.

Table A.11: Monotonicity Relations Generated by Unordered Monotonicity and Propensity Scores

	Values of Z-pairs	T	Propensity Score Comparison					Respective Unordered Monotonicity Relations				
Relation 1	(z_c, z_8)	t_h	$P(T=t_h\|Z=z_c)=0.82$	$>$	$0.34=P(T=t_h\|Z=z_8)$	$\Rightarrow$		$\mathbf{1}[T_\omega(z_c)=t_h]$	$\geq$	$\mathbf{1}[T_\omega(z_8)=t_h]$		
Relation 2	(z_8, z_e)	t_h	$P(T=t_h\|Z=z_8)=0.34$	$<$	$0.44=P(T=t_h\|Z=z_e)$	$\Rightarrow$		$\mathbf{1}[T_\omega(z_8)=t_h]$	$\leq$	$\mathbf{1}[T_\omega(z_e)=t_h]$		
Relation 3	(z_e, z_c)	t_h	$P(T=t_h\|Z=z_e)=0.44$	$<$	$0.82=P(T=t_h\|Z=z_c)$	$\Rightarrow$		$\mathbf{1}[T_\omega(z_e)=t_h]$	$\leq$	$\mathbf{1}[T_\omega(z_c)=t_h]$		
Relation 4	(z_c, z_8)	t_m	$P(T=t_m\|Z=z_c)=0.15$	$<$	$0.57=P(T=t_m\|Z=z_8)$	$\Rightarrow$		$\mathbf{1}[T_\omega(z_c)=t_m]$	$\leq$	$\mathbf{1}[T_\omega(z_8)=t_m]$		
Relation 5	(z_8, z_e)	t_m	$P(T=t_m\|Z=z_8)=0.57$	$>$	$0.07=P(T=t_m\|Z=z_e)$	$\Rightarrow$		$\mathbf{1}[T_\omega(z_8)=t_m]$	$\geq$	$\mathbf{1}[T_\omega(z_e)=t_m]$		
Relation 6	(z_e, z_c)	t_m	$P(T=t_m\|Z=z_e)=0.07$	$<$	$0.15=P(T=t_m\|Z=z_c)$	$\Rightarrow$		$\mathbf{1}[T_\omega(z_e)=t_m]$	$\leq$	$\mathbf{1}[T_\omega(z_c)=t_m]$		
Relation 7	(z_c, z_8)	t_l	$P(T=t_l\|Z=z_c)=0.03$	$<$	$0.09=P(T=t_l\|Z=z_8)$	$\Rightarrow$		$\mathbf{1}[T_\omega(z_c)=t_l]$	$\leq$	$\mathbf{1}[T_\omega(z_8)=t_l]$		
Relation 8	(z_8, z_e)	t_l	$P(T=t_l\|Z=z_8)=0.09$	$<$	$0.49=P(T=t_l\|Z=z_e)$	$\Rightarrow$		$\mathbf{1}[T_\omega(z_8)=t_l]$	$\leq$	$\mathbf{1}[T_\omega(z_e)=t_l]$		
Relation 9	(z_e, z_c)	t_l	$P(T=t_l\|Z=z_e)=0.49$	$>$	$0.03=P(T=t_l\|Z=z_c)$	$\Rightarrow$		$\mathbf{1}[T_\omega(z_e)=t_l]$	$\geq$	$\mathbf{1}[T_\omega(z_c)=t_l]$		

Table A.12 describes the elimination process under unordered monotonicity. It shows that the surviving response-types constitute the same response matrix generated by the revealed preference analysis in Lemma **L-3**. This result adds credibility to the identification result advocated here, but also incites to question whether both approaches are simply different facades of the same underlying identification strategy.

F.1 Understanding Revealed Preference and Monotonicity Approaches

The unordered monotonicity and revealed preference strategies do not imply each other. Here we compare two identification strategies. The first one uses revealed preference analysis and the incentive matrix. The second assumes unordered monotonicity and obtain the direction of the monotonicity relation through propensity score inequalities.

As mentioned, a key property of unordered monotonicity is that it allows the researcher to use information from observed data (i.e the ranking of propensity scores) as a source of identification.

Table A.12: Elimination of MTO Response-types Under Unordered Monotonicity

Counterfactual Choices	**All 27 Possible Response-types**																										
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
$T_\omega(z_c)$	t_h	t_h	t_h	t_h	t_h	t_h	t_h	t_h	t_h	t_m	t_m	t_m	t_m	t_m	t_m	t_m	t_m	t_m	t_l	t_l	t_l	t_l	t_l	t_l	t_l	t_l	t_l
$T_\omega(z_8)$	t_h	t_h	t_h	t_m	t_m	t_m	t_l	t_l	t_l	t_h	t_h	t_h	t_m	t_m	t_m	t_l	t_l	t_l	t_h	t_h	t_h	t_m	t_m	t_m	t_l	t_l	t_l
$T_\omega(z_e)$	t_h	t_m	t_l	t_h	t_m	t_l	t_h	t_m	t_l	t_h	t_m	t_l	t_h	t_m	t_l	t_h	t_m	t_l	t_h	t_m	t_l	t_h	t_m	t_l	t_h	t_m	t_l
Monotonicity 1	✓	✓	✓	✓	✓	✓	✓	✓	✓	✗	✗	✗	✓	✓	✓	✓	✓	✓	✗	✗	✗	✓	✓	✓	✓	✓	✓
Monotonicity 2	✓	✗	✗	✓	✓	✓	✓	✓	✓	✓	✗	✗	✓	✓	✓	✓	✓	✓	✓	✗	✗	✓	✓	✓	✓	✓	✓
Monotonicity 3	✓	✓	✓	✓	✓	✓	✓	✓	✓	✗	✓	✓	✗	✓	✓	✗	✓	✓	✗	✓	✓	✗	✓	✓	✗	✓	✓
Monotonicity 4	✓	✓	✓	✓	✓	✓	✓	✓	✓	✗	✗	✗	✓	✓	✓	✗	✗	✗	✓	✓	✓	✓	✓	✓	✓	✓	✓
Monotonicity 5	✓	✗	✓	✓	✓	✓	✓	✗	✓	✓	✗	✓	✓	✓	✓	✓	✗	✓	✓	✗	✓	✓	✓	✓	✓	✗	✓
Monotonicity 6	✓	✗	✓	✓	✗	✓	✓	✗	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✗	✓	✓	✗	✓	✓	✗	✓
Monotonicity 7	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✗	✗	✗	✗	✗	✗	✓	✓	✓
Monotonicity 8	✓	✓	✓	✓	✓	✓	✗	✗	✓	✓	✓	✓	✓	✓	✓	✗	✗	✓	✓	✓	✓	✓	✓	✓	✗	✗	✓
Monotonicity 9	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✗	✗	✓	✗	✗	✓	✗	✗	✓
Not Eliminated	1			4		6			9					14	15												27

The top section of this table lists all the 27 possible response-types that the response variable $S_\omega = [T_\omega(z_c)\,,\, T_\omega(z_8)\,,\, T_\omega(z_e)]$ can take. Rows present the counterfactual neighborhood choices that would arise if a family were assigned to control group, Section 8 and experimental group, that is $T_\omega(z_c), T_\omega(z_8)$ and $T_\omega(z_e)$ respectively. Columns present all the values of response-type as choices range over $\text{supp}(T) = \{t_h, t_m, t_l\}$. The second section of this table indicate whether the response-type in the column of the first panel violates any of the following monotonicity relations:

	Z-pairs	Values of T	Unordered Monotonicity Relations		
Monotonicity Relation 1	(z_c, z_8)	t_h	$\mathbf{1}[T_\omega(z_c) = t_h]$	$\geq$	$\mathbf{1}[T_\omega(z_8) = t_h]$
Monotonicity Relation 2	(z_8, z_e)	t_h	$\mathbf{1}[T_\omega(z_8) = t_h]$	$\leq$	$\mathbf{1}[T_\omega(z_e) = t_h]$
Monotonicity Relation 3	(z_e, z_c)	t_h	$\mathbf{1}[T_\omega(z_c) = t_h]$	$\geq$	$\mathbf{1}[T_\omega(z_e) = t_h]$
Monotonicity Relation 4	(z_c, z_8)	t_m	$\mathbf{1}[T_\omega(z_c) = t_m]$	$\leq$	$\mathbf{1}[T_\omega(z_8) = t_m]$
Monotonicity Relation 5	(z_8, z_e)	t_m	$\mathbf{1}[T_\omega(z_8) = t_m]$	$\geq$	$\mathbf{1}[T_\omega(z_e) = t_m]$
Monotonicity Relation 6	(z_e, z_c)	t_m	$\mathbf{1}[T_\omega(z_e) = t_m]$	$\leq$	$\mathbf{1}[T_\omega(z_c) = t_m]$
Monotonicity Relation 7	(z_c, z_8)	t_l	$\mathbf{1}[T_\omega(z_c) = t_l]$	$\leq$	$\mathbf{1}[T_\omega(z_8) = t_l]$
Monotonicity Relation 8	(z_8, z_e)	t_l	$\mathbf{1}[T_\omega(z_8) = t_l]$	$\leq$	$\mathbf{1}[T_\omega(z_e) = t_l]$
Monotonicity Relation 9	(z_e, z_c)	t_l	$\mathbf{1}[T_\omega(z_e) = t_l]$	$\geq$	$\mathbf{1}[T_\omega(z_c) = t_l]$

A check mark sign indicates that the response-type indicated by the column in the top of the table does not violate the choice restriction indicated by the row. A cross sign indicates that the associated response-type violates the relation. The last row of the panel indicates the response-types that are not eliminated by any of the monotonicity relatinons.

Different rankings generated distinct sets of surviving response-types. Moreover, the unordered monotonicity analysis is not affected by the incentive design of MTO given the propensity scores ranking. On the other hand, the revealed preference relies on the information of MTO design and is not affected by propensity score rankings. Distinct designs produce different sets of surviving response-types. To clarify ideas, it is useful to represent the MTO design and the propensity score rankings as binary matrices.

The *Incentive Matrix* is a 3×3 binary matrix that characterise the MTO Design. Rows are associated with instrumental values $[z_c, z_8, z_e]$ and columns with neighborhood choices $[t_h, t_m, t_l]$. If voucher $z \in \{z_c, z_8, z_e\}$ incentives neighborhood choice $t \in \{t_h, t_m, t_l\}$ than the associated (z, t) element of the Incentive Matrix takes value 1, otherwise it takes values 0. The resulting matrix is presented in Table A.13. Control assignment (first row) provides no incentives which generate a row of zero elements. Section 8 which incentivizes relocation to medium and low-poverty neighborhoods is represented by the second row $[0, 1, 1]$. Experimental voucher incentivizes low-poverty neighborhood relocation and is represented by third row $[0, 1, 1]$.

The Monotonicity Matrix characterizes the information on the propensity score rankings that is used by unordered monotonicity. Rows are associated with pairs of instrumental values $[(z_c, z_8), (z_8, z_e), (z_e, z_c)]$ and columns with neighborhood choices (t_h, t_m, t_l). Each element associated with a Z-pair $(z, z') \in \{(z_c, z_8), (z_8, z_e), (z_e, z_c)\}$ and a choice $t \in \{t_h, t_m, t_l\}$ takes value 1 if $P(T = t|Z = z) > P(T = t|Z = z')$ and zero otherwise. Thus, the element on the first row and first column takes value 1 if $P(T = t_h|Z = z_c) > P(T = t_h|Z = z_8)$ and zero otherwise. The Monotonicity matrix generated by the MTO data is presented in Table A.13.

Table A.13: MTO Incentive Matrix and MTO Monotonicity Matrix

Group Assignment	Z-values	Incentive Matrix t_h	t_m	t_l	Z-pairs	Monotonicity Matrix t_h	t_m	t_l
Control	z_c	0	0	0	(z_c, z_8)	1	0	0
Section 8	z_8	0	1	1	(z_8, z_e)	0	1	0
Experimental	z_e	0	0	1	(z_e, z_c)	0	0	1

The Incentive Matrix characterise the MTO Design. Rows are associated with instrumental values $[z_c, z_8, z_e]$ and columns with neighborhood choices $[t_h, t_m, t_l]$. If voucher $z \in \{z_c, z_8, z_e\}$ incentives neighborhood choice $t \in \{t_h, t_m, t_l\}$ than the associated (z, t) element of the Incentive Matrix takes value 1, otherwise it takes values 0. The Monotonicity Matrix summarizes the propensity score rankings. Rows are associated with pairs of instrumental values $[(z_c, z_8), (z_8, z_e), (z_e, z_c)]$ and columns with neighborhood choices (t_h, t_m, t_l). Each element associated with a Z-pair $(z, z') \in \{(z_c, z_8), (z_8, z_e), (z_e, z_c)\}$ and a choice $t \in \{t_h, t_m, t_l\}$ takes value 1 if $P(T = t|Z = z) > P(T = t|Z = z')$ and zero otherwise.

The incentive matrix of Table A.13 contains all the information that, when combined with revealed preference analysis, generates the response matrix in Lemma **L-3**. On the same token,

The monotonicity matrix of Table A.13 contains all the information that, when combined with unordered monotonicity, generates the same set of response-types that constitute the response matrix in Lemma **L-3**. A natural question is to inquiry if for every monotonic incentive matrix there is a monotonicity matrix (and vice -versa) that generates the same set of surviving response-types.

Both Incentive and Monotonicity matrices are 3×3 binary matrices. Under no restrictions, each one of the 9 elements of the matrix can take values 0 or 1. This generates $9^2 = 512$ possible binary matrices. However distinct values of the instrument variable induce different incentives. This imply that no two rows of the incentive matrix are identical. This restriction generates 336 possible incentive matrices, each one associated to a particular incentive design. Those, in turn, generate 168 distinct sets of response-types under revealed preference analysis.

The monotonicity matrix describes the ranking of propensity scores. It cannot be the case that all elements in a column are equal to one (nor all equal to zero), otherwise the following impossible cyclical inequality would arise:

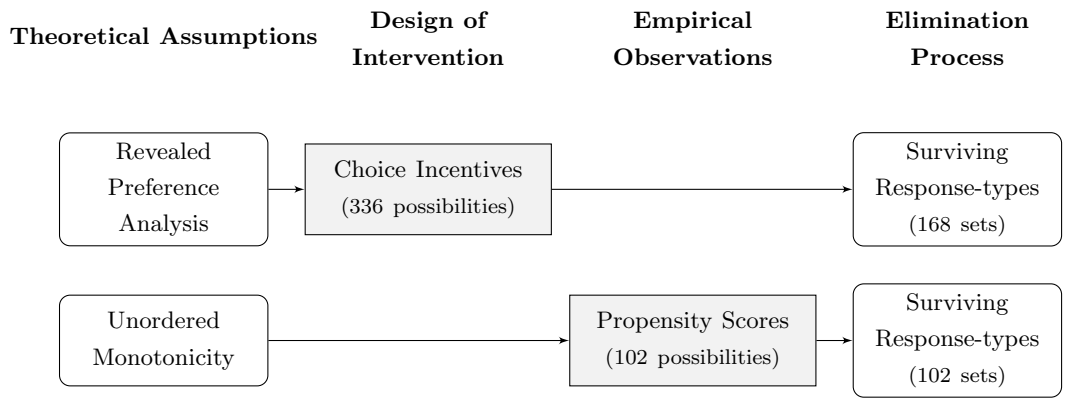
$$P(T = t|Z = z_c) < P(T = t|Z = z_8) < P(T = t|Z = z_e) < P(T = t|Z = z_c) \text{ for some } t \in \{t_h, t_m, t_l\}.$$

It cannot be the case that all elements in a row are equal to one (nor all equal to zero) either, otherwise another impossible inequality would arise:

$$1 = \sum_{t \in \{t_h, t_m, t_l\}} P(T = t|Z = z) < \sum_{t \in \{t_h, t_m, t_l\}} P(T = t|Z = z') = 1 \text{ for some } (z, z') \in \{(z_c, z_8), (z_8, z_e), (z_e, z_c)\}.$$

Therefore the number of configurations of propensity score rankings is equal to the number of possible 3×3 binary matrices whose rows or columns are not all equal to one (or zero). Standard combinatorics shows that there are 102 possible monotonicity matrices. Under unordered monotonicity, these rankings generate 102 distinct sets of response-types. There are 30 pairs of monotonicity and incentive matrices that generates the same set of surviving response-types. One of these pairs represents the case of MTO. Figure 15 summarises this information.

Figure 15: **Summary of Identification Strategies**



G Verifying Condition (14) for MTO Response Matrix of L-3

Theorem (T-1) states that Unordered monotonicity (13) holds if and only if verifying condition (282) is zero.

$$\sum_{t\in\{t_l,t_m,t_h\}} \mathbf{1}'_{3,1}\Big(\big(\boldsymbol{B}'_t(\mathbf{1}_{3,7}-\boldsymbol{B}_t)\big) \odot \big(\boldsymbol{B}'_t(\mathbf{1}_{3,7}-\boldsymbol{B}_t)\big)'\Big)\mathbf{1}_{3,1} = 0. \tag{282}$$

In this section I show that the summation terms $\mathbf{1}'_{3,1}\big(\big(\boldsymbol{B}'_t(\mathbf{1}_{3,7}-\boldsymbol{B}_t)\big) \odot \big(\boldsymbol{B}'_t(\mathbf{1}_{3,7}-\boldsymbol{B}_t)\big)'\big)\mathbf{1}_{3,1}$ for each $t \in \{t_l, t_m, t_h\}$ is zero.

Response matrix $\boldsymbol{R}$ of Lemma L-3 is displayed below:

$$\boldsymbol{R} = \begin{array}{c} \begin{matrix} \boldsymbol{s}_1 & \boldsymbol{s}_2 & \boldsymbol{s}_3 & \boldsymbol{s}_4 & \boldsymbol{s}_5 & \boldsymbol{s}_6 & \boldsymbol{s}_7 \end{matrix} \\ \begin{bmatrix} t_h & t_m & t_l & t_h & t_h & t_m & t_h \\ t_h & t_m & t_l & t_m & t_l & t_m & t_m \\ t_h & t_m & t_l & t_l & t_l & t_l & t_h \end{bmatrix} \end{array} \begin{array}{c} \\ T_\omega(z_c) \\ T_\omega(z_8) \\ T_\omega(z_e) \end{array} \tag{283}$$

Note that the binary matrix $\boldsymbol{B}_t = \mathbf{1}[\boldsymbol{R} = t]; t \in \{t_l, t_m, t_h\}$ stands for matrix that indicates if the elements in $\boldsymbol{R}$ are equal to t. Thus the first term of the summation (282), namelly, $\mathbf{1}'_{3,1}\big(\big(\boldsymbol{B}'_{t_h}(\mathbf{1}_{3,7}-\boldsymbol{B}_{t_h})\big) \odot \big(\boldsymbol{B}'_{t_h}(\mathbf{1}_{3,7}-\boldsymbol{B}_{t_h})\big)'\big)\mathbf{1}_{3,1}$, is given by:

$$\boldsymbol{B}_{t_h} = \begin{array}{c} \begin{matrix} \boldsymbol{s}_1 & \boldsymbol{s}_2 & \boldsymbol{s}_3 & \boldsymbol{s}_4 & \boldsymbol{s}_5 & \boldsymbol{s}_6 & \boldsymbol{s}_7 \end{matrix} \\ \begin{bmatrix} 1 & 0 & 0 & 1 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \end{array} \tag{284}$$

$$\Rightarrow (\mathbf{1}_{3,7}-\boldsymbol{B}_{t_h}) = \begin{array}{c} \begin{matrix} \boldsymbol{s}_1 & \boldsymbol{s}_2 & \boldsymbol{s}_3 & \boldsymbol{s}_4 & \boldsymbol{s}_5 & \boldsymbol{s}_6 & \boldsymbol{s}_7 \end{matrix} \\ \begin{bmatrix} 0 & 1 & 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 & 1 & 1 & 0 \end{bmatrix} \end{array}, \tag{285}$$

$$\therefore \big(\boldsymbol{B}'_{t_h}(\mathbf{1}_{3,7}-\boldsymbol{B}_{t_h})\big) = \begin{bmatrix} 0 & 3 & 3 & 2 & 2 & 3 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 2 & 2 & 1 & 1 & 2 & 0 \end{bmatrix} \tag{286}$$

$$\Rightarrow \big(\boldsymbol{B}'_{t_h}(\mathbf{1}_{3,7}-\boldsymbol{B}_{t_h})\big) \odot \big(\boldsymbol{B}'_{t_h}(\mathbf{1}_{3,7}-\boldsymbol{B}_{t_h})\big)' = \mathbf{0}_{7,7} \tag{287}$$

$$\Rightarrow \mathbf{1}'_{3,1}\big(\big(\boldsymbol{B}'_{t_h}(\mathbf{1}_{3,7}-\boldsymbol{B}_{t_h})\big) \odot \big(\boldsymbol{B}'_{t_h}(\mathbf{1}_{3,7}-\boldsymbol{B}_{t_h})\big)'\big)\mathbf{1}_{3,1} = 0. \tag{288}$$

Equation (284) comes from the definition of $\boldsymbol{B}_t; t \in \{t_h, t_m, t_l\}$. Equality (286) is generated by simple matrix multiplication. Let $\boldsymbol{G}_{t_h} \equiv \big(\boldsymbol{B}'_{t_h}(\mathbf{1}_{3,7}-\boldsymbol{B}_{t_h})\big)$ denotes the matrix multiplication in (286). Matrix $\boldsymbol{G}_{t_h}$ is a square matrix whose dimension is 7×7. Let $\boldsymbol{G}_{t_h}[i,j]$ denotes the element in the i-th row and $j-th$ column of matrix $\boldsymbol{G}_{t_h}$. Note that either $\boldsymbol{G}_{t_h}[i,j] = 0$ or $\boldsymbol{G}_{t_h}[j,i] = 0$ (or

both). Thus it must be the case that $\boldsymbol{G}_{t_h} \odot \boldsymbol{G}'_{t_h} = \mathbf{0}_{7,7}$, where $\mathbf{0}_{7,7}$ denotes a 7×7 matrix of 0s. Otherwise stated, $\big(\boldsymbol{B}'_{t_h}(\mathbf{1}_{3,7} - \boldsymbol{B}_{t_h})\big) \odot \big(\boldsymbol{B}'_{t_h}(\mathbf{1}_{3,7} - \boldsymbol{B}_{t_h})\big)'\big)$ in (287) is equal to $\mathbf{0}_{7,7}$, which implies that $\mathbf{1}'_{3,1}\big(\big(\boldsymbol{B}'_{t_h}(\mathbf{1}_{3,7} - \boldsymbol{B}_{t_h})\big) \odot \big(\boldsymbol{B}'_{t_h}(\mathbf{1}_{3,7} - \boldsymbol{B}_{t_h})\big)'\big)\mathbf{1}_{3,1} = 0$ as stated in (288). The remaining terms for t_m and t_l are also zero and its computation follows the same steps in (284)–(288).

The second term of the summation (282) is $\mathbf{1}'_{3,1}\big(\big(\boldsymbol{B}'_{t_m}(\mathbf{1}_{3,7} - \boldsymbol{B}_{t_m})\big) \odot \big(\boldsymbol{B}'_{t_m}(\mathbf{1}_{3,7} - \boldsymbol{B}_{t_m})\big)'\big)\mathbf{1}_{3,1}$, is given by:

$$\boldsymbol{B}_{t_m} = \begin{array}{c} \begin{matrix} \boldsymbol{s}_1 & \boldsymbol{s}_2 & \boldsymbol{s}_3 & \boldsymbol{s}_4 & \boldsymbol{s}_5 & \boldsymbol{s}_6 & \boldsymbol{s}_7 \end{matrix} \\ \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \end{array} \tag{289}$$

$$\Rightarrow \big(\mathbf{1}_{3,7} - \boldsymbol{B}_{t_m}\big) = \begin{array}{c} \begin{matrix} \boldsymbol{s}_1 & \boldsymbol{s}_2 & \boldsymbol{s}_3 & \boldsymbol{s}_4 & \boldsymbol{s}_5 & \boldsymbol{s}_6 & \boldsymbol{s}_7 \end{matrix} \\ \begin{bmatrix} 1 & 0 & 1 & 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 1 & 1 & 1 \end{bmatrix} \end{array}, \tag{290}$$

$$\therefore \big(\boldsymbol{B}'_{t_m}(\mathbf{1}_{3,7} - \boldsymbol{B}_{t_m})\big) = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 3 & 0 & 3 & 2 & 3 & 1 & 2 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 2 & 0 & 2 & 1 & 2 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 \end{bmatrix} \tag{291}$$

$$\Rightarrow \big(\boldsymbol{B}'_{t_m}(\mathbf{1}_{3,7} - \boldsymbol{B}_{t_m})\big) \odot \big(\boldsymbol{B}'_{t_m}(\mathbf{1}_{3,7} - \boldsymbol{B}_{t_m})\big)'\big) = \mathbf{0}_{7,7} \tag{292}$$

$$\Rightarrow \mathbf{1}'_{3,1}\big(\big(\boldsymbol{B}'_{t_m}(\mathbf{1}_{3,7} - \boldsymbol{B}_{t_m})\big) \odot \big(\boldsymbol{B}'_{t_m}(\mathbf{1}_{3,7} - \boldsymbol{B}_{t_m})\big)'\big)\mathbf{1}_{3,1} = 0. \tag{293}$$

Equation (289) comes from the definition of $\boldsymbol{B}_t; t \in \{t_m, t_m, t_l\}$. Equality (291) is generated by simple matrix multiplication. Let $\boldsymbol{G}_{t_m} \equiv \big(\boldsymbol{B}'_{t_m}(\mathbf{1}_{3,7} - \boldsymbol{B}_{t_m})\big)$ denotes the matrix multiplication in (291). Matrix $\boldsymbol{G}_{t_m}$ is such that either $\boldsymbol{G}_{t_m}[i,j] = 0$ or $\boldsymbol{G}_{t_m}[j,i] = 0$ (or both). Thus it must be the case that $\boldsymbol{G}_{t_m} \odot \boldsymbol{G}'_{t_m} = \mathbf{0}_{7,7}$, where $\mathbf{0}_{7,7}$ denotes a 7×7 matrix of 0s. Otherwise stated, $\big(\boldsymbol{B}'_{t_m}(\mathbf{1}_{3,7} - \boldsymbol{B}_{t_m})\big) \odot \big(\boldsymbol{B}'_{t_m}(\mathbf{1}_{3,7} - \boldsymbol{B}_{t_m})\big)'\big)$ in (292) is equal to $\mathbf{0}_{7,7}$, which implies that $\mathbf{1}'_{3,1}\big(\big(\boldsymbol{B}'_{t_m}(\mathbf{1}_{3,7} - \boldsymbol{B}_{t_m})\big) \odot \big(\boldsymbol{B}'_{t_m}(\mathbf{1}_{3,7} - \boldsymbol{B}_{t_m})\big)'\big)\mathbf{1}_{3,1} = 0$ as stated in (293).

The third term of the summation (282) is $\mathbf{1}'_{3,1}\big(\big(\boldsymbol{B}'_{t_l}(\mathbf{1}_{3,7} - \boldsymbol{B}_{t_l})\big) \odot \big(\boldsymbol{B}'_{t_l}(\mathbf{1}_{3,7} - \boldsymbol{B}_{t_l})\big)'\big)\mathbf{1}_{3,1}$, is given by:

$$\boldsymbol{B}_{t_l} = \begin{array}{c} \begin{matrix} \boldsymbol{s}_1 & \boldsymbol{s}_2 & \boldsymbol{s}_3 & \boldsymbol{s}_4 & \boldsymbol{s}_5 & \boldsymbol{s}_6 & \boldsymbol{s}_7 \end{matrix} \\ \begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 1 & 0 \end{bmatrix} \end{array} \tag{294}$$

$$\Rightarrow \big(\mathbf{1}_{3,7} - \boldsymbol{B}_{t_l}\big) = \begin{array}{c} \begin{matrix} \boldsymbol{s}_1 & \boldsymbol{s}_2 & \boldsymbol{s}_3 & \boldsymbol{s}_4 & \boldsymbol{s}_5 & \boldsymbol{s}_6 & \boldsymbol{s}_7 \end{matrix} \\ \begin{bmatrix} 1 & 1 & 0 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \end{array}, \tag{295}$$

$$\therefore \left(\boldsymbol{B}'_{t_l}(\mathbf{1}_{3,7} - \boldsymbol{B}_{t_l})\right) = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 3 & 3 & 0 & 2 & 1 & 2 & 3 \\ 1 & 1 & 0 & 0 & 0 & 0 & 1 \\ 2 & 2 & 0 & 1 & 0 & 1 & 2 \\ 1 & 1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \tag{296}$$

$$\Rightarrow \left(\boldsymbol{B}'_{t_l}(\mathbf{1}_{3,7} - \boldsymbol{B}_{t_l})\right) \odot \left(\boldsymbol{B}'_{t_l}(\mathbf{1}_{3,7} - \boldsymbol{B}_{t_l})\right)' = \mathbf{0}_{7,7} \tag{297}$$

$$\Rightarrow \mathbf{1}'_{3,1}\Big(\left(\boldsymbol{B}'_{t_l}(\mathbf{1}_{3,7} - \boldsymbol{B}_{t_l})\right) \odot \left(\boldsymbol{B}'_{t_l}(\mathbf{1}_{3,7} - \boldsymbol{B}_{t_l})\right)'\Big)\mathbf{1}_{3,1} = 0. \tag{298}$$

Equation (294) comes from the definition of $\boldsymbol{B}_t; t \in \{t_l, t_l, t_l\}$. Equality (296) is generated by simple matrix multiplication. Let $\boldsymbol{G}_{t_l} \equiv \left(\boldsymbol{B}'_{t_l}(\mathbf{1}_{3,7} - \boldsymbol{B}_{t_l})\right)$ denotes the matrix multiplication in (296). Matrix $\boldsymbol{G}_{t_l}$ shares the same property of $\boldsymbol{G}_{t_h}$ and $\boldsymbol{G}_{t_m}$, that is to say that either $\boldsymbol{G}_{t_l}[i,j] = 0$ or $\boldsymbol{G}_{t_l}[j,i] = 0$ (or both). Thus it must be the case that $\boldsymbol{G}_{t_l} \odot \boldsymbol{G}'_{t_l} = \mathbf{0}_{7,7}$, where $\mathbf{0}_{7,7}$ denotes a 7×7 matrix of 0s. Otherwise stated, $\left(\boldsymbol{B}'_{t_l}(\mathbf{1}_{3,7} - \boldsymbol{B}_{t_l})\right) \odot \left(\boldsymbol{B}'_{t_l}(\mathbf{1}_{3,7} - \boldsymbol{B}_{t_l})\right)'\big)$ in (297) is equal to $\mathbf{0}_{7,7}$, which implies that $\mathbf{1}'_{3,1}\Big(\left(\boldsymbol{B}'_{t_l}(\mathbf{1}_{3,7} - \boldsymbol{B}_{t_l})\right) \odot \left(\boldsymbol{B}'_{t_l}(\mathbf{1}_{3,7} - \boldsymbol{B}_{t_l})\right)'\Big)\mathbf{1}_{3,1} = 0$ as stated in (298).

We conclude that the verifying condition is equal to zero and by Theorem (**T-1**), Unordered Monotonicity (13) holds.

H An Alternative Strategy that Also Generates MTO Response Matrix

Response matrix $\boldsymbol{R}$ in L-3 stems from combining MTO incentives and preference analysis. The response matrix may also arise under alternative strategies. I show that the same response matrix can be generated based on the simple premise that families share similar behavior. Specifically that a change in the instrument that induces a family towards a choice cannot induce another family against the same choice.

Choice assumptions aim to capture similarities on the manner that families decide among neighborhood choices. I explore the assumption that a change in the instrument $z \to z'$ that induces a family ω towards a choice t cannot inducing another family ω' against the same choice t. This assumption is equivalent to the elimination of defiers in the binary choice model. Some notation is needed to generalize this idea.

Let $D_{t,\omega}(z) = \mathbf{1}[T_\omega(z) = t]$ be the indicator that takes value one if family ω chooses neighborhood $t \in \{t_c, t_8, t_f\}$ under voucher $z \in \{z_c, z_8, z_e\}$ and zero otherwise. Consider two families ω and ω' such that family ω chooses t under voucher z but does not choose t under voucher z', while family ω' does the opposite, that is:

$$\text{Family } \omega\text{: } D_{t,\omega}(z) = 1 \text{ and } D_{t,\omega}(z') = 0, \tag{299}$$

$$\text{Family } \omega'\text{: } D_{t,\omega'}(z) = 0 \text{ and } D_{t,\omega'}(z') = 1, \tag{300}$$

$$\text{where } t \in \{t_h, t_m, t_l\}, \quad z, z' \in \{z_c, z_8, z_e\} \text{ and } z \neq z'.$$

If we label family ω as a t-complier, then family ω' is a t-definer.

Now suppose that a change in the instrument influences all families towards or against a choice, then for each treatment choice t and for any instrument change $z \to z'$, there must be only t-defiers or t-compliers. It is not clear which type of family is to be eliminated. We can still exploit the core idea of the binary case by assuming that for a given neighborhood choice and IV change, there may exists only compliers or defiers. Assumption A-3 formalizes this criteria. It states that for any instrumental change $z \to z'$, and any choice t, there cannot be a family ω that shifts its choice towards t while another family w' shift its choice against t.

Assumption A-3. For each $t \in \{t_h, t_m, t_l\}$, and all $z, z' \in \{z_c, z_8, z_e\}$; $z \neq z'$ we have that:

$$\mathbf{1}\big[D_{t,\omega}(z) = 1 \text{ and } D_{t,\omega}(z') = 0\big] \cdot \mathbf{1}\big[D_{t,\omega'}(z) = 0 \text{ and } D_{t,\omega'}(z') = 1\big] = 0 \, \forall \omega, \omega' \in \Omega. \tag{301}$$

Assumption A-3 enable us to the retrieve monotonicity inequalities from propensity score relations. Assumption A-3 states that either $\mathbf{1}\big[D_{t,\omega}(z) = 0 \text{ and } D_{t,\omega}(z') = 1\big] = 0$ holds for all families $\omega \in \Omega$ or $\mathbf{1}\big[D_{t,\omega}(z) = 1 \text{ and } D_{t,\omega}(z') = 0\big] = 0$ holds for all $\omega \in \Omega$. Suppose the restriction $\mathbf{1}\big[D_{t,\omega}(z) = 0 \text{ and } D_{t,\omega}(z') = 1\big] = 0$ is true. Thus the indicators $(D_{t,\omega}(z), D_{t,\omega}(z'))$ can only take values in $\{(0,0),(0,1),(1,1)\}$. This is equivalent to state the following monotonicity inequality:

$$\mathbf{1}[T_\omega(z) = t] \equiv D_{t,\omega}(z) \leq D_{t,\omega}(z') \equiv \mathbf{1}[T_\omega(z') = t] \text{ holds for all } \omega \in \Omega.$$

This monotonicity inequality implies the propensity score relation $P(T = t|Z = z) < P(T = t|Z =$

z'). Lemma **L-14** states this rationale:

Lemma L-14. Under Assumption **A-3** we have that:

$$\begin{aligned} P(T=t|Z=z) < P(T=t|Z=z') \quad &\Rightarrow \quad \mathbf{1}[T_\omega(z)=t] \leq \mathbf{1}[T_\omega(z')=t] \,\forall\, \omega \in \Omega \\ P(T=t|Z=z) > P(T=t|Z=z') \quad &\Rightarrow \quad \mathbf{1}[T_\omega(z)=t] \geq \mathbf{1}[T_\omega(z')=t] \,\forall\, \omega \in \Omega \end{aligned}$$

Lemma **L-14** can be used to generate all monotonicity inequalities based on propensity score comparisons. There are three propensity score relations for each treatment choice $t \in \{t_h, t_m, t_l\}$. This combination produces nine monotonicity inequalities. Those inequalities are exactly the ones displayed in Property **P-1**, and, according to Lemma **L-4**, these inequalities uniquely generate the MTO response matrix $\boldsymbol{R}$ in **L-3**.

There is a key distinction between the method presented in this section and the revealed preference analysis of Sections 4.1–6. Assumption **A-3** relies on propensity scores to generate monotonicity restrictions. In contrast, the revealed preference analysis exploit the incentives of the experiment and uses choice axioms to produce monotonicity restrictions. A benefit of this analysis is that it enables the use of propensity scores to test if these choice axioms are empirically sound. Next section shows that the identification of causal parameters depends only on the properties of response-matrix $\boldsymbol{R}$.

I Examining the Impact of Counseling

The framework proposed here can be used to better understand the nuances of the MTO design. The primary incentive of the experimental voucher is the rent subsidy for families that agree to relocate to low-poverty neighborhoods. A secondary incentive of the experimental voucher is a counseling service offered by local institutions for neighborhood relocation. The MTO response matrix $\boldsymbol{R}$ in Lemma **L-3** is useful to investigate the impact of a variation of counseling intensity. An increase in the intensity of counseling services induce some families to choose a low-poverty neighborhood when assigned to experimental voucher. According to the MTO response matrix, it could affect families of two types $\boldsymbol{s}_2$ and $\boldsymbol{s}_7$. Specifically, an increase in counseling services induce (1) a share of $\boldsymbol{s}_2$-type families to turn into $\boldsymbol{s}_6$-type, that is, $\boldsymbol{s}_2 = [t_m, t_m, t_m]' \mapsto \boldsymbol{s}_2 = [t_m, t_m, t_m]'$ and $\boldsymbol{s}_6 = [t_m, t_m, t_l]'$; and (2) a share of $\boldsymbol{s}_2$-type families to turn into $\boldsymbol{s}_7$-type, that is, $\boldsymbol{s}_7 = [t_h, t_m, t_h]' \mapsto \boldsymbol{s}_7 = [t_h, t_m, t_h]'$ and $\boldsymbol{s}_4 = [t_h, t_m, t_l]'$. Therefore an increase in counseling services reduces the probabilities $P(\boldsymbol{S} = \boldsymbol{s}_2)$ and $P(\boldsymbol{S} = \boldsymbol{s}_7)$ and increases $P(\boldsymbol{S} = \boldsymbol{s}_4)$ and $P(\boldsymbol{S} = \boldsymbol{s}_6)$.

Counseling provides an opportunity to use the framework of Section 4 to model more complex incentives schemes. We can make the case that the experimental voucher has more incentives than Section 8 regarding low-poverty neighborhood relocation. This additional incentive scheme can be easily modeled by the incentive matrix in Table A.14. It differs from the original incentive matrix (2) of Section 4 in the value associated with choice t_l and voucher z_e, that is $\boldsymbol{L}[z_e, t_l] = 2$, that is greater than $\boldsymbol{L}[z_8, t_l] = 1$. The matrix is ordinary, any value bigger than 1 suffices.

Table A.14: MTO Incentive Matrix with Experimental Voucher Counseling

Group Assignment	Z-values	Matrix t_h	Matrix t_m	Matrix t_l
Control	z_c	0	0	0
Section 8	z_8	0	1	1
Experimental	z_e	0	0	2

We can inquire how the incentive structure of Table A.14 changes the response matrix of **L-3**, which relies on the monotonic incentives of matrix 2. Theorem **L-15** presents the response matrix generated by the new incentive matrix in Table A.14.

Lemma L-15. WARP and Normal Choice **A-2** applied to the Incentive matrix of Table A.14 generate the following response matrix:

	Response Matrix							
	$\boldsymbol{s}_1$	$\boldsymbol{s}_2$	$\boldsymbol{s}_3$	$\boldsymbol{s}_4$	$\boldsymbol{s}_5$	$\boldsymbol{s}_6$	$\boldsymbol{s}_7$	$\boldsymbol{s}_8$
$T_\omega(z_c)$	t_h	t_l	t_m	t_h	t_h	t_m	t_h	t_h
$T_\omega(z_8)$	t_h	t_l	t_m	t_m	t_l	t_m	t_m	t_h
$T_\omega(z_e)$	t_h	t_l	t_m	t_l	t_l	t_l	t_h	t_l

Proof. The response matrix is generated by the same elimination process that generates the response matrix of Lemma **L-3**. □

The response matrix in **L-3** differs from the one in **L-3** by the additional response-type $\boldsymbol{s}_8 = [t_h\ ,\ t_h\ ,\ t_l]'$. We can enhance the interpretation of this response-type by comparing the response-matrices under these two incentive schemes. Table A.15 presents the mapping of response-types between response matrices in **L-3** and **L-15**. Under additional incentives, a share of $\boldsymbol{s}_1$-type families that always choose t_h to turn into $\boldsymbol{s}_8$-type that choose t_l when facing the experimental voucher. The probability $P(\boldsymbol{S} = \boldsymbol{s}_8)$ accounts for the share of families that, without counseling, would always choose a high-poverty neighborhood t_h under any assignment. Moreover, under the experimental voucher, those families would switch their choice to a low-poverty due only to counseling, that is to say that they would not switch their choice if only the rent-subsidy were offered.

Table A.15: Mapping of Response-types for Across Incentive Designs

No/Low Counseling Incentive Matrix 2	maps to	High Counseling Incentive Matrix A.14		
$\boldsymbol{s}_2 = [t_m, t_m, t_m]'$	$\mapsto$	$\boldsymbol{s}_2 = [t_m, t_m, t_m]'$	and	$\boldsymbol{s}_6 = [t_m, t_m, t_l]'$
$\boldsymbol{s}_7 = [t_h, t_m, t_h]'$	$\mapsto$	$\boldsymbol{s}_7 = [t_h, t_m, t_h]'$	and	$\boldsymbol{s}_4 = [t_h, t_m, t_l]'$
$\boldsymbol{s}_1 = [t_h, t_h, t_h]'$	$\mapsto$	$\boldsymbol{s}_1 = [t_h, t_h, t_h]'$	and	$\boldsymbol{s}_8 = [t_h, t_h, t_l]'$

J Applying New Tools to The LATE Model

This example is serves as an illustration of the method presented in Section 7.1 of the main paper. It summarizes a discussion also mentioned in Pinto (2016).

Consider a social experiment that randomly assign tuition discounts for prospective students deciding to go to college. Choice T_ω takes value t_1 if student ω decides to go to college and $T_\omega = t_0$ otherwise. The instrument Z that takes value $Z_\omega = z_1$ if student ω receives a tuition discount and $Z_\omega = z_0$ otherwise. The observed outcome is given by $Y_\omega = Y_\omega(t_1)\mathbf{1}[T_\omega = t_1] + Y_\omega(t_0)\mathbf{1}[T_\omega = t_0]$ where $Y_\omega(t_1), Y_\omega(t_0)$ denote the potential outcomes of student ω for college enrollment and no college respectively.

The binary LATE model can be characterized by the binary *incentive matrix* $\boldsymbol{L}$ in (302) whose columns are associated with the choice options and rows with the instrumental values.

$$\text{LATE Incentive Matrix } \boldsymbol{L} = \begin{array}{c} \begin{matrix} t_0 & t_1 \end{matrix} \\ \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \end{array} \begin{matrix} z_0 \\ z_1 \end{matrix} \tag{302}$$

According to (302), we have that $\boldsymbol{L}$ is binary and $\boldsymbol{L}[z_0, t] \leq \boldsymbol{L}[z_1, t]$ for all $t \in \{t_0, t_1\}$. Thus it is a case of monotonic incentives.

Now consider a preferences-based approach to examine the LATE model. The choice set consist of the college choice $t \in \{0,1\}$ and consumption goods $x \in \mathcal{X}$. Let $\mathcal{B}_\omega(z,t) \subset \mathcal{X}$ denotes the budget set of consumer goods for student ω when the college choice is fixed at t and the tuition assignment is fixed at z. A tuition discount would enlarge the student budget conditioned on college enrollment – $\mathcal{B}_\omega(z_0,1) \subset \mathcal{B}_\omega(z_1,1)$ – and has no budget impact when choice is fixed at no college – $\mathcal{B}_\omega(z_0,0) = \mathcal{B}_\omega(z_1,0)$. In summary, budget and incentives are connected by the following rule $\mathbf{L}[z,t] \leq \mathbf{L}[z',t] \Rightarrow \mathcal{B}_\omega(z,t) \subseteq \mathcal{B}_\omega(z',t);\ z,z' \in \{z_1,z_0\}, t \in \{0,1\}$.

The Weak Axiom of Revealed Preferences (WARP) states that if a bundle (t,x) is chosen over (t',x') when both were available then (t,x) is (strictly) revealed preferred to (t',x'), then (t',x') cannot be (strictly) revealed preferred to (t,x), in short: $(t,x) \succ^d_\omega (t',x') \Rightarrow (t',x') \not\succ^d_\omega (t,x)$. If student ω chooses college under no discount, $T_\omega(z_0) = t_1$, then the student prefers some (unobserved) bundle $(t_1,x); x \in \mathcal{B}_\omega(z_1,t_0)$ to all bundles, $(t_0,x'); x' \in \mathcal{B}_\omega(z_0,t_0)$. Under no discount, the student faces the same consumption budget $\mathcal{B}_\omega(z_0,t_0) = \mathcal{B}_\omega(z_1,t_0)$, and WARP implies that $(t_1,x) \succ^d_\omega (t_0,x'''); x''' \in \mathcal{B}_\omega(z_1,t_0)$. Moreover bundle $(t_1,x); x \in \mathcal{B}_\omega(z_0,t_1)$ is available as $\mathcal{B}_\omega(z_0,t_1) \subseteq \mathcal{B}_\omega(z_1,t_1)$. Therefore WARP implies that if a student chooses college under no discount, he will still chooses college if a discount is offered. Notationally, WARP ensures the following choice restriction:

$$T_\omega(z_0) = t_1 \Rightarrow T_\omega(z_1) = t_1. \tag{303}$$

Choice restriction (303) can be equivalently expressed as $\mathbf{1}[T_\omega(z_0) = t_1] \leq \mathbf{1}[T_\omega(z_1) = t_1]$ or $\mathbf{1}[T_\omega(z_0) = t_0] \geq \mathbf{1}[T_\omega(z_1) = t_0]$, therefore unordered monotonicity (13) holds. The unordered monotonicity condition is also checked using the necessary and sufficient condition of Theorem **T-1** further in this appendix.

Let *Response Vector* $\boldsymbol{S} = [T(t_0), T(t_1)]'$ is the 2-dimensional random vector of unobserved counterfactual choices when the instruments fixed at z_0, z_1. The support of S_ω consist of four vectors termed *response-types*: $\boldsymbol{s}_1 = [t_0,t_0]$ (never-takers), $\boldsymbol{s}_2 = [t_0,t_1]$ (compliers), $\boldsymbol{s}_3 = [t_1,t_1]$ (always-takers), $\boldsymbol{s}_4 = [t_1,t_0]$ (defiers). Choice restriction (303) eliminates the defiers $\boldsymbol{s_4}$. The *Response matrix* $\boldsymbol{R}$ in (304) consists on the array of the remaining response-types:

$$\boldsymbol{R} = \begin{matrix} \boldsymbol{s}_1 & \boldsymbol{s}_2 & \boldsymbol{s}_3 \end{matrix} \begin{bmatrix} t_0 & t_0 & t_1 \\ t_0 & t_1 & t_1 \end{bmatrix} \begin{matrix} T_\omega(z_0) \\ T_\omega(z_1) \end{matrix} \tag{304}$$

The vectors of outcome means $\boldsymbol{Q}_Z(t); t \in \{t_0,t_1\}$ can be evaluated through data and are given by:

$$\boldsymbol{Q}_Z(t_0) = \begin{bmatrix} E(Y \cdot \mathbf{1}[T=t_0] | Z = z_0) \\ E(Y \cdot \mathbf{1}[T=t_0] | Z = z_1) \end{bmatrix}, \boldsymbol{Q}_Z(t_1) = \begin{bmatrix} E(Y \cdot \mathbf{1}[T=t_1] | Z = z_0) \\ E(Y \cdot \mathbf{1}[T=t_1] | Z = z_1) \end{bmatrix}. \tag{305}$$

The vectors of propensity scores $\boldsymbol{P}_Z(t); t \in \{t_0,t_1\}$ can also be evaluated though observed data

and are given by:

$$\boldsymbol{P}_Z(t_0) = \begin{bmatrix} P(T=t_0|Z=z_0) \\ P(T=t_0|Z=z_1) \end{bmatrix}, \quad \boldsymbol{P}_Z(t_1) = \begin{bmatrix} P(T=t_1|Z=z_0) \\ P(T=t_1|Z=z_1) \end{bmatrix}. \tag{306}$$

We ought to identify the vector of response-type probabilities $\boldsymbol{P}_S$ and unobserved vectors counterfactual outcome means $\boldsymbol{Q}_S(t_0), \boldsymbol{Q}_S(t_1)$ that are defined as:

$$\boldsymbol{P}_S = \begin{bmatrix} P(\boldsymbol{S}=\boldsymbol{s}_1) \\ P(\boldsymbol{S}=\boldsymbol{s}_2) \\ P(\boldsymbol{S}=\boldsymbol{s}_3) \end{bmatrix}, \boldsymbol{Q}_S(t_0) = \begin{bmatrix} E(Y(t_0)|\boldsymbol{S}=\boldsymbol{s}_1)\,P(\boldsymbol{S}=\boldsymbol{s}_1) \\ E(Y(t_0)|\boldsymbol{S}=\boldsymbol{s}_2)\,P(\boldsymbol{S}=\boldsymbol{s}_2) \\ E(Y(t_0)|\boldsymbol{S}=\boldsymbol{s}_3)\,P(\boldsymbol{S}=\boldsymbol{s}_3) \end{bmatrix}, \boldsymbol{Q}_S(t_1) = \begin{bmatrix} E(Y(t_1)|\boldsymbol{S}=\boldsymbol{s}_1)\,P(\boldsymbol{S}=\boldsymbol{s}_1) \\ E(Y(t_1)|\boldsymbol{S}=\boldsymbol{s}_2)\,P(\boldsymbol{S}=\boldsymbol{s}_2) \\ E(Y(t_1)|\boldsymbol{S}=\boldsymbol{s}_3)\,P(\boldsymbol{S}=\boldsymbol{s}_3) \end{bmatrix}. \tag{307}$$

Our goal is to generate the formulas for the identified counterfactual outcomes in LATE using general formulas that stem from the response matrix $\boldsymbol{R}$. Let $\boldsymbol{B}_{t_0} = \mathbf{1}[\boldsymbol{R}=t_0]$ and $\boldsymbol{B}_{t_1} = \mathbf{1}[\boldsymbol{R}=t_1]$ be the binary matrices that indicates if the elements in $\boldsymbol{R}$ are equal to t_0 and t_1 respectively. Let $\boldsymbol{B}_t = \boldsymbol{C}_t\boldsymbol{A}_t; t \in \{t_0, t_1\}$ stands for the decomposition of the binary matrices $\boldsymbol{B}_t$ generated upon the response matrix $\boldsymbol{R}$. Matrix $\boldsymbol{C}_t$ comprise the non-zero vectors of the respective binary matrix $\boldsymbol{B}_t$, while $\boldsymbol{A}_t$ is a mapping of these vectors into $\boldsymbol{B}_t$. In the LATE example, we have that:

$$\boldsymbol{B}_{t_0} = \mathbf{1}[\boldsymbol{R}=t_0] = \overset{\begin{matrix}\boldsymbol{s}_1 & \boldsymbol{s}_2 & \boldsymbol{s}_3\end{matrix}}{\begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}} = \underbrace{\overset{\begin{matrix}\boldsymbol{s}_2 & \boldsymbol{s}_1\end{matrix}}{\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}}}_{\boldsymbol{C}_{t_0}} \cdot \underbrace{\overset{\begin{matrix}\boldsymbol{s}_1 & \boldsymbol{s}_2 & \boldsymbol{s}_3\end{matrix}}{\begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}}}_{\boldsymbol{A}_{t_0}}, \tag{308}$$

$$\boldsymbol{B}_{t_1} = \mathbf{1}[\boldsymbol{R}=t_1] = \overset{\begin{matrix}\boldsymbol{s}_1 & \boldsymbol{s}_2 & \boldsymbol{s}_3\end{matrix}}{\begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}} = \underbrace{\overset{\begin{matrix}\boldsymbol{s}_2 & \boldsymbol{s}_3\end{matrix}}{\begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix}}}_{\boldsymbol{C}_{t_1}} \cdot \underbrace{\overset{\begin{matrix}\boldsymbol{s}_1 & \boldsymbol{s}_2 & \boldsymbol{s}_3\end{matrix}}{\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}}}_{\boldsymbol{A}_{t_1}}. \tag{309}$$

Theorem **T-1** states that unordered monotonicity holds if and only if the following expression is equal to zero:

$$\sum_{t \in \{t_l, t_m, t_h\}} \boldsymbol{\iota}' \Big((\boldsymbol{C}_t' \bar{\boldsymbol{C}}_t) \odot (\bar{\boldsymbol{C}}_t' \boldsymbol{C}_t) \Big) \boldsymbol{\iota} = 0, \tag{310}$$

where $\boldsymbol{C}_t$ comes from decompositions (10)–(12), $\boldsymbol{\iota}$ stand for vectors of element ones,[40] $\bar{\boldsymbol{C}}_t \equiv (\boldsymbol{\iota}\boldsymbol{\iota}' - \boldsymbol{C}_t)$ is the complement of binary matrix $\boldsymbol{C}_t$, and $\odot$ denotes the Hadamard (element-wise)

[40] Vectors $\boldsymbol{\iota}$ have dimension 3×1 in the case of MTO $\boldsymbol{C}_t$ matrices.

multiplication. In the case of LATE, the equations related to $\boldsymbol{C}_{t_0}$ are given by:

$$\boldsymbol{C}_{t_0} = \begin{bmatrix} \boldsymbol{s}_2 & \boldsymbol{s}_1 \\ 1 & 1 \\ 0 & 1 \end{bmatrix} \Rightarrow (\mathbf{1}_{2,2} - \boldsymbol{C}_{t_0}) = \begin{bmatrix} \boldsymbol{s}_2 & \boldsymbol{s}_1 \\ 0 & 0 \\ 1 & 0 \end{bmatrix} \tag{311}$$

$$\therefore (\boldsymbol{C}'_{t_0}(\mathbf{1}_{2,2} - \boldsymbol{C}_{t_0})) = \begin{bmatrix} \boldsymbol{s}_2 & \boldsymbol{s}_1 \\ 0 & 0 \\ 1 & 0 \end{bmatrix} \text{ and } ((\mathbf{1}_{2,2} - \boldsymbol{C}_{t_0})'\boldsymbol{C}_{t_0}) = \begin{bmatrix} \boldsymbol{s}_2 & \boldsymbol{s}_1 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} \tag{312}$$

$$\Rightarrow \mathbf{1}'_{2,1}\big((\boldsymbol{C}'_{t_0}(\mathbf{1}_{2,2} - \boldsymbol{C}_{t_0})) \odot (\boldsymbol{C}'_{t_0}(\mathbf{1}_{2,2} - \boldsymbol{C}_{t_0}))'\big)\mathbf{1}_{2,1} = 0. \tag{313}$$

The respective equations for $\boldsymbol{C}_{t_1}$ are given by:

$$\boldsymbol{C}_{t_1} = \begin{bmatrix} \boldsymbol{s}_2 & \boldsymbol{s}_3 \\ 0 & 1 \\ 1 & 1 \end{bmatrix} \Rightarrow (\mathbf{1}_{2,2} - \boldsymbol{C}_{t_1}) = \begin{bmatrix} \boldsymbol{s}_2 & \boldsymbol{s}_3 \\ 1 & 0 \\ 0 & 0 \end{bmatrix} \tag{314}$$

$$\therefore (\boldsymbol{C}'_{t_1}(\mathbf{1}_{2,2} - \boldsymbol{C}_{t_1})) = \begin{bmatrix} \boldsymbol{s}_2 & \boldsymbol{s}_3 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} \text{ and } ((\mathbf{1}_{2,2} - \boldsymbol{C}_{t_1})'\boldsymbol{C}_{t_1}) = \begin{bmatrix} \boldsymbol{s}_2 & \boldsymbol{s}_3 \\ 0 & 0 \\ 1 & 0 \end{bmatrix} \tag{315}$$

$$\Rightarrow \mathbf{1}'_{2,1}\big((\boldsymbol{C}'_{t_1}(\mathbf{1}_{2,2} - \boldsymbol{C}_{t_1})) \odot (\boldsymbol{C}'_{t_1}(\mathbf{1}_{2,2} - \boldsymbol{C}_{t_1}))'\big)\mathbf{1}_{2,1} = 0. \tag{316}$$

According to Theorem **T-1**, equations (313) and (313) imply that unordered monotonicity holds, as expected. Now if unordered monotonicity works, then the identified counterfactual outcomes are given by the following formula:

$$\underbrace{\boldsymbol{A}_t\boldsymbol{Q}_S(t) \div \boldsymbol{A}_t\boldsymbol{P}_S(t)}_{\text{Identified Counterfactual Outcomes}} = \underbrace{(\boldsymbol{C}'_t\boldsymbol{C}_t)^{-1}\boldsymbol{C}'_t \cdot \boldsymbol{Q}_Z(t) \div (\boldsymbol{C}'_t\boldsymbol{C}_t)^{-1}\boldsymbol{C}'_t \cdot \boldsymbol{P}_Z(t)}_{\text{Identification Formulas}}; \quad t \in \{t_0, t_1\}, \tag{317}$$

where $\boldsymbol{A}_t\boldsymbol{Q}_S(t)$ is identified by $\boldsymbol{A}_t\boldsymbol{Q}_S(t) = \left(\boldsymbol{C}'_t\boldsymbol{C}_t\right)^{-1}\boldsymbol{C}'_t\boldsymbol{Q}_Z(t)$ and $\boldsymbol{A}_t\boldsymbol{P}_S$ is identified by $\boldsymbol{A}_t\boldsymbol{P}_S = \left(\boldsymbol{C}'_t\boldsymbol{C}_t\right)^{-1}\boldsymbol{C}'_t\boldsymbol{P}_Z(t)$. The left-hand side of Equations (317) for $t = t_0$ is given in equations (318)–(320).

$$\boldsymbol{A}'_{t_0}\boldsymbol{Q}_S(t_0) = \begin{bmatrix} E(Y(t_0)|\boldsymbol{S} = \boldsymbol{s}_2)\, P(\boldsymbol{S} = \boldsymbol{s}_2) \\ E(Y(t_0)|\boldsymbol{S} = \boldsymbol{s}_1)\, P(\boldsymbol{S} = \boldsymbol{s}_1) \end{bmatrix}, \tag{318}$$

$$\boldsymbol{A}'_{t_0}\boldsymbol{P}_S = \begin{bmatrix} P(\boldsymbol{S} = \boldsymbol{s}_2) \\ P(\boldsymbol{S} = \boldsymbol{s}_1) \end{bmatrix}, \tag{319}$$

$$\therefore \boldsymbol{A}_{t_0}\boldsymbol{Q}_S(t_0) \div \boldsymbol{A}_{t_0}\boldsymbol{P}_S = \begin{bmatrix} E(Y(t_0)|\boldsymbol{S} = \boldsymbol{s}_2) \\ E(Y(t_0)|\boldsymbol{S} = \boldsymbol{s}_1) \end{bmatrix}. \tag{320}$$

Following the symmetric approach of (318)–(320), we obtain the following equation for t_1 :

$$\boldsymbol{A}_{t_1}\boldsymbol{Q}_S(t_1) \div \boldsymbol{A}_{t_1}\boldsymbol{P}_S = \begin{bmatrix} E(Y(t_1)|\boldsymbol{S} = \boldsymbol{s}_2) \\ E(Y(t_1)|\boldsymbol{S} = \boldsymbol{s}_3) \end{bmatrix}. \tag{321}$$

Equations (326)–(324) describe the right-hand side of equation (317) for $t = t_0$:

$$(\boldsymbol{C}'_{t_0}\boldsymbol{C}_{t_0})^{-1}\boldsymbol{C}'_{t_0} = \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix}, \tag{322}$$

$$\Rightarrow (\boldsymbol{C}'_{t_0}\boldsymbol{C}_{t_0})\boldsymbol{C}'_{t_0}\boldsymbol{Q}_Z(t_0) = \begin{bmatrix} E(Y \cdot \mathbf{1}[T = t_0]|Z = z_0) - E(Y \cdot \mathbf{1}[T = t_0]|Z = z_1) \\ E(Y \cdot \mathbf{1}[T = t_0]|Z = z_1) \end{bmatrix}, \tag{323}$$

$$\text{and } (\boldsymbol{C}'_{t_1}\boldsymbol{C}_{t_1})\boldsymbol{C}'_{t_1}\boldsymbol{P}_Z(t_0) = \begin{bmatrix} P(T = t_0|Z = z_0) - P(T = t_0|Z = z_1) \\ P(T = t_0|Z = z_1) \end{bmatrix}, \tag{324}$$

$$\Rightarrow \boldsymbol{C}_{t_0}^{-1} = (\boldsymbol{C}'_t\boldsymbol{C}_{t_0})^{-1}\boldsymbol{C}'_{t_0} \cdot \boldsymbol{Q}_Z(t_0) \div (\boldsymbol{C}'_{t_0}\boldsymbol{C}_{t_0})^{-1}\boldsymbol{C}'_{t_0} \cdot \boldsymbol{P}_Z(t_0) = \begin{bmatrix} \frac{E(Y \cdot \mathbf{1}[T=t_0]|Z=z_0) - E(Y \cdot \mathbf{1}[T=t_0]|Z=z_1)}{P(T=t_0|Z=z_0) - P(T=t_0|Z=z_1)} \\ \frac{E(Y \cdot \mathbf{1}[T=t_0]|Z=z_1)}{P(T=t_0|Z=z_1)} \end{bmatrix}. \tag{325}$$

The right-hand side of equation (317) for $t = t_1$ can be obtained following the same steps in (326)–(325):

$$\boldsymbol{C}_{t_1}^{-1} = (\boldsymbol{C}'_{t_1}\boldsymbol{C}_{t_1})^{-1}\boldsymbol{C}'_{t_1} = \begin{bmatrix} -1 & 1 \\ 1 & 0 \end{bmatrix}, \tag{326}$$

$$\Rightarrow (\boldsymbol{C}'_t\boldsymbol{C}_{t_1})^{-1}\boldsymbol{C}'_{t_1} \cdot \boldsymbol{Q}_Z(t_1) \div (\boldsymbol{C}'_{t_1}\boldsymbol{C}_{t_1})^{-1}\boldsymbol{C}'_{t_1} \cdot \boldsymbol{P}_Z(t_1) = \begin{bmatrix} \frac{E(Y \cdot \mathbf{1}[T=t_1]|Z=z_1) - E(Y \cdot \mathbf{1}[T=t_1]|Z=z_0)}{P(T=t_1|Z=z_1) - P(T=t_1|Z=z_0)} \\ \frac{E(Y \cdot \mathbf{1}[T=t_1]|Z=z_0)}{P(T=t_1|Z=z_0)} \end{bmatrix}. \tag{327}$$

The final equation for t_0 arises by equating (320) and (325). The respective equation for t_1 is obtained by equating (321) and (327):

$$\begin{bmatrix} E(Y(t_0)|\boldsymbol{S} = \boldsymbol{s}_2) \\ E(Y(t_0)|\boldsymbol{S} = \boldsymbol{s}_1) \end{bmatrix} = \begin{bmatrix} \frac{E(Y \cdot \mathbf{1}[T=t_0]|Z=z_0) - E(Y \cdot \mathbf{1}[T=t_0]|Z=z_1)}{P(T=t_0|Z=z_0) - P(T=t_0|Z=z_1)} \\ \frac{E(Y \cdot \mathbf{1}[T=t_0]|Z=z_1)}{P(T=t_0|Z=z_1)} \end{bmatrix}, \tag{328}$$

$$\begin{bmatrix} E(Y(t_1)|\boldsymbol{S} = \boldsymbol{s}_2) \\ E(Y(t_1)|\boldsymbol{S} = \boldsymbol{s}_3) \end{bmatrix} = \begin{bmatrix} \frac{E(Y \cdot \mathbf{1}[T=t_1]|Z=z_1) - E(Y \cdot \mathbf{1}[T=t_1]|Z=z_0)}{P(T=t_1|Z=z_1) - P(T=t_0|Z=z_0)} \\ \frac{E(Y \cdot \mathbf{1}[T=t_1]|Z=z_0)}{P(T=t_1|Z=z_0)} \end{bmatrix}. \tag{329}$$

As expected, the counterfactual outcomes $E(Y(t_0)|\boldsymbol{s}_1)$, $E(Y(t_0)|\boldsymbol{s}_2)$, $E(Y(t_1)|\boldsymbol{s}_2)$, $E(Y(t_1)|\boldsymbol{s}_3)$ and the response-type probabilities $P(\boldsymbol{s}_1), P(\boldsymbol{s}_2), P(\boldsymbol{s}_3)$ are identified. Moreover, $E(Y(t_1)|\boldsymbol{s}_2) -$

$E(Y(t_0)|\boldsymbol{s}_2)$ generates the well-known LATE expression:

$$E(Y(t_1)|\boldsymbol{s}_2) = \frac{E(Y \cdot \mathbf{1}[T = t_0]|Z = z_0) - E(Y \cdot \mathbf{1}[T = t_0]|Z = z_1)}{P(T = t_0|Z = z_0) - P(T = t_0|Z = z_1)} \tag{330}$$

$$E(Y(t_0)|\boldsymbol{s}_2) = \frac{E(Y \cdot \mathbf{1}[T = t_0]|Z = z_0) - E(Y \cdot \mathbf{1}[T = t_0]|Z = z_1)}{P(T = t_0|Z = z_0) - P(T = t_0|Z = z_1)} \tag{331}$$

$$= \frac{E(Y \cdot \mathbf{1}[T = t_0]|Z = z_0) - E(Y \cdot \mathbf{1}[T = t_0]|Z = z_1)}{1 - P(T = t_1|Z = z_0) - 1 + P(T = t_0|Z = z_1)} \tag{332}$$

$$= \frac{E(Y \cdot \mathbf{1}[T = t_0]|Z = z_0) - E(Y \cdot \mathbf{1}[T = t_0]|Z = z_1)}{P(T = t_1|Z = z_1) - P(T = t_0|Z = z_0)} \tag{333}$$

$$\therefore E(Y(t_1) - Y(t_0)|\boldsymbol{s}_2) = \frac{E(Y \cdot (\mathbf{1}[T = t_0] + \mathbf{1}[T = t_1])|Z = z_1) - E(Y \cdot (\mathbf{1}[T = t_1] + \mathbf{1}[T = t_0])|Z = z_0)}{P(T = t_1|Z = z_1) - P(T = t_1|Z = z_0)} \tag{334}$$

$$= \frac{E(Y|Z = z_1) - E(Y|Z = z_0)}{P(T = t_0|Z = z_0) - P(T = t_0|Z = z_1)} \tag{335}$$

K The Binary Roy Model and LIV

The well-known Roy model is commonly defined as a binary choice model with a continuous instrument and a separable choice equation. It is useful to discuss the features of the Roy model in light of a general IV framework consisting of three observed variables Z, T, Y, an unobserved random vector $\boldsymbol{V}$ and an error term ϵ_Y defined over the probability space $(\Omega, \sigma - \mathfrak{A}, P)$. The causal relations among these variables are defined by the Choice Equation (336), the Outcome Equation (337) and the Independent Condition (338). This general IV framework is completed by the regularity conditions (339)–(340).

$$\textbf{Choice Equation:} \quad T = f_T(Z, \boldsymbol{V}), \tag{336}$$

$$\textbf{Outcome Equation :} \quad Y = f_Y(T, \boldsymbol{V}, \epsilon_Y), \tag{337}$$

$$\textbf{Independence Condition :} \quad \boldsymbol{V}, Z, \epsilon_Y \text{ are mutually independent}, \tag{338}$$

$$\textbf{Regularity Conditions :} \quad Y(t) = f_Y(t, \boldsymbol{V}, \epsilon_Y); \forall\ t \in \text{supp}(T) \text{ has finite moments}, \tag{339}$$

$$\boldsymbol{V} \text{ is absolutely continuous}, \tag{340}$$

$$P(T = t|Z = z) \neq P(T = t|Z = z') \forall\, z, z' \in \text{supp}(Z), t \in \text{supp}(T). \tag{341}$$

The random variable $Y(t); t \in \text{supp}(T)$ denotes the counterfactual outcome when the treatment (choice) variable is fixed at value t. Independence Condition (338) implies the IV exclusion restriction $Y(t) \perp\!\!\!\perp Z;\ t \in \text{supp}(T)$. Condition (341) is also termed IV relevance, which means that a variation in the values that the instrument takes affects the likelihood of choosing any treatment value t in the support of T.

The Roy model arise from the framework (336)–(340) by adding the four features: (1) instrument Z is continuous (342); (2) choice T is binary (343); (3) there is enough variability in instrument Z to assure full support of the propensity score $P(T = t_1|Z)$; and (4) the choice T is governed by an indicator function that is separable in the observed instrument Z and unobserved confounding vector $\boldsymbol{V}$. Those features are respectively defined by equations (345)–(346).

$$\textbf{Continuous IV:} \quad Z \text{ is absolutely continuous}, \tag{342}$$

$$\textbf{Binary Treatment:} \quad \text{supp}\, T = \{t_0, t_1\}, \tag{343}$$

$$\textbf{Full Support :} \quad \text{for any } p \in [0,1], \text{ there exists } z \in \text{supp}(Z);\ P(T = t_1|Z = z) = p, \tag{344}$$

$$\textbf{Separability :} \quad T = D \cdot t_1 + (1 - D) \cdot t_0, \tag{345}$$

$$\text{such that} \quad D = \mathbf{1}[\zeta(Z) \geq \varphi(\boldsymbol{V})]. \tag{346}$$

The remaining of this section examines the local instrumental variable (LIV) parameter of Heckman and Vytlacil (1999).

Transforming Variables Let $F_{\varphi(\boldsymbol{V})}(u) = P(\varphi(\boldsymbol{V}) \leq u)$ be the CDF of random variable $\varphi(\boldsymbol{V})$ and $U = F_{\varphi(\boldsymbol{V})}(\varphi(\boldsymbol{V}))$ be a transformation of $\varphi(\boldsymbol{V})$ that has uniform distribution $U \sim [0,1]$ due to

continuity of $\boldsymbol{V}$. In this notation, the indicator D in (346) can be restated as:

$$D = \mathbf{1}[\zeta(Z) \geq \varphi(\boldsymbol{V})] = \mathbf{1}[F_{\varphi(\boldsymbol{V})}(\zeta(Z)) \geq F_{\varphi(\boldsymbol{V})}(\varphi(\boldsymbol{V}))] \equiv \mathbf{1}[P(Z) \geq U], \tag{347}$$

$$\therefore E(\mathbf{1}[T = t_1]|Z = z) = P(D = 1|Z = z) = E(\mathbf{1}[P(Z) \geq U]|Z = z) \equiv P(z), \tag{348}$$

where (347) is due to Separability (345) and the fact that $F_{\varphi(\boldsymbol{V})}(u)$ is strictly monotone. Equality (348) arises from the uniform distribution of U. The probability $P(T = t_1|Z = z)$ in (348) is called the propensity score and is a function of $z \in \text{supp}(Z)$ and, fr sake of notational simplicity, it is denoted by $P(z)$.

K.1 Causal Effects

Unobserved variable U causes both the treatment choice T and the outcome Y. It is the source of selection bias and it is also a matching variable, that is to say that counterfactual outcomes are independent of the treatment choice conditioned on U, i.e. $Y(t) \perp\!\!\!\perp T|U$; $t \in \{t_0, t_1\}$. The identification of treatment effects relies on the ability to control for unobserved variable U. If U were observed, then the average treatment effect $E(Y(t_1) - Y(t_0))$ could be evaluated by a weighted average of the difference-in-means conditioned on U over its distribution:

$$E(Y(t_1) - Y(t_0)) = \int_0^1 E(Y(t_1) - Y(t_0)|U = u)du \tag{349}$$

The identification of causal effects requires the evaluation of the expectation of counterfactual outcomes conditional on confounding variable U. It is useful to define the following parameters:

$$\Delta_{t_1}(p) = E(Y(t_1)|U = p), \tag{350}$$

$$\Delta_{t_0}(p) = E(Y(t_0)|U = p), \tag{351}$$

$$\Delta_{MTE}(p) = E(Y(t_1) - Y(t_0)|U = p) = \Delta_{t_1}(p) - \Delta_{t_0}(p). \tag{352}$$

$\Delta_{MTE}(p)$ in (352) is termed the marginal treatment effect. It can be interpreted as the average treatment effect for the set of agents that are indifferent between choosing t_1 or t_0 for a given instrumental value. Indeed, let the instrumental variable Z takes a value $z \in \text{supp}(Z)$. The propensity score is given by $P(z)$, and according to the choice equation (347), and agents that are indifferent between t_1 andt_0 must be the ones whose value U is equal to $U = P(z)$.

Heckman and Vytlacil (2005) show that a range of causal parameters can be expressed as a weighted average of the the marginal treatment effect $\Delta_{MTE}(p)$. The equations below describe the average treatment effect Δ_{ATE}, the treatment on the treated Δ_{TT}, and the treatment on the

untreated Δ_{TT} and functions of the marginal treatment effect $\Delta_{MTE}(p)$.

$$\Delta_{ATE} = E(Y(t_1) - Y(t_0)) = \int_0^1 E(Y(t_1) - Y(t_0)|U = p)dp = \int_0^1 \Delta_{MTE}(p)dp, \tag{353}$$

$$\Delta_{TT} = E(Y(t_1) - Y(t_0)|T = t_1) = \frac{E\Big((Y(t_1) - Y(t_0)) \cdot \mathbf{1}[P(Z) \geq U]\Big)}{E\big(\mathbf{1}[P(Z) \geq U]\big)} = \int_0^1 \Delta_{MTE}(p) \left(\frac{1 - F_{P(Z)}(p)}{P(T = t_1)}\right) dp,$$

$$\Delta_{TUT} = E(Y(t_1) - Y(t_0)|T = t_0) = \frac{E\Big((Y(t_1) - Y(t_0)) \cdot \mathbf{1}[P(Z) < U]\Big)}{E\big(\mathbf{1}[P(Z) < U]\big)} = \int_0^1 \Delta_{MTE}(p) \left(\frac{F_{P(Z)}(p)}{P(T = t_0)}\right) dp,$$

where $F_{P(Z)}(p) \equiv P(P(Z) \leq p)$ denotes the CDF of the propensity score $P(Z)$.

K.2 Identification

Equations (354)–(359) describe the identification of the conditional expectation of counterfactual outcome $E(Y(t_1)|U = p)$. Expectation $E(Y \cdot \mathbf{1}[T = t_1]|P(Z) = p)$ can be evaluated through observed data. Equality (354) applies the separability condition (348). Equality (355) is due to the independence relation $(U, Y(t_1), Y(t_0)) \perp\!\!\!\perp P(Z)$ which is a consequence of the IV exclusion restriction. In summary, equations (354)–(355) show that this expectation is equivalent to integrating $E(Y(t_1)|U = u)$ over $[0, p]$. Equation (356) uses the fact that U is absolutely continuous and thereby the Lebesgue differentiation theorem holds. Equation (356) states that the counterfactual expectation conditional on $U = p; p \in [0, 1]$ can be evaluated by the partial derivative of $E(Y \cdot \mathbf{1}[T = t_1]|P(Z) = p)$ with respect to the propensity score $P(Z)$ at value p.

$$E\Big(Y \cdot \mathbf{1}[T = t_1]|P(Z) = p\Big) = E\Big(Y(t_1) \cdot \mathbf{1}[P(Z) \geq U]|P(Z) = p\Big), \tag{354}$$

$$= E\Big(Y(t_1) \cdot \mathbf{1}[p \geq U]\Big) = \int_0^p E(Y(t_1)|U = u)du. \tag{355}$$

$$\Rightarrow \frac{\partial E(Y \cdot \mathbf{1}[T = t_1]|P(Z) = p)}{\partial p} = E(Y(t_1)|U = p) \equiv \Delta_{t_1}(p). \tag{356}$$

Equations (357)–(358) shows that $E(Y \cdot \mathbf{1}[T = t_1]|P(Z) = p)$ can be equivalently expressed as $E(Y|T = t_1, P(Z) = p)p$, which imply the identifying equation (359).

$$E(Y \cdot \mathbf{1}[T = t_1]|P(Z) = p) = E(Y|T = t_1, P(Z) = p)E(\mathbf{1}[T = t_1]|P(Z) = p) \tag{357}$$

$$= E(Y|T = t_1, P(Z) = p)p \tag{358}$$

$$\therefore \frac{\partial E(Y|T = t_1, P(Z) = p)p}{\partial p} = E(Y(t_1)|U = p) \tag{359}$$

Equations (360)–(365) describe the identification of the conditional expectation of counterfactual outcome $E(Y(t_0)|U = p)$. Those equations follow the same method that enables the identification

of $E(Y(t_1)|U=p)$ in (354)–(359).

$$E\Big(Y\cdot\mathbf{1}[T=t_0]|P(Z)=p\Big)=E\Big(Y(t_0)\cdot\mathbf{1}[P(Z)\geq U]|P(Z)=p\Big) \tag{360}$$

$$=E\Big(Y(t_0)\cdot\mathbf{1}[p<U]\Big)=\int_p^1 E(Y(t_0)|U=u)du \tag{361}$$

$$\Rightarrow\frac{\partial E(Y\cdot\mathbf{1}[T=t_0]|P(Z)=p)}{\partial p}=-E(Y(t_0)|U=p)\equiv-\Delta_{t_0}(p) \tag{362}$$

$$\text{also } E(Y\cdot\mathbf{1}[T=t_0]|P(Z)=p)=E(Y|T=t_0,P(Z)=p)E(\mathbf{1}[T=t_0]|P(Z)=p) \tag{363}$$

$$=E(Y|T=t_0,P(Z)=p)(1-p) \tag{364}$$

$$\therefore\frac{\partial E(Y|T=t_0,P(Z)=p)(1-p)}{\partial p}=-E(Y(t_0)|U=p) \tag{365}$$

Equations (366)–(367) state that the identification of the marginal treatment effect $\Delta_{MTE}(p)$ can be obtained by the derivative of $E(Y|P(Z)=p)$ with respect to p.

$$E\Big(Y\cdot\big(\mathbf{1}[T=t_1]+\mathbf{1}[T=t_0]\big)|P(Z)=p\Big)=E\Big(Y|P(Z)=p\Big), \tag{366}$$

$$\Rightarrow\frac{\partial E\Big(Y|P(Z)=p\Big)}{\partial p}=E(Y(t_1)-Y(t_0)|U=p)\equiv\Delta_{MTE}(p) \tag{367}$$

K.3 Estimation

A typical procedure to estimate the Roy model consists of four steps.

1. The first step consists of estimating the propensity score function $P(Z)$.
2. The second step requires to evaluate the statistical relation between the outcome Y and the propensity score $P(Z)$. It is often the case that outcome Y is estimated as a polynomial function of propensity scores.
3. We can then estimate the marginal treatment $\Delta_{MTE}(p)$ effect by combining the partial derivatives listed in (356), (359), (362), (365), and (367).
4. The average treatment effect is obtained by integrating the $\Delta_{MTE}(p)$ across its support $[0,1]$.

Let $i\in\Omega$ denotes an economic agent i in the sample space Ω. Let $\hat{p}_i$ be the estimated value of the propensity score $P(T=t_1|Z)$ for $Z=z_i$.

Procedure based on $E(Y|P(Z)=p)$ We discuss three procedures that estimate the average treatment effect using distinct identification equations. Let the expectation $E(Y|P(Z)=p)$ be

approximated by a polynomial function of p :

$$E(Y|P(Z)=p) = \kappa_0 + \sum_{j=1}^{J} \kappa_j p^j, \tag{368}$$

$$\Rightarrow \Delta_{MTE}(p) = \frac{\partial E(Y|P(Z)=p)}{\partial p} = \sum_{j=1}^{J} j\kappa_j p^{j-1}, \tag{369}$$

$$\therefore \Delta_{ATE} = \int_0^1 \Delta_{MTE}(u)du = \sum_{j=1}^{J} \kappa_j. \tag{370}$$

The is approach can be estimated by the following procedure:

1. Evaluate the regression $Y_i = \kappa_0 + \sum_{j=1}^{J} \kappa_j \hat{p}^j + \epsilon_i$ using all data sample $\{Y_i, \hat{p}_i;\ i \in \Omega\}$.
2. The estimated average treatment effect is given by $\hat{\Delta}_{ATE} = \sum_{j=1}^{J} \hat{\kappa}_j$, where $\hat{\kappa}_j; j = 1, \ldots, J$ denotes the least squares estimates.

Procedure based on $E(Y\mathbf{1}[T=t]|P(Z)=p)$ Another approach explores approximations of the expectation $E(Y\mathbf{1}[T=t]|P(Z)=p); t \in \{t_0, t_1\}$ by a polynomial functions of p :

$$E(Y\mathbf{1}[T=t_1]|P(Z)=p) = \varphi_{0,t_1} + \sum_{j=1}^{J} \varphi_{j,t_1} p^j, \tag{371}$$

$$\Rightarrow \Delta_{t_1}(p) = \frac{\partial E(Y\mathbf{1}[T=t_1]|P(Z)=p)}{\partial p} = \sum_{j=1}^{J} j\varphi_{j,t_1} p^{j-1}, \tag{372}$$

$$E(Y\mathbf{1}[T=t_0]|P(Z)=p) = \varphi_{0,t_0} + \sum_{j=1}^{J} \varphi_{j,t_0} p^j, \tag{373}$$

$$\Rightarrow \Delta_{t_0}(p) = -\frac{\partial E(Y\mathbf{1}[T=t_0]|P(Z)=p)}{\partial p} = -\sum_{j=1}^{J} j\varphi_{j,t_0} p^{j-1}, \tag{374}$$

$$\text{thus } \Delta_{MTE}(p) = \Big(\Delta_{t_1}(p) - \Delta_{t_0}(p)\Big) = \sum_{j=1}^{J} j\big(\varphi_{j,t_1} + \varphi_{j,t_0}\big)p^{j-1} \tag{375}$$

$$\therefore \Delta_{ATE} = \int_0^1 \Delta_{MTE}(u)du = \sum_{j=1}^{J} \big(\varphi_{j,t_1} + \varphi_{j,t_0}\big). \tag{376}$$

The is approach can be estimated by the following procedure:

1. Use the transformed outcome $Y_i\mathbf{1}[T_i = t_1]$ (instead of Y_i) and evaluate the regression $Y_i\mathbf{1}[T_i = t_1] = \varphi_{0,t_1} + \sum_{j=1}^{J} \varphi_{j,t_1}\hat{p}^j + \epsilon_i$ using all data sample $\{Y_i, \hat{p}_i;\ i \in \Omega\}$.
2. Use the transformed outcome $Y_i\mathbf{1}[T_i = t_0]$ (instead of Y_i) and evaluate the regression $Y_i\mathbf{1}[T_i = t_0] = \varphi_{0,t_0} + \sum_{j=1}^{J} \varphi_{j,t_0}\hat{p}^j + \epsilon_i$ using all data sample $\{Y_i, \hat{p}_i;\ i \in \Omega\}$.
3. The estimated average treatment effect is given by $\hat{\Delta}_{ATE} = \sum_{j=1}^{J} \hat{\varphi}_{j,t_1} + \hat{\varphi}_{j,t_0}$, where $\hat{\varphi}_{j,t_1}, \hat{\varphi}_{j,t_0}; j = 1, \ldots, J$ denotes the least squares estimates.

Procedure based on $E(Y|T=t,P(Z)=p)$ Let the expectation $E(Y|T=t,P(Z)=p); t \in \{t_0,t_1\}$ by a polynomial functions of p :

$$E(Y|T=t_1,P(Z)=p) = \tau_{0,t_1} + \sum_{j=1}^{J} \tau_{j,t_1} p^j, \tag{377}$$

$$\Rightarrow E(Y|T=t_1,P(Z)=p)p = \tau_{0,t_1} p + \sum_{j=1}^{J} \tau_{j,t_1} p^{j+1}, \tag{378}$$

$$\Rightarrow \Delta_{t_1}(p) = \frac{\partial E(Y|T=t_1,P(Z)=p)p}{\partial p} = \tau_{0,t_1} + \sum_{j=1}^{J} (j+1)\tau_{j,t_1} p^j, \tag{379}$$

$$E(Y|T=t_0,P(Z)=p) = \tau_{0,t_0} + \sum_{j=1}^{J} \tau_{j,t_0} (1-p)^j, \tag{380}$$

$$\Rightarrow E(Y|T=t_0,P(Z)=p)(1-p) = \tau_{0,t_0}(1-p) + \sum_{j=1}^{J} \tau_{j,t_0}(1-p)^{j+1}, \text{ expressed as a function of 1-}p \tag{381}$$

$$\Rightarrow \Delta_{t_0}(p) = -\frac{\partial E(Y\mathbf{1}[T=t_0]|P(Z)=p)(1-p)}{\partial p} \tag{382}$$

$$= \frac{\partial E(Y\mathbf{1}[T=t_0]|P(Z)=p)(1-p)}{\partial (1-p)} = \tau_{0,t_0} + \sum_{j=1}^{J} j\tau_{j,t_0}(1-p)^{j-1}, \tag{383}$$

$$\text{thus } \Delta_{MTE}(p) = \Delta_{t_1}(p) - \Delta_{t_0}(p) = \left(\tau_{1,t_0} - \tau_{0,t_0}\right) + \sum_{j=1}^{J} j\left(\tau_{j,t_1} p^j - \tau_{j,t_0}(1-p)^j\right) \tag{384}$$

$$\therefore \Delta_{ATE} = \int_0^1 \Delta_{MTE}(u)du = \left(\tau_{1,t_0} - \tau_{0,t_0}\right) + \sum_{j=1}^{J} \left(\tau_{j,t_1} - \tau_{j,t_0}\right). \tag{385}$$

The is approach can be estimated by the following procedure:

1. Estimate $E(Y|T=t_1,P(Z)=p)$ as a polynomial function of p by evaluating the regression $Y_i = \tau_{0,t_1} + \sum_{j=1}^{J} \tau_{j,t_1} \hat{p}^j + \epsilon_i$ using only the data for treated participants, that is, $\{Y_i, \hat{p}_i \text{ such that } ; T_i = t_1 \text{ for } i \in \Omega\}$.

2. Estimate $E(Y|T=t_0,P(Z)=p)$ as a polynomial function of $1-p$ by evaluating the regression $Y_i = \tau_{0,t_0} + \sum_{j=1}^{J} \tau_{j,t_0}(1-\hat{p}^j) + \epsilon_i$ using only the data for control participants, that is, $\{Y_i, \hat{p}_i \text{ such that } ; T_i = t_0 \text{ for } i \in \Omega\}$.

3. The estimated average treatment effect is given by $\hat{\Delta}_{ATE} = \left(\hat{\tau}_{0,t_1} - \hat{\tau}_{0,t_0}\right) \sum_{j=1}^{J} \hat{\tau}_{j,t_1} - \hat{\tau}_{j,t_0}$, where $\hat{\tau}_{j,t_1}, \hat{\tau}_{j,t_0}; j = 0,\ldots,J$ denotes the least squares estimates.

L Examples that Apply the Estimator in T-6

Theorem T-6 describes a general estimation procedure that applies to case of multiple treatments. The theorem characterizes a broad class of estimation methods including the binary case .

Suppose the IV that takes N_Z values in $z_1, ..., z_{N_Z}$. Each value z renders a propensity score $P_t(z) \equiv P(T=t|Z=z)$ for a choice $t \in \text{supp}(T)$. Let $z_{t,1}, ..., z_{t,N_Z}$, be the ordered sequence of IV

values according to increasing values of the propensity scores for choice t, namely:

$$P_t(z_{t,i}) \equiv P(T=t|Z=z_{t,i}) < P(T=t|Z=z_{t,i+1}) \equiv P_t(z_{t,i+1}) \text{ for } i=1,...,N_Z-1.$$

Note that different choices $t, t' \in \text{supp}(T)$ generate distinct ordering $z_{t,i},...,z_{t,N_Z}$ and $z_{t',i},...,z_{t',N_Z}$ of the same IV values.

Let $h_{t,i} \equiv H_t(z_{t,i})$; $z_{t,i} \in \{z_{t,1},...,z_{t,2}\}$ be a transformation of the IV-values where $f_t : \text{supp}(Z) \to \mathbb{R}$ denotes any function that produces N_Z distinct numbers $h_{t,1},...,h_{t,N_Z}$. In particular, $h_{t,i}$ may stand for the propensity scores $P_t(z_{t,i})$ or any transformation of these propensity scores. For example, $\boldsymbol{\lambda}(x) = \left[x, x^2, x^3\right]$ characterizes a polynomial of degree 3 for $N_Z = 3$.

Each family (agent) $\omega \in \Omega$ is assigned to a instrumental value $Z_\omega \in \{z_{t,1},...,z_{t,N_Z}\}$, which determines the transformation $H_{t,\omega} = H_t(Z_\omega) \in \{h_{t,1},...,h_{t,N_Z}\}$ for family ω. In short, $Z_\omega = z_{t,i} \Rightarrow H_{t,\omega} = h_{t,i}$. Let $\boldsymbol{\lambda}_{t,\omega} = \boldsymbol{\lambda}(H_{t,\omega})$ denotes the vector $\boldsymbol{\lambda}(x)$ for agent ω evaluated at $x = H_{t,\omega}$. The set $\Sigma_t(z)$ consists of response-types that take value t when the IV is set to a value z, i.e. $\Sigma_t(z) = \{\boldsymbol{s}; \boldsymbol{R}[z, \boldsymbol{s}] = t\}$. I use this notation to state an equivalence result across estimation procedures:

Let $P_{t_h}(z_8), P_{t_h}(z_e), P_{t_h}(z_c)$ be the propensity scores for choice t_h and IV values z_8, z_e, z_c respectively. These can be estimated by the methods described in Section 8.2. The values $h_{t_h,z_8}, h_{t_h,z_e}, h_{t_h,z_c}$ are the transformations of the propensity scores $h_{t_h,z} = f(P_{t_h}(z)); z \in \{z_8, z_e, z_c\}$. Examples of function $f(\cdot)$ are the identify $f(p) = p$, the inverse CDF of the Normal distribution $f(p) = \Phi^{-1}(p)$, or the logit function $f(p) = \text{logit}(p)$.[41] Variable $H_{t_h,\omega} = f(P_{t_h}(Z_\omega))$ assigns values h_{t_h,z_8} , h_{t_h,z_e} , h_{t_h,z_c} to each family ω. Setting $\boldsymbol{\lambda}(x)$ to the polynomial[42] $\boldsymbol{\lambda}(x) = \left[x, x^2, x^3\right]$ yields the following linear regressions:

$$D_{\omega,t_h} = \theta_{t_h,1} \cdot H_{t_h,\omega} + \theta_{t_h,2} \cdot H^2_{t_h,\omega} + \theta_{t_h,3} \cdot H^3_{t_h,\omega} + \epsilon_{\omega,D} \tag{386}$$

$$Y_\omega \cdot D_{t,\omega} = \beta_{t_h,1} \cdot H^1_{t_h,\omega} + \beta_{t_h,2} \cdot H^2_{t_h,\omega} + \beta_{t_h,3} \cdot H^3_{t_h,\omega} + \epsilon_{\omega,D} \tag{387}$$

According to **T-6**, there are three identified counterfactual outcomes $\Delta_{t_h,1}, \Delta_{t_h,2}, \Delta_{t_h,3}$ that depend on the ranking of the propensity scores. For the neighborhood choice t_h, the propensity scores are ordered as $P_{t_h}(z_8) < P_{t_h}(z_e) < P_{t_h}(z_c)$. This ranking yields the following identified counterfactual outcomes and respective estimates:

$$\Delta_{t_h,3} = E(Y(t_h)|\boldsymbol{S} \in \boldsymbol{s}_4, \boldsymbol{s}_5), \quad \text{and} \quad \widehat{\Delta_{t_h,3}} = \frac{\sum_{k=1}^3 \left((h_{t_h,z_c})^k - (h_{t_h,z_e})^k \right) \hat{\beta}_{t_h,k}}{\sum_{k=1}^3 \left((h_{t_h,z_c})^k - (h_{t_h,z_e})^k \right) \hat{\theta}_{t_h,k}} \tag{388}$$

$$\Delta_{t_h,2} = E(Y(t_h)|\boldsymbol{S} = \boldsymbol{s}_7), \quad \text{and} \quad \widehat{\Delta_{t_h,2}} = \frac{\sum_{k=1}^3 \left((h_{t_h,z_e})^k - (h_{t_h,z_8})^k \right) \hat{\beta}_{t_h,k}}{\sum_{k=1}^3 \left((h_{t_h,z_e})^k - (h_{t_h,z_8})^k \right) \hat{\theta}_{t_h,k}} \tag{389}$$

$$\Delta_{t_h,1} = E(Y(t_h)|\boldsymbol{S} = \boldsymbol{s}_1), \quad \text{and} \quad \widehat{\Delta_{t_h,1}} = \frac{\sum_{k=1}^3 \left((h_{t_h,z_8})^k - 0^k \right) \hat{\beta}_{t_h,k}}{\sum_{k=1}^3 \left((h_{t_h,z_8})^k - 0^k \right) \hat{\theta}_{t_h,k}}. \tag{390}$$

[41] The logit function, or logistic transformation, is the inverse of logistic CDF and is given by the logarithm of the odds-ration, that is, $\text{logit}(p) = \log\left(\frac{p}{1-p}\right)$.

[42] The Taylor expansion justifies the adoption of a local polynomial to approximate a function. This approach is common in the literature of policy evaluation that corrects for endogeneity by estimating a function of the outcomes on the propensity scores. See, for instance, Fan and Gijbels (1996); Heckman (1980); Heckman et al. (1985).

M Connecting T-6 with Identifying Equations (38) of Section 7.1

Estimator **T-6** can be expressed as a matrix format using the following notation. Let the 3×1 vector $\boldsymbol{H}_t = [H_t(z_c); H_t(z_8); H_t(z_e)]$ be the propensity score estimates for $t \in \{t_h, t_m, t_l\}$ that can be evaluated as $\boldsymbol{H}_t = \boldsymbol{B}_t \cdot \widehat{\boldsymbol{P}}_S$ where $\widehat{\boldsymbol{P}}_S$ is estimated as in **L-7**. Let the 3×3 matrix λ_t be defined as $\boldsymbol{\lambda}_t = [\boldsymbol{\lambda}(H_t(z_c)), \boldsymbol{\lambda}(H_t(z_8)), \boldsymbol{\lambda}(H_t(z_e))]'$. Let $\boldsymbol{\lambda}_{t,\Omega} = [\boldsymbol{\lambda}'_{t,\omega}; \omega \in \Omega]$ be the matrix that stacks the transpose of the 3×1 vector $\boldsymbol{\lambda}_{t,\omega} \equiv \boldsymbol{\lambda}(H_t(Z_\omega))$ across all families $\omega \in \Omega$. In the same fashion, $\mathbf{D}_{t,\Omega}$ is the vector that stacks the indicator $D_{t,\omega}$ across all families $\omega \in \Omega$ and $\mathbf{Y}_\Omega$ stacks the outcome Y_ω across families $\omega \in \Omega$. **T-6** estimate for the identified counterfactual outcomes $\big(\boldsymbol{A}_t \boldsymbol{Q}_S(t)\big) \div \big(\boldsymbol{A}_t \boldsymbol{P}_S\big)$ of equation (38) is obtained by the following equation:

$$\underbrace{\Big[\big(\boldsymbol{C}_t^{-1}\boldsymbol{\lambda}_t\big)\Big(\boldsymbol{\lambda}'_{t,\Omega}\boldsymbol{\lambda}_{t,\Omega}\Big)^{-1}\boldsymbol{\lambda}'_{t,\Omega}\Big(\mathbf{Y}_\Omega \odot \mathbf{D}_{t,\Omega}\Big)\Big]}_{\text{Estimate of } \boldsymbol{A}_t\boldsymbol{Q}_S(t)} \div \underbrace{\Big[\big(\boldsymbol{C}_t^{-1}\boldsymbol{\lambda}_t\big)\Big(\boldsymbol{\lambda}'_{t,\Omega}\boldsymbol{\lambda}_{t,\Omega}\Big)^{-1}\boldsymbol{\lambda}'_{t,\Omega}\mathbf{D}_{t,\Omega}\Big]}_{\text{Estimate of } \boldsymbol{A}_t\boldsymbol{P}_S} \tag{391}$$

N Connection of the Estimator in T-6 and the Control Function Approach

The estimator (82) can be interpreted as a control function. For examples of the literature in policy evaluation that uses control functions, see Ahn and Powell (1993); Heckman and Robb (1985); Heckman and Urzúa (2010); Powell (1994); Wooldridge (2015).

The outcome $Y = f_Y(T, \boldsymbol{V}, \epsilon_Y)$ is a function of choice T and unobserved variables $\boldsymbol{V}$ that induce selection bias. Let conditional expectation $E(Y|T = t, P_t = p)$ be written as the sum of the a constant term $\beta_{1,t}$ and a term $E(\xi_t(\boldsymbol{V})|T = t, P_t = p)$ that depends on $\boldsymbol{V}$ as in (392). Under unordered monotonicity, $T = t$ is equivalent to $P_t \geq U_t$. Moreover, $P_t \perp\!\!\!\perp \boldsymbol{V}$ due to $\boldsymbol{V} \perp\!\!\!\perp Z$ in (25). This enables to express the expectation $E(\xi_t(\boldsymbol{V})|T = t, P_t = p) = E(\xi_t(\boldsymbol{V})|U_t \leq p)$ as a control function $K_t(p)$ in equation (392).

$$E(Y|T = t, P_t = p) = \beta_{1,t} + E(\xi_t(\boldsymbol{V})|T = t, P_t = p) = \beta_{1,t} + E(\xi_t(\boldsymbol{V})|U_t \leq p) = \beta_{1,t} + K_t(p). \tag{392}$$

Let the control function be approximated by the local polynomial $K_t(p) = \sum_{k=2}^{K} \beta_{k,t} p^{k-1}$. Equation (393) shows that the expectation $E(Y \cdot \mathbf{1}[T = t]|P_t = p)$ is equal to $\boldsymbol{\lambda}(p)\boldsymbol{\beta}_t$ where $\boldsymbol{\lambda}(h) = [h, h^2, ..., h^K]$ and $\boldsymbol{\beta}_t = [\beta_{1,t}, ..., \beta_{K,t}]'$.

$$E(Y \cdot \mathbf{1}[T = t]|P_t = p) = E(Y|T = t, P_t = p)p = \beta_{1,t}p + \left(\sum_{k=2}^{K} \beta_{k,t} p^{k-1}\right) p = \sum_{k=1}^{K} \beta_{k,t} p^k = \boldsymbol{\lambda}(p)\boldsymbol{\beta}_t. \tag{393}$$

The propensity score $P_t(Z) = P(T = t|Z = p)$ is a one-to-one (injective) function of the instrument Z. Thus $E(Y \cdot \mathbf{1}[T = t]|P_t = p) = E(Y \cdot \mathbf{1}[T = t]|Z = z)$ for $z \in \text{supp}(Z)$ such that $P_t(z) = p$. We can also write the probability $P(T = t|Z = z) = \boldsymbol{\lambda}(p)\boldsymbol{\theta}_t$ where $\boldsymbol{\theta}_t = [1, 0, ..., 0]'$ and $P_t(z) = p; z \in \text{supp}(Z)$. Thereby replacing $E(Y \cdot \mathbf{1}[T = t]|Z = z)$ and $P(T = t|Z = z)$ in (80)–(81) by the expressions $\boldsymbol{\lambda}(p)\boldsymbol{\beta}_t$ and $\boldsymbol{\lambda}(p)\boldsymbol{\theta}_t$ generates the estimator (82) in **T-6**.

The estimation procedure in **T-6** and the 2SLS of **T-4** produce the same counterfactual outcome

estimates. In the case of the 2SLS, the counterfactual outcomes are obtained by the estimates of the parameter β in the second stage equation (55). In contrast, **T-6** evaluates a function of the outcome-choice $Y \cdot T$ interaction on the transformed propensity scores h_t. Counterfactual estimate $\Delta_{t,i}$ is obtained by evaluating the outcome function at values $h_{t,i}, h_{t,i-1}$. The function can also be used to disentangle the counterfactual outcomes that are conditional on two response-types.

O Detailed Description of the Estimation Procedures for Identified Parameters

The estimates presented in the Empirical Section 11 are based on both the observed data and on the properties of the response matrix $\boldsymbol{R}$ generated by the revealed preference analysis. There are four types of estimates:

1. Response-type Probabilities (Figure 5).
2. Pre-program variable means conditioned on response-types (Table 12).
3. Counterfactual outcomes conditioned on response-types (Figures 7–9).
4. Causal effects (Figure 10).

The estimates are based on either the ordinary least square regression (OLS) or the two stage least square regression (2SLS). The estimation of Response-type Probabilities (Figure 5) is based on Lemma **L-7** and based on an OLS regression. These estimates are described in Section O.3.

The estimation of Pre-program variable means conditioned on response-types of Table 12 are also obtained via OLS. The procedure builds upon Lemma **L-7** as described in Appendix A.13. A detailed description of these estimates is presented in Section O.4.

Counterfactual outcomes conditioned on response-types (Figures 7–9) are based on Theorems **T-4**–**T-6** and are obtained via 2SLS. A detailed description of these estimates is presented in Section O.5 and also in Appendix P.

For sake of notational simplicity, the theoretical results focus on unconditional estimates that assign equal weights to observed data. However, the estimates displayed in Section 11 are conditioned on pre-program variables and are weighted according to the adult survey weights of the MTO Interim Impacts Evaluation (2003), Appendix B. Sections O.3–O.5 explain how to account for conditioning and weighting in each of the estimation procedures.

The input of each estimation is based on transformations of the observed data. The notation for observed data is presented in Section O.1 below. The notation for the data transformations is presented in Section O.2. The notation of these two sections is used in Sections O.3–O.5 which describe the estimation procedures.

O.1 Notation for Observed Data

The observed variables are given by pre-program variables X, instrumental variable Z, choice T, outcome Y and weights W. Those are represented by the following notation:

1. Let the sample size be N and the observed variables in the data set be indexed by $\omega \in \Omega$ where Ω stands for the indexing set defined as $\Omega = \{1, ..., N\}$.
2. Let $\mathbb{T}$ denotes the $N \times 1$ vector of choices. In the case of MTO, there are three possible choices represented by t_h, t_m and t_l.
3. Let $\mathbb{Z}$ denotes the $N \times 1$ vector of instrumental values. In the case of MTO, there are three possible instrumental values represented by z_c, z_8 and t_e.
4. Let $\mathbb{X}$ denotes the $N \times K$ matrix that represents K pre-program variables we wish to control for.
5. Let $\mathbb{X}_k$ denotes the k-th column of the pre-program matrix $\mathbb{X}$.
6. Let $\mathbb{X}_0$ denotes the $N \times K$ matrix that represents the standardized version of the K pre-program variables in $\mathbb{X}$. Each vector in $\mathbb{X}_0$ has sample mean equal to zero.
7. Let $\mathbb{S}$ denotes the $N \times 4$ binary matrix of site indicators for 4 out of the 5 intervention sites.
8. Let $\mathbb{S}_0$ denotes the standardize version of binary matrix $\mathbb{S}$.
9. Let $\mathbb{Y}$ denotes the $N \times 1$ vector of the outcome of interest.
10. Let $\mathbb{W}$ denotes the $N \times 1$ vector of positive weights.
11. Let diag($\mathbb{W}$) be the $N \times N$ diagonal matrix whose diagonal elements are given by the vector $\mathbb{W}$.

Pre-program Variables

As mentioned, all estimates are conditioned on pre-program variables of the MTO intervention. These baseline variables are classified into the following categories: site, family characteristics, mobility, neighborhood safety and neighborhood satisfaction. The list below describe the variables in each category:

1. Site:
   The intervention was implemented in five cities – Baltimore, Boston, Chicago, Los Angeles, and New York. Each regression has four dummy variables that indicate if the resident lives in Baltimore, Boston, Chicago or Los Angeles.
2. Family characteristics:

(a) If resident ever married.

(b) If resident has no teenagers in the household.

(c) If resident has a family member that is disabled.

3. Mobility:

(a) If resident had ever applied for a Section 8 housing voucher in the past.

(b) If resident has moved at least 3 times within 5 years previous to the onset of the intervention.

4. Neighborhood safety:

(a) If resident was a victim of battery (being beaten) or assaulted in the past 6 months prior to the intervention.

(b) If resident has ever moved to another location due gang (criminal) activity.

(c) If the resident feels unsafe at night in the neighborhood.

5. Neighborhood satisfaction:

(a) If resident reported that has no friends in the neighborhood.

(b) If resident has watched for neighbor's children.

(c) If resident has no family in the neighborhood.

(d) If resident chats with neighbor.

(e) The neighborhood dissatisfaction index.

O.2 Data Transformations

The data transformations necessary to perform the estimation procedures are of four types.

1. Indicators associated with the values that the instrumental variable $\mathbb{Z}$ and choice $\mathbb{T}$ take.

2. Element-wise multiplication of the choice indicators and the outcome.

3. Binary matrices based on the response matrix $\boldsymbol{R}$.

4. Stacked vectors and matrices of observed data.

These three types of data transformations are described below.

1. Variable Indicators

The elements of the observed vector $\mathbb{Z}$ that represents the instrumental variable Z can take values in $\{z_c, z_8, z_e\}$. The elements of the observed vector $\mathbb{T}$ that represents choice T can take values in $\{t_l, t_m, t_h\}$. The information contained in these two variables is assessed through the following binary vectors:

- Let $\mathbb{I}_{z_c},\ \mathbb{I}_{z_8},\ \mathbb{I}_{z_e}$ denote the $N \times 1$ vectors that indicate whether $\mathbb{Z}$ takes the values z_c, z_8 and z_e respectively.

- Let $\mathbb{D}_{t_h},\ \mathbb{D}_{t_m},\ \mathbb{D}_{t_l}$ denote the $N \times 1$ vectors that indicate whether $\mathbb{T}$ takes the values t_h, t_m and t_l respectively.

2. Element-wise Multiplication

The dependent variable used in the estimation of counterfactual outcomes is the element-wise multiplication between the outcome $\mathbb{Y}$ and each of the binary vectors that indicate choice, that is, $\mathbb{D}_t; t \in \{t_h, t_m, t_l\}$. These element-wise multiplications are denote by $\mathbb{Y} \odot \mathbb{D}_t$, where $\odot$ denotes the Hadamard multiplication.

- Let $\mathbb{Y} \odot \mathbb{D}_{t_h},\ \mathbb{Y} \odot \mathbb{D}_{t_m},\ \mathbb{Y} \odot \mathbb{D}_{t_l}$ denote the $N \times 1$ vectors of element-wise multiplication between outcome $\mathbb{Y}$ and choice indicators $\mathbb{D}_{t_h}, \mathbb{D}_{t_m}, \mathbb{D}_{t_l}$ respectively.

- Let $\mathbb{X}_k \odot \mathbb{D}_{t_h},\ \mathbb{X}_k \odot \mathbb{D}_{t_m},\ \mathbb{X}_k \odot \mathbb{D}_{t_l}$ denote the $N \times 1$ vectors of element-wise multiplication between outcome the k-th column of baseline data in $\mathbb{X}$, that is $\mathbb{X}[\cdot, k]$, and choice indicators $\mathbb{D}_{t_h}, \mathbb{D}_{t_m}, \mathbb{D}_{t_l}$ respectively.

3. Response Matrix Indicators

The response matrix the 3×7 matrix generated by the revealed preference analysis displayed in (394):

$$\begin{array}{c} \\ \boldsymbol{R} = \end{array} \begin{array}{c} \begin{array}{ccccccc} \boldsymbol{s}_1 & \boldsymbol{s}_2 & \boldsymbol{s}_3 & \boldsymbol{s}_4 & \boldsymbol{s}_5 & \boldsymbol{s}_6 & \boldsymbol{s}_7 \end{array} \\ \begin{bmatrix} t_h & t_m & t_l & t_h & t_h & t_m & t_h \\ t_h & t_m & t_l & t_m & t_l & t_m & t_m \\ t_h & t_m & t_l & t_l & t_l & t_l & t_h \end{bmatrix} \end{array} \begin{array}{c} \\ T_\omega(z_c) \\ T_\omega(z_8) \\ T_\omega(z_e) \end{array} \tag{394}$$

The estimation assess the content of the response matrix $\boldsymbol{R}$ via binary matrices $\boldsymbol{B}_t = 1[\boldsymbol{R} = t]$ that have the same dimension of $\boldsymbol{R}$ and take the value 1 if the respective element in $\boldsymbol{R}$ is equal to choice t. In the case of MTO, there are three possible choices $\{t_h, t_m, t_l\}$ which are associated to three binary matrices $\boldsymbol{B}_{t_h}, \boldsymbol{B}_{t_m}, \boldsymbol{B}_{t_l}$ listed below.

$$\boldsymbol{B}_{t_h} = \begin{array}{c} \begin{matrix} \boldsymbol{s}_1 & \boldsymbol{s}_2 & \boldsymbol{s}_3 & \boldsymbol{s}_4 & \boldsymbol{s}_5 & \boldsymbol{s}_6 & \boldsymbol{s}_7 \end{matrix} \\ \begin{bmatrix} 1 & 0 & 0 & 1 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \end{array} \begin{array}{l} \mathbf{1}[T_\omega(z_c) = t_h] \\ \mathbf{1}[T_\omega(z_8) = t_h] \\ \mathbf{1}[T_\omega(z_e) = t_h] \end{array} \tag{395}$$

$$\boldsymbol{B}_{t_m} = \begin{array}{c} \begin{matrix} \boldsymbol{s}_1 & \boldsymbol{s}_2 & \boldsymbol{s}_3 & \boldsymbol{s}_4 & \boldsymbol{s}_5 & \boldsymbol{s}_6 & \boldsymbol{s}_7 \end{matrix} \\ \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \end{array} \begin{array}{l} \mathbf{1}[T_\omega(z_c) = t_m] \\ \mathbf{1}[T_\omega(z_8) = t_m] \\ \mathbf{1}[T_\omega(z_e) = t_m] \end{array} \tag{396}$$

$$\boldsymbol{B}_{t_l} = \begin{array}{c} \begin{matrix} \boldsymbol{s}_1 & \boldsymbol{s}_2 & \boldsymbol{s}_3 & \boldsymbol{s}_4 & \boldsymbol{s}_5 & \boldsymbol{s}_6 & \boldsymbol{s}_7 \end{matrix} \\ \begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 1 & 0 \end{bmatrix} \end{array} \begin{array}{l} \mathbf{1}[T_\omega(z_c) = t_l] \\ \mathbf{1}[T_\omega(z_8) = t_l] \\ \mathbf{1}[T_\omega(z_e) = t_l] \end{array} \tag{397}$$

The first row of each the binary matrices $\boldsymbol{B}_t; t \in \{t_h, t_m, t_l\}$ is associated with the instrumental value z_c; the second with z_8; and the last row with z_e. Each ω-th element of the vector $\mathbb{Z}[\omega, 1]$ may also be z_c, z_8, or z_e. Now $\mathbb{B}_t; t \in \{t_h, t_m, t_l\}$ are the $N \times 7$ binary matrices that stack z-rows of the 3×7 matrix $\boldsymbol{B}_t$ corresponding to the instrumental values in the $N \times 1$ vector $\mathbb{Z}$. Specifically, if $\mathbb{Z}[\omega, 1] = z_c$, then $\mathbb{B}_t[\omega, \cdot] = \boldsymbol{B}_{[}z_c, \cdot]$; if $\mathbb{Z}[\omega, 1] = z_8$, then $\mathbb{B}_t[\omega, \cdot] = \boldsymbol{B}_{[}z_8, \cdot]$; and if $\mathbb{Z}[\omega, 1] = z_e$, then $\mathbb{B}_t[\omega, \cdot] = \boldsymbol{B}_{[}z_e, \cdot]$. We can use the vector indicators $\mathbb{I}_z; z \in \{z_c, z_8, z_e\}$ to define binary matrices $\mathbb{B}_t; t \in \{t_h, t_m, t_l\}$ as following:

- Let $\mathbb{B}_{t_c}$ be the $N \times 7$ binary matrix defined by $\mathbb{B}_{t_h} \equiv [\mathbb{I}_{z_c}, \mathbb{I}_{z_8}, \mathbb{I}_{z_e}] \cdot \boldsymbol{B}_{t_h}$.
- Let $\mathbb{B}_{t_m}$ be the $N \times 7$ binary matrix defined by $\mathbb{B}_{t_m} \equiv [\mathbb{I}_{z_c}, \mathbb{I}_{z_8}, \mathbb{I}_{z_e}] \cdot \boldsymbol{B}_{t_m}$.
- Let $\mathbb{B}_{t_l}$ be the $N \times 7$ binary matrix defined by $\mathbb{B}_{t_h} \equiv [\mathbb{I}_{z_c}, \mathbb{I}_{z_8}, \mathbb{I}_{z_e}] \cdot \boldsymbol{B}_{t_l}$.

The equation below exemplifies the construction of binary matrices $\mathbb{B}_{t_h}, \mathbb{B}_{t_m}$ and $\mathbb{B}_{t_l}$:

$$\text{If } \mathbb{Z} = \begin{bmatrix} z_e \\ z_c \\ z_e \\ z_e \\ z_c \\ z_8 \\ z_c \\ z_c \\ z_8 \end{bmatrix}, \text{ then } \mathbb{B}_{t_h} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \mathbb{B}_{t_m} = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 & 1 \end{bmatrix}, \mathbb{B}_{t_l} = \begin{bmatrix} 0 & 0 & 1 & 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 0 \end{bmatrix}.$$

4. Stacked Data

- Let $\mathbb{D} = [\mathbb{D}'_{t_h}, \mathbb{D}'_{t_m}, \mathbb{D}'_{t_l}]'$ be the $3N \times 1$ binary vector that stacks the indicator vectors $\mathbb{D}_{t_h}, \mathbb{D}_{t_m}$ and $\mathbb{D}_{t_l}$.

- Let $\mathbb{B} = [\mathbb{B}'_{t_h}, \mathbb{B}'_{t_m}, \mathbb{B}'_{t_l}]'$ be the $3N \times 7$ binary matrix that stacks the binary matrices $\mathbb{B}_{t_h}, \mathbb{B}_{t_m}$ and $\mathbb{B}_{t_l}$.

- Let $\left(\mathbf{1}_{3,1} \otimes \mathbb{X}_0\right) = [\mathbb{X}'_0, \mathbb{X}'_0, \mathbb{X}'_0]'$ be the $3N \times K$ matrix that stacks the matrix of standardized baseline variables $\mathbb{X}_0$ three times. In this notation, $\otimes$ denotes the Kronecker multiplication and $\mathbf{1}_{3,1} \equiv [1, 1, 1]'$ is the 3×1 vector of elements one.

- Let $\left(\mathbf{1}_{3,1} \otimes \mathbb{X}_k\right) \odot \mathbb{D}$ stands for the $3N \times 1$ vector that stakes the vectors $\mathbb{X}_k \odot \mathbb{D}_{t_h}$, $\mathbb{X}_k \odot \mathbb{D}_{t_m}$, and $\mathbb{X}_k \odot \mathbb{D}_{t_l}$.

O.3 Estimation of Response-type Probabilities

Response-type probabilities are estimated by a weighted OLS regression that controls for baseline variables in $\mathbb{X}_0$. The components of this regression are given by:

1. The dependent variable is given by the vector $\mathbb{D} \equiv [\mathbb{D}'_{t_h}, \mathbb{D}'_{t_m}, \mathbb{D}'_{t_l}]'$ whose dimension is $3N \times 1$.

2. The first set of regressors is given by the binary matrices $\mathbb{B} = [\mathbb{B}'_{t_h}, \mathbb{B}'_{t_m}, \mathbb{B}'_{t_l}]'$ whose dimension is $3N \times 7$.

3. The second set of regressors is the stacked matrix of K standardized baseline variables $\left(\mathbf{1}_{3,1} \otimes \mathbb{X}_0\right)$ whose dimension is $3N \times K$.

The regression can be expressed in matrix notation as:

$$\mathbb{D} = \mathbb{B}\boldsymbol{\beta} + \left(\mathbf{1}_{3,1} \otimes \mathbb{X}_0\right)\boldsymbol{\gamma} + \boldsymbol{\epsilon} \tag{398}$$

The estimator for the response-type probabilities is given by $\widehat{\boldsymbol{\beta}}$ of the following weighted estimator:

$$\begin{pmatrix} \widehat{\boldsymbol{\beta}} \\ \widehat{\boldsymbol{\gamma}} \end{pmatrix} = \Big(\left[\mathbb{B}\,,\ \left(\mathbf{1}_{3,1} \otimes \mathbb{X}_0\right)\right]' \Big(\mathbf{1}_{3,1} \otimes \text{diag}(\mathbb{W})\Big) \left[\mathbb{B}\,,\ \left(\mathbf{1}_{3,1} \otimes \mathbb{X}_0\right)\right] \Big)^{-1} \left[\mathbb{B}\,,\ \left(\mathbf{1}_{3,1} \otimes \mathbb{X}_0\right)\right]' \Big(\mathbf{1}_{3,1} \otimes \text{diag}(\mathbb{W})\Big)\mathbb{D}, \tag{399}$$

where $\mathbb{W}$ is the $N \times 1$ vector of weights. Note that the regression does not have an intercept as it would be collinear to the columns of matrix $\mathbb{B}$. Moreover, each of the columns of the matrix of baseline variables $\mathbb{X}_0$ must have zero mean. This is necessary to guarantee that the sum of the estimates of response-types probabilities will sum to one. Inference can be obtained by the method of bootstrap.

In practice, regression (398) can be evaluated by the following steps:

1. Generate the choice indicators $\mathbb{D}_{t_h}, \mathbb{D}_{t_m}, \mathbb{D}_{t_l}$.

2. Generate the Binary matrices $\mathbb{B}_{t_h}, \mathbb{B}_{t_m}, \mathbb{B}_{t_l}$.

3. Standardized baseline the baseline variables to obtain $\mathbb{X}_0$.

4. Stack $\mathbb{D}_{t_h}, \mathbb{D}_{t_m}, \mathbb{D}_{t_l}$ to generate $\mathbb{D}$.

5. Stack $\mathbb{B}_{t_h}, \mathbb{B}_{t_m}, \mathbb{B}_{t_l}$ to generate $\mathbb{B}$.

6. Stack $\mathbb{X}_0$ and the weights $\boldsymbol{W}$ three times.

7. Evaluate a weighted regression of $\boldsymbol{D}$ on $\mathbb{B}$ and the stacked $\mathbb{X}_0$.

8. Response-type probabilities are the estimates associated with $\mathbb{B}$.

O.4 Estimating Pre-program Variable Means Conditioned on Response-types

The estimation of pre-program variable means stems from the estimation result of Appendix A.13. Pre-program Variable Means are estimated by a weighted OLS regression that controls for site of intervention. Let $\mathbb{X}_k$ be the $N \times 1$ vector of pre-program variable we are interested in. The dependent variable of the regression is the stacked vectors of the element-wise multiplication of the pre-program variable $\mathbb{X}_k$ and each of the choice indicators $\mathbb{D}_t; t \in \{t_h, t_m, t_l\}$.

There are two sets of regressors. The first set of regressors is generated upon the binary matrices $\mathbb{B}_t; t \in \{t_h, t_m, t_l\}$ and the estimated response-type probabilities. Specifically, let $\boldsymbol{P}_S$ be the 7×1 vector of response-type probabilities estimates described in the previous section. Let $\mathbf{1}_{N,1}\boldsymbol{P}_S'$ be the $N \times 7$ matrix whose columns are constant and take the value of the estimates of the response-type probabilities. Each matrix $\mathbb{B}_t; t \in \{t_h, t_m, t_l\}$ has dimension $N \times 7$. Thus $\mathbb{B}_t \odot \left(\mathbf{1}_{N,1}\boldsymbol{P}_S'\right)$ stands for the matrix generated by multiplying each of the seven columns of $\mathbb{B}_t$ by its respective estimate of response-type probability. The first regressor is obtained by stacking these matrices across the choices t_h, t_m, t_l, that is:

$$\mathbb{B} \odot \mathbf{1}_{3N,1}\boldsymbol{P}_S' = [\mathbb{B}_{t_h} \odot \left(\mathbf{1}_{N,1}\boldsymbol{P}_S'\right) \, , \, \mathbb{B}_{t_m} \odot \left(\mathbf{1}_{N,1}\boldsymbol{P}_S'\right) \, , \,]\mathbb{B}_{t_l} \odot \left(\mathbf{1}_{N,1}\boldsymbol{P}_S'\right)].$$

The second set of estimator is given by the stacking the standardised matrix of site indicators $\mathbb{S}_0$, that is $\left(\mathbf{1}_{3,1} \odot \mathbb{S}_0\right)$.

In summary, the components of the regression are:

1. The dependent variable is given by the vector:
$$\left(\mathbf{1}_{3,1} \otimes \mathbb{X}_k\right) \odot \mathbb{D} \equiv [\mathbb{X}_k \odot \mathbb{D}_{t_h}' \, , \, \mathbb{X}_k \odot \mathbb{D}_{t_m}' \, , \, \mathbb{X}_k \odot \mathbb{D}_{t_l}']'$$
whose dimension is $3N \times 1$.

2. The first set of regressors is given by the matrix $\mathbb{B} \odot \mathbf{1}_{3N,1}\boldsymbol{P}_S'$, which is obtained by multiplying each column of the binary matrices $\mathbb{B}_{t_h}, \mathbb{B}_{t_m}, \mathbb{B}_{t_h}$ by the response-type estimates and then stacking those. This procedure generates a matrix that has dimension $3N \times 7$.

3. The second set of regressors is the stacked matrix of standardized site indicators $\left(\mathbf{1}_{3,1} \otimes \mathbb{S}_0\right)$ whose dimension is $3N \times 4$.

The regression can be expressed in matrix notation as:

$$\left(\mathbf{1}_{3,1} \otimes \mathbb{X}_k\right) \odot \mathbb{D} = \left(\mathbb{B} \odot \mathbf{1}_{3N,1}\boldsymbol{P}_S'\right)\boldsymbol{\beta} + \left(\mathbf{1}_{3,1} \otimes \mathbb{S}_0\right)\boldsymbol{\gamma} + \boldsymbol{\epsilon} \tag{400}$$

Thus, the estimator for the mean of the pre-program variable $\mathbb{X}_k$ is given by $\widehat{\boldsymbol{\beta}}$ of the following weighted estimator:

$$\begin{pmatrix} \widehat{\boldsymbol{\beta}} \\ \widehat{\boldsymbol{\gamma}} \end{pmatrix} = \Big(\big[(\mathbb{B} \odot \mathbf{1}_{3N,1} \boldsymbol{P}_S') \, , \, (\mathbf{1}_{3,1} \otimes \mathbb{S}_0) \big]' \Big(\mathbf{1}_{3,1} \otimes \text{diag}(\mathbb{W}) \Big) \big[(\mathbb{B} \odot \mathbf{1}_{3N,1} \boldsymbol{P}_S') \, , \, (\mathbf{1}_{3,1} \otimes \mathbb{S}_0) \big] \Big)^{-1} \cdot \tag{401}$$

$$\big[(\mathbb{B} \odot \mathbf{1}_{3N,1} \boldsymbol{P}_S') \, , \, (\mathbf{1}_{3,1} \otimes \mathbb{S}_0) \big]' \Big(\mathbf{1}_{3,1} \otimes \text{diag}(\mathbb{W}) \Big) (\mathbf{1}_{3,1} \otimes \mathbb{X}_k) \odot \mathbb{D}, \tag{402}$$

where $\mathbb{W}$ is the $N \times 1$ vector of weights. As mentioned, the regression does not have an intercept as it would be collinear to the columns of matrix $\mathbb{B}$. Moreover, each of the columns of the matrix of the standardized site indicators $\mathbb{S}_0$ must have zero mean. In practice, regression 400 can be evaluated by the following steps:

1. Generate the choice indicators $\mathbb{D}_{t_h}, \mathbb{D}_{t_m}, \mathbb{D}_{t_l}$.
2. Multiple each element of these indicators by the respective element in the pre-program variable $\mathbb{X}_k$, that is, $\mathbb{X}_k \odot \mathbb{D}_{t_h}$, $\mathbb{X}_k \odot \mathbb{D}_{t_m}$, and $\mathbb{X}_k \odot \mathbb{D}_{t_l}$.
3. Stack $\mathbb{X}_k \odot \mathbb{D}_{t_h}, \mathbb{X}_k \odot \mathbb{D}_{t_m}, \mathbb{X}_k \odot \mathbb{D}_{t_l}$ to generate the dependent variables given by:
$$\big(\mathbf{1}_{3,1} \otimes \mathbb{X}_k \big) \odot \mathbb{D} = [\mathbb{X}_k \odot \mathbb{D}_{t_h}, \mathbb{X}_k \odot \mathbb{D}_{t_m}, \mathbb{X}_k \odot \mathbb{D}_{t_l}]. \tag{403}$$
4. Generate the Binary matrices $\mathbb{B}_{t_h}, \mathbb{B}_{t_m}, \mathbb{B}_{t_l}$.
5. Estimate the response-type probabilities $\boldsymbol{P}_S$ described in the previous section.
6. Multiply each of the 7 columns of $\mathbb{B}_t; t \in \{t_h, t_m, t_l\}$ by its respective estimate of the response-type probability, namely $\mathbb{B}_t \odot \big(\mathbf{1}_{N,1} \boldsymbol{P}_S' \big); t \in \{t_h, t_m, t_l\}$.
7. Stack the matrices $\mathbb{B}_t \odot \big(\mathbf{1}_{N,1} \boldsymbol{P}_S' \big); t \in \{t_h, t_m, t_l\}$ to generate the first set of regressors given by:
$$\mathbb{B} \odot \big(\mathbf{1}_{3N,1} \otimes \boldsymbol{P}_S \big) = [\mathbb{B}_{t_h} \odot \big(\mathbf{1}_{N,1} \boldsymbol{P}_S' \big) \, , \, \mathbb{B}_{t_m} \odot \big(\mathbf{1}_{N,1} \boldsymbol{P}_S' \big) \, , \, \mathbb{B}_{t_l} \odot \big(\mathbf{1}_{N,1} \boldsymbol{P}_S' \big)]. \tag{404}$$
8. Standardized the 4 site indicators to obtain $\mathbb{S}_0$.
9. Stack $\mathbb{S}_0$ and the weights $\boldsymbol{W}$ three times, that is $\mathbf{1}_{3,1} \otimes \mathbb{S}_0$.
10. Evaluate a weighted regression that sets the dependent variable as $\big(\mathbf{1}_{3,1} \otimes \mathbb{X}_k \big) \odot \mathbb{D}$ in (403), the regressors as $\mathbb{B} \odot \big(\mathbf{1}_{3N,1} \otimes \boldsymbol{P}_S \big)$ in (404) in addition to the standardized site indicators $\big(\mathbf{1}_{3,1} \otimes \mathbb{S}_0 \big)$ using the stacked weighting vectors $\boldsymbol{W}$ as weights.
11. Pre-program variable means are the 7×1 estimates associated with the regressor $\mathbb{B} \odot \big(\mathbf{1}_{3N,1} \otimes \boldsymbol{P}_S \big)$ in (404).

O.5 Estimating Counterfactual Outcomes Conditioned on Response-types

Theorem **T-4** explains that each counterfactual outcome mean that is nonparametrically identified can be obtained by a 2SLS regression such that:

1. The dependent variable is given by the multiplication of the outcome of interest Y and a choice indicator $\mathbf{1}[T=t]; t \in \{t_h, t_m, t_l\}$.

2. The endogenous variable is given by the choice $\mathbf{1}[T=t]$.

3. The instrument consists of the two instrumental variable indicators $\mathbf{1}[Z=z], \mathbf{1}[Z=z']$, where $(z, z') \in \{(z_c, z_8), (z_c, z_e), (z_8, z_e)\}$.

Table 9 maps the 2SLS with its respective counterfactual outcome estimate.

Let matrix $[\mathbb{I}_z, \mathbb{I}_{z'}]$ stands for the instrumental variable indicators. Let $\mathbb{W}$ be the vector of positive weights with dimension $N \times 1$ and let $\text{diag}(\mathbb{W})^{1/2}$ be the $N \times N$ diagonal matrix whose diagonal elements are the squared root of the weights. It is useful to define $\widetilde{\mathbb{Z}} \equiv \text{diag}(\mathbb{W})^{1/2}[\mathbb{I}_z, \mathbb{I}_{z'}]$ as the weighted version of the instrumental variable matrix. Let $\mathbb{X}_0$ be the $N \times K$ standardized matrix of baseline variable we wish to control. Each column of $\mathbb{X}_0$ must have zero mean. Let $\widetilde{\mathbb{X}}_0 \equiv \text{diag}(\mathbb{W})^{1/2}\mathbb{X}_0$ be the weighted version of the standardized baseline variables. Let $\mathbb{D}_t$ be the vector of choice indicator for $t \in \{t_h, t_m, t_l\}$ and $\widetilde{\mathbb{D}}_t \equiv \text{diag}(\mathbb{W})^{1/2}\mathbb{D}_t$ be its weighted version. Let $\mathbb{D}_t \odot \mathbb{Y}$ be the $N \times 1$ vector that stands for our dependent variable and let $\widetilde{\mathbb{Y}} \equiv \text{diag}(\mathbb{W})^{1/2}\big(\mathbb{D}_t \odot \mathbb{Y}\big)$ be its weighted version.

In this notation, the projection on the space generated by the columns of $[\widetilde{\mathbb{Z}}, \widetilde{\mathbb{X}}_0]$ is given by:

$$\mathbb{P}_{\widetilde{\mathbb{Z}}, \widetilde{\mathbb{X}}_0} = [\widetilde{\mathbb{Z}}, \widetilde{\mathbb{X}}_0]\Big([\widetilde{\mathbb{Z}}, \widetilde{\mathbb{X}}_0]'[\widetilde{\mathbb{Z}}, \widetilde{\mathbb{X}}_0]\Big)^{-1}[\widetilde{\mathbb{Z}}, \widetilde{\mathbb{X}}_0]'.$$

Let the second stage of the 2SLS regression can be represented in matrix notation by:

$$\widetilde{\mathbb{Y}} = \widetilde{\mathbb{D}}_t\boldsymbol{\beta} + \widetilde{\mathbb{X}}_0\boldsymbol{\gamma} + \mathbf{1}_{N,1}\kappa + \boldsymbol{\epsilon}. \tag{405}$$

The 2SLS estimator is therefore given by:

$$\begin{pmatrix} \widehat{\boldsymbol{\beta}} \\ \widehat{\boldsymbol{\gamma}} \\ \widehat{\kappa} \end{pmatrix} = \left(\Big[\widetilde{\mathbb{D}}_t\,,\, \widetilde{\mathbb{X}}_0\,,\, \mathbf{1}_{N,1}\Big]' \mathbb{P}_{\widetilde{\mathbb{Z}}, \widetilde{\mathbb{X}}_0} \Big[\widetilde{\mathbb{D}}_t\,,\, \widetilde{\mathbb{X}}_0\,,\, \mathbf{1}_{N,1}\Big]\right)^{-1} \left(\Big[\widetilde{\mathbb{D}}_t\,,\, \widetilde{\mathbb{X}}_0\,,\, \mathbf{1}_{N,1}\Big]' \mathbb{P}_{\widetilde{\mathbb{Z}}, \widetilde{\mathbb{X}}_0} \widetilde{\mathbb{Y}}\right), \tag{406}$$

where $\widehat{\boldsymbol{\beta}}$ is the estimate of the counterfactual outcome mean $E(Y(t)|\boldsymbol{S} \in \Sigma_t(z) \oplus \Sigma_t(z'))$ described by Table 9. Note that the first stage of the 2SLS does not have an intercept while the second stage has. Inference can be obtained by the method of bootstrap.

In practice, the 2SLS in (406) can be obtained by the following steps:

1. Select a choices $t \in \{t_h, t_m, t_l\}$ and generate the choice indicator $\mathbb{D}_t$.

2. Multiply the outcome of interest $\mathbb{Y}$ by the indicator $\mathbb{D}_t$, that is $\mathbb{Y} \odot \mathbb{D}_t$.

3. Select two instrumental values $(z, z') \in \{(z_c, z_8), (z_c, z_e), (z_8, z_e)\}$ and generate the two IV indicators $[\mathbb{I}_z, \mathbb{I}_{z'}]$.

4. Standardized baseline the baseline variables to obtain $\mathbb{X}_0$.

5. Perform a weighted 2SLS of $\mathbb{Y} \odot \mathbb{D}_t$ on $\mathbb{D}_t, \mathbb{X}_0$ using the IV indicators as instrumental variables and $\mathbb{W}$ as weights.

P Estimating Counterfactual Outcomes of T-6 Using 2SLS

Theorem **T-4** in Section 8.1 states that each counterfactual outcome mean that is identified can be estimate by a 2SLS procedure. Specifically, the counterfactual mean $E(Y(t)|\boldsymbol{S} \in \Sigma_t(z) \oplus \Sigma_t(z'))$ is identified (see Lemma **L-6**) and can be estimated by the 2SLS that uses $Y \cdot D_t$ as dependent variable, $D_t$as endogenous variable and two indicators $\mathbf{1}[Z = z], \mathbf{1}[Z = z']$ as instrumental variables. Theorem **T-6** in Section 9.4 states that these counterfactual outcomes can be also evaluated by a ratio of OLS estimates.

This section presents a method that evaluates the ratio of Theorem **T-6** as a new 2SLS regression. The regression differs from the one in **T-4** as it is based on a more complex transformation of the instrumental variable. Nevertheless, the new 2SLS regression generates the same numerical estimates for all identified counterfactual outcome means. The new 2SLS estimator is not to meant to replace the 2SLS method of Theorem **T-4**. Instead, the benefit of the new estimator is to aid on the evaluation of counterfactual outcomes that are partially identified.

P.1 Defining Basic Notation

This section follows the notation of Appendix A.15. Let the observed data be indexed by the set $\Omega = \{1, ..., N\}$ and the support of the instrumental variable Z be given by $\text{supp}(Z) = \{z_1, ..., z_{N_Z}\}$. Theorem **T-6** is based on functions $H_t(z); z \in \text{supp}(Z), t \in \text{supp}(T)$ and the vector-valued function $\boldsymbol{\lambda}$ that has dimension $N_Z \times 1$. Thus consider the following notation for each $t \in \text{supp}(T)$:

1. Let $H_t(z_i); i = 1, ..., N_Z$ be the values that function $H_t(z)$ takes across the values $z \in \text{supp}(Z)$. A common choice for function $H_t(z)$ is the propensity score. Thus, for sake of clarity, $H_t(z)$ are set to be the propensity score $H_t(z) = P(T = t|Z = z) \equiv P_t(z)$ henceforward. The unconditional the propensity score probability can be nonparametrically estimated by:
$$P_t(z) = \frac{\sum_{\omega=1}^{N} \mathbf{1}[T_\omega = t]\mathbf{1}[Z_\omega = z]}{\sum_{\omega=1}^{N} \mathbf{1}[Z_\omega = z]}. \tag{407}$$

2. Let $\boldsymbol{\lambda}(x) = [\lambda_1(x), ..., \lambda_{N_Z}(x)]'; i = 1, ..., N_Z$ be the $N_Z \times 1$ vector. A common choice for the vector-valued function $\boldsymbol{\lambda}(x)$ is a N_Z-polynomial transformation of the propensity score. In the case of MTO, $N_Z = 3$ and the vector $\boldsymbol{\lambda}(x)$ is set to $\boldsymbol{\lambda}(x) = [\lambda_1(x), \lambda_2(x), \lambda_3(x)]$, where $\lambda_n(x) = x^{n^{-1}}; x \in \mathbb{R}, n \in \mathbb{N}$.

3. Let $\boldsymbol{M}_t = [\boldsymbol{\lambda}(P_t(z_1)), ..., \boldsymbol{\lambda}(P_t(z_{N_Z}))]'$ be the $N_Z \times N_Z$ matrix generated by the rectangular array of vectors $\boldsymbol{\lambda}(P_t(z_{N_Z})); i = 1, ..., N_Z$. The i-th row of matrix $\boldsymbol{M}$ stands for the transpose of the vector $\boldsymbol{\lambda}(P_t(z_1))$.

4. Let $\boldsymbol{\iota}_z$ be the $N_Z \times 1$ vector that is associated with the instrumental values and takes value 1 for the element associated with value z and zero otherwise. In the case of MTO, the order of the IV is (z_c, z_8, z_e) therefore we have that $\boldsymbol{\iota}_{z_c} = [1, 0, 0]'$, $\boldsymbol{\iota}_{z_8} = [0, 1, 0]'$, $\boldsymbol{\iota}_{z_e} = [0, 0, 1]'$.

5. Each agent $w \in \Omega \equiv \{1, \ldots, N\}$, is assigned to an instrumental value $Z_\omega \in \text{supp}(Z)$. Let $P_{t,\omega} \equiv P_t(Z_\omega)$ denotes the propensity score for choice $t \in \text{supp}(T)$ assigned to agent $w \in \Omega$. Following this notation, let $\boldsymbol{\lambda}_{t,\omega} \equiv \boldsymbol{\lambda}(P_{t,\omega})$ be the vector of functions of the propensity score assigned to agent ω. This vector will play the role of covariates in the 2SLS regression.

6. Let $\mathbb{M}_t = [\boldsymbol{\lambda}_{t,\omega}; \omega \in \Omega]$ be the $N \times N_Z$ matrix that stacks each agent $w \in \Omega \equiv \{1, \ldots, N\}$, is assigned to an instrumental value $Z_\omega \in \text{supp}(Z)$. In the case of MTO, we have the matrices $\mathbb{M}_{t_h}, \mathbb{M}_{t_m}, \mathbb{M}_{t_l}$ each of them consists of a $N \times 3$ matrix whose ω-th row is given by $[\lambda_1(P_{t,\omega}), \lambda_2(P_{t,\omega}), \lambda_3(P_{t,\omega})]$ for choice $t \in \{t_h, t_m, t_l\}$.

7. Let $\mathbb{I}_z = \big[\mathbf{1}[Z_\omega = z; \omega \in \Omega]\big]$ be the $N \times 1$ vector that stacks the indicator if Z_ω takes the value $z \in \text{supp}(Z)$ across agents $\omega \in \Omega$.

8. Let $\mathbb{I} = [\mathbb{I}_{z_1}, ..., \mathbb{I}_{z_{N_Z}}]$ be the $N \times N_Z$ matrix generated by the rectangular array of vectors $\mathbb{I}_z; z \in \text{supp}(Z)$. In the case of MTO, $\mathbb{I}$ has dimension $N \times 3$ and is given by $\mathbb{I} = [\mathbb{I}_{z_c}, \mathbb{I}_{z_8}, \mathbb{I}_{z_e}]$.

9. In this notation, matrix $\mathbb{M}_t$ can be expressed as $\mathbb{M}_t = \mathbb{I}\boldsymbol{M}_t$.

10. Let $\mathbb{P}_{\mathbb{M}_t} = \mathbb{M}_t\Big(\mathbb{M}_t'\mathbb{M}_t\Big)^{-1}\mathbb{M}_t'$ be the projection onto the columns of $\mathbb{M}_t$.

11. Let $\widetilde{\mathbb{I}}_z = \mathbb{P}_{\mathbb{M}_t}\mathbb{I}_z$ be the prediction of the instrumental value indicator $\mathbb{I}_z$ by the matrix of covariates $\mathbb{M}_t$.

12. Let $\mathbb{Y}$ be the $N \times 1$ vector of observed outcomes, that is $\mathbb{Y} = [Y_\omega; \omega \in \Omega]$.

13. Let $\mathbb{D}_t$ be the $N \times 1$ vector that indicates if the treatment choice across agents ω is equal to $t \in \text{supp}(T)$, that is $\mathbb{D}_t = [\mathbf{T}_\omega = \mathbf{t}; \omega \in \Omega]$.

14. Let $P_z \equiv P(Z = z); z \in \text{supp}(Z)$ be the probability that the instrumental variable Z takes value z. The unconditional probability can be nonparametrically estimated by:

$$P_z = \frac{\sum_{\omega=1}^{N} \mathbf{1}[Z_\omega = z]\mathbf{1}[Z_\omega = z]}{N}. \tag{408}$$

P.2 A new 2SLS Regression to Evaluate Counterfactual Outcome Means in T-4

Theorem T-6 states that the counterfactual outcome $\Lambda_t(z, z') \equiv E(Y(t)|\boldsymbol{S} \in \Sigma_t(z) \oplus \Sigma_t(z'))$ can be estimated by:

$$\widehat{\Lambda}_t(z, z') = \frac{\Big(\boldsymbol{\lambda}(P_t(z)) - \boldsymbol{\lambda}(P_t(z'))\Big)'\widehat{\boldsymbol{\beta}}_t}{\Big(\boldsymbol{\lambda}(P_t(z)) - \boldsymbol{\lambda}(P_t(z'))\Big)'\widehat{\boldsymbol{\theta}}_t}, \tag{409}$$

$$\tag{410}$$

where parameters $\widehat{\boldsymbol{\beta}}_t$ and $\widehat{\boldsymbol{\theta}}_t$ are given by:

$$\widehat{\boldsymbol{\beta}}_t = \Big(\mathbb{M}_t'\mathbb{M}_t\Big)^{-1}\mathbb{M}_t'\Big(\mathbb{D}_t \odot \mathbb{Y}\Big), \tag{411}$$

$$\widehat{\boldsymbol{\theta}}_t = \Big(\mathbb{M}_t'\mathbb{M}_t\Big)^{-1}\mathbb{M}_t'\mathbb{D}_t. \tag{412}$$

It is useful to express the parameter $\widehat{\Lambda}_t(z,z')$ in (409) as following:

$$\widehat{\Lambda}_t(z,z') = \frac{\left(\boldsymbol{\iota}_z - \boldsymbol{\iota}_{z'}\right)' \boldsymbol{M}_t \widehat{\boldsymbol{\beta}}_t}{\left(\boldsymbol{\iota}_z - \boldsymbol{\iota}_{z'}\right)' \boldsymbol{M}_t \widehat{\boldsymbol{\theta}}_t} \tag{413}$$

The goal of the section is to show that estimator 413 can be evaluated by a 2SLS regression based on a trans formation of the instrumental variable Z. Specifically let $\widetilde{\mathbb{Z}}_{z,z'}$ be a vector of a transformation of the instrumental variable Z defined by:

$$\widetilde{\mathbb{Z}}_{z,z'} \equiv \widetilde{\mathbb{I}}_z \cdot (1/P_z) - \widetilde{\mathbb{I}}_{z'} \cdot (1/P_{z'}) \tag{414}$$

The transformed IV $\widetilde{\mathbb{Z}}_{z,z'}$ in (414) is a $N \times 1$ vector generated by the an inverted probability weighed difference between the predictions $\widetilde{\mathbb{I}}_z, \widetilde{\mathbb{I}}_{z'}$. Weights are given by the inverse of the instrumental variable probabilities $P_z, P_{z'}$ and $\widetilde{\mathbb{I}}_z, \widetilde{\mathbb{I}}_{z'}$ are the fitted values of the instrumental variable indicators $\mathbb{I}_z, \mathbb{I}_{z'}$ that are projected in the space generated by the columns on the matrix $\mathbb{M}_t$, namely $\widetilde{\mathbb{I}}_z = \mathbb{P}_{\mathbb{M}_t}\mathbb{I}_z$.

We claim that 2SLS regression that uses $\widetilde{\mathbb{Z}}_{z,z'}$ as the vector of instrumental variable, $\mathbb{D}_t$ as the endogenous indicator and $\mathbb{D}_t \odot \mathbb{Y}$ and the dependent outcome produces an IV estimate that is numerically the same as the estimator $\widehat{\Lambda}_t(z,z')$ in (409). The following lemmas are useful to show that this statement holds.

Lemma L-16. The vector of transformed instrumental variable $\widetilde{\mathbb{Z}}_{z,z'}$ can be expressed as:

$$\widetilde{\mathbb{Z}}'_{z,z'} = N\left(\boldsymbol{\iota}_z - \boldsymbol{\iota}_{z'}\right)' \boldsymbol{M}_t \left(\mathbb{M}_t'\mathbb{M}_t\right)^{-1} \mathbb{M}_t'.$$

Proof.

$$\begin{aligned}
\widetilde{\mathbb{Z}}'_{z,z'} &= \left(\widetilde{\mathbb{I}}_z \cdot (1/P_z) - \widetilde{\mathbb{I}}_{z'} \cdot (1/P_{z'})\right)' \\
&= \left(\mathbb{P}_{\mathbb{M}_t}[\mathbb{I}_z, \mathbb{I}_{z'}][1/P_z, 1/P_{z'}]'\right)' \\
&= \left(\mathbb{M}_t\left(\mathbb{M}_t'\mathbb{M}_t\right)^{-1}\mathbb{M}_t'[\mathbb{I}_z, \mathbb{I}_{z'}][1/P_z, 1/P_{z'}]'\right)' \\
&= [1/P_z, 1/P_{z'}][\mathbb{I}_z, -\mathbb{I}_{z'}]'\mathbb{M}_t\left(\mathbb{M}_t'\mathbb{M}_t\right)^{-1}\mathbb{M}_t' \\
&= [1/P_z, 1/P_{z'}]\left([\mathbb{I}_z, -\mathbb{I}_{z'}]'\mathbb{M}_t\right)\left(\mathbb{M}_t'\mathbb{M}_t\right)^{-1}\mathbb{M}_t' \\
&= [1/P_z, 1/P_{z'}]\left([\mathbb{I}_z'\mathbb{M}_t, -\mathbb{I}_{z'}\mathbb{M}_t]'\right)\left(\mathbb{M}_t'\mathbb{M}_t\right)^{-1}\mathbb{M}_t' \\
&= [1/P_z, 1/P_{z'}]\left[(N \cdot P_z)\boldsymbol{\iota}_z\boldsymbol{M}_t, -(N \cdot P_{z'})\boldsymbol{\iota}_{z'}\boldsymbol{M}_t\right]'\left(\mathbb{M}_t'\mathbb{M}_t\right)^{-1}\mathbb{M}_t' \\
&= \left[\left(\frac{N \cdot P_z}{P_z}\right)\boldsymbol{\iota}_z\boldsymbol{M}_t - \left(\frac{N \cdot P_{z'}}{P_{z'}}\right)\boldsymbol{\iota}_{z'}\boldsymbol{M}_t\right]'\left(\mathbb{M}_t'\mathbb{M}_t\right)^{-1}\mathbb{M}_t' \\
&= N\left(\boldsymbol{\iota}_z - \boldsymbol{\iota}_{z'}\right)'\boldsymbol{M}_t\left(\mathbb{M}_t'\mathbb{M}_t\right)^{-1}\mathbb{M}_t'.
\end{aligned}$$

□

As mentioned, consider the 2SLS estimator $\widetilde{\Lambda}_t(z,z')$ that uses $\widetilde{\mathbb{Z}}_{z,z'}$ as the vector of instrumental

variable, $\mathbb{D}_t$ as the endogenous indicator and $\mathbb{D}_t \odot \mathbb{Y}$ and the dependent outcome produces an IV estimate that is numerically the same as the estimator $\widehat{\Lambda}_t(z, z')$. This IV estimator is expressed in matrix form by:

$$\widetilde{\Lambda}_t(z, z') = \left(\mathbb{D}_t' \mathbb{P}_{\widetilde{\mathbb{Z}}_{z,z'}} \mathbb{D}_t\right)^{-1} \mathbb{D}_t' \mathbb{P}_{\widetilde{\mathbb{Z}}_{z,z'}} \left(\mathbb{D}_t \odot \mathbb{Y}\right) \tag{415}$$

$$\text{where } \mathbb{P}_{\widetilde{\mathbb{Z}}_{z,z'}} = \widetilde{\mathbb{Z}}_{z,z'} \left(\widetilde{\mathbb{Z}}_{z,z'}' \widetilde{\mathbb{Z}}_{z,z'}\right)^{-1} \widetilde{\mathbb{Z}}_{z,z'}'. \tag{416}$$

Next lemma shows that $\widetilde{\Lambda}_t(z, z')$ in (415) and $\widehat{\Lambda}_t(z, z')$ in (409) produce the same estimates.

Lemma L-17. $\widetilde{\Lambda}_t(z, z')$ in (415) and $\widehat{\Lambda}_t(z, z')$ in (409) are numerically identical.

Proof. According to equation (413), we have that

$$\widehat{\Lambda}_t(z, z') = \frac{\left(\boldsymbol{\iota}_z - \boldsymbol{\iota}_{z'}\right)' \boldsymbol{M}_t \widehat{\boldsymbol{\beta}}_t}{\left(\boldsymbol{\iota}_z - \boldsymbol{\iota}_{z'}\right)' \boldsymbol{M}_t \widehat{\boldsymbol{\theta}}_t}.$$

Thus it suffices to prove that:

$$\left(\mathbb{D}_t' \mathbb{P}_{\widetilde{\mathbb{Z}}_{z,z'}} \mathbb{D}_t\right)^{-1} \mathbb{D}_t' \mathbb{P}_{\widetilde{\mathbb{Z}}_{z,z'}} \left(\mathbb{D}_t \odot \mathbb{Y}\right) = \frac{\left(\boldsymbol{\iota}_z - \boldsymbol{\iota}_{z'}\right)' \boldsymbol{M}_t \widehat{\boldsymbol{\beta}}_t}{\left(\boldsymbol{\iota}_z - \boldsymbol{\iota}_{z'}\right)' \boldsymbol{M}_t \widehat{\boldsymbol{\theta}}_t}.$$

The proof is obtained by the following sequence of linear algebra calculations:

$$\begin{aligned}
\widetilde{\Lambda}_t(z, z') &= \left(\mathbb{D}_t' \widetilde{\mathbb{Z}}_{z,z'} \left(\widetilde{\mathbb{Z}}_{z,z'}' \widetilde{\mathbb{Z}}_{z,z'}\right)^{-1} \widetilde{\mathbb{Z}}_{z,z'}' \mathbb{D}_t\right)^{-1} \mathbb{D}_t' \widetilde{\mathbb{Z}}_{z,z'} \left(\widetilde{\mathbb{Z}}_{z,z'}' \widetilde{\mathbb{Z}}_{z,z'}\right)^{-1} \widetilde{\mathbb{Z}}_{z,z'}' \left(\mathbb{D}_t \odot \mathbb{Y}\right) \\
&= \left(\mathbb{D}_t' \widetilde{\mathbb{Z}}_{z,z'} \widetilde{\mathbb{Z}}_{z,z'}' \mathbb{D}_t\right)^{-1} \mathbb{D}_t' \widetilde{\mathbb{Z}}_{z,z'} \widetilde{\mathbb{Z}}_{z,z'}' \left(\mathbb{D}_t \odot \mathbb{Y}\right) \\
&= \left(\widetilde{\mathbb{Z}}_{z,z'}' \mathbb{D}_t\right)^{-1} \widetilde{\mathbb{Z}}_{z,z'}' \left(\mathbb{D}_t \odot \mathbb{Y}\right) \\
&= \frac{\widetilde{\mathbb{Z}}_{z,z'}' \left(\mathbb{D}_t \odot \mathbb{Y}\right)}{\widetilde{\mathbb{Z}}_{z,z'}' \mathbb{D}_t} \\
&= \frac{N \left(\boldsymbol{\iota}_z - \boldsymbol{\iota}_{z'}\right)' \boldsymbol{M}_t \left(\mathbb{M}_t' \mathbb{M}_t\right)^{-1} \mathbb{M}_t' \left(\mathbb{D}_t \odot \mathbb{Y}\right)}{N \left(\boldsymbol{\iota}_z - \boldsymbol{\iota}_{z'}\right)' \boldsymbol{M}_t \left(\mathbb{M}_t' \mathbb{M}_t\right)^{-1} \mathbb{M}_t' \mathbb{D}_t} \\
&= \frac{\left(\boldsymbol{\iota}_z - \boldsymbol{\iota}_{z'}\right)' \boldsymbol{M}_t \widehat{\boldsymbol{\beta}}_t}{\left(\boldsymbol{\iota}_z - \boldsymbol{\iota}_{z'}\right)' \boldsymbol{M}_t \widehat{\boldsymbol{\theta}}_t} \\
&= \widehat{\Lambda}_t(z, z').
\end{aligned}$$

The first equality is due to (416). The second equality uses the fact that $\widetilde{\mathbb{Z}}_{z,z'}$ is a vector, thereby $\widetilde{\mathbb{Z}}_{z,z'}' \widetilde{\mathbb{Z}}_{z,z'}$ is a scalar and can be eliminated. The third equality uses the fact that $\mathbb{D}_t' \widetilde{\mathbb{Z}}_{z,z'}$ is a scalar and can be eliminated. The fourth equality uses the fact that $\widetilde{\mathbb{Z}}_{z,z'}' \mathbb{D}_t$ is a scalar, thereby the matrix inversion can be expressed as a ratio of matrix multiplications. The fifth equality applies Lemma **L-16** twice. The six equality eliminates the sample size N in the numerator and denominator of the ratio. The last equality is due to (413). □

Theorem **T-7** uses Lemma **L-17** to evaluate the estimator $\widehat{\Lambda}_t(z, z')$ in (409) as a Three Stage

Least Square Regression (3SLS) that uses the prediction of instrumental variable indicators as the instrumental variable followed by a 2SLS regression.

Theorem T-7. *Consider a 3SLS regression described in (417)–(420). The first stage consists of a OLS regression that uses the instrumental variable indicator $D_{z,\omega} \equiv \mathbf{1}[Z_\omega = z]$ as dependent variable and the local polynomial of propensity scores $\boldsymbol{\lambda}_{t,\omega}$ (described in Section P.1) as exogenous variable.*

$$\textit{First Stage: } D_{z,\omega} = \boldsymbol{\lambda}_{t,\omega}\boldsymbol{\theta}_{t,z} + \epsilon_{\omega,D} \textit{ for } z, z' \in \operatorname{supp}(Z). \tag{417}$$

The regression (417) is performed for instrumental values $z, z' \in \operatorname{supp} Z$. Let $\hat{D}_{z,\omega}$ be the forecast of the z-regression for agent ω and let $\tilde{Z}_{z,z',\omega}$ in (418) be the difference of the forecasts $\hat{D}_{z,\omega}, \hat{D}_{z',\omega}$ weighted by the inverse of the probability that the instrumental variable Z takes values z, z' respectively:

$$\textit{Forecast: } \tilde{Z}_{z,z',\omega} = \hat{D}_{z,\omega}\frac{1}{P_z} - \hat{D}_{z',\omega}\frac{1}{P_{z'}} \tag{418}$$

The second and third stages perform a standard 2SLS regression that uses $\tilde{Z}_{z,z',\omega}$ as the instrument, the treatment indicator $D_{t,\omega} \equiv \mathbf{1}[T_\omega = t]$ as the endogenous variable and $D_{t,\omega} \odot Y_\omega$ and the dependent outcome:

$$\textit{Second Stage: } D_{t,\omega} = \gamma \cdot \tilde{Z}_{z,z',\omega} + \epsilon_{\omega,D} \tag{419}$$

$$\textit{Third Stage: } Y_\omega \cdot D_{t,\omega} = \beta \cdot D_{t,\omega} + \epsilon_{\omega,Y}. \tag{420}$$

*If unordered monotonicity (13) hold, then for any $t \in \operatorname{supp}(T)$ and any two values $z, z' \in \operatorname{supp}(Z)$, the estimator for β and is equal to the estimator $\widehat{\Lambda}_t(z, z')$ in (82) of Theorem **T-6** which evaluates the counterfactual mean $E(Y(t)|\boldsymbol{S} \in \Sigma_t(z) \oplus \Sigma_t(z'))$.*

Proof. The theorem is a direct consequence of the equality stated in Lemma **L-17**. For instance, the vector of forecasts defined in (418) is given by $\widetilde{\mathbb{Z}}_{z,z'}$ in (414). Thus the estimator for β in (420) is given by the equation (415). According to Lemma **L-17**, the estimator is equal to $\widehat{\Lambda}_t(z, z')$ in (409). According to Theorem **T-6**, this estimator evaluates the counterfactual mean $E(Y(t)|\boldsymbol{S} \in \Sigma_t(z) \oplus \Sigma_t(z'))$. □

P.3 Estimating Partially Identified Counterfactual Outcomes Using 2SLS

The MTO response matrix in **L-3** produces three point identified counterfactual means conditioned on two response-types. Section 9 offers a solution to disentangle each of these counterfactual means into two counterfactual means conditioned on a single response-type. Specifically, we have that:

$$E(Y(t_h)|\boldsymbol{S} \in \{\boldsymbol{s}_4, \boldsymbol{s}_5\}) \text{ disentangled into } E(Y(t_h)|\boldsymbol{S} = \boldsymbol{s}_4) \text{ and } E(Y(t_h)|\boldsymbol{S} = \boldsymbol{s}_5) \tag{421}$$

$$E(Y(t_m)|\boldsymbol{S} \in \{\boldsymbol{s}_4, \boldsymbol{s}_7\}) \text{ disentangled into } E(Y(t_m)|\boldsymbol{S} = \boldsymbol{s}_4) \text{ and } E(Y(t_m)|\boldsymbol{S} = \boldsymbol{s}_7) \tag{422}$$

$$E(Y(t_l)|\boldsymbol{S} \in \{\boldsymbol{s}_4, \boldsymbol{s}_6\}) \text{ disentangled into } E(Y(t_l)|\boldsymbol{S} = \boldsymbol{s}_4) \text{ and } E(Y(t_l)|\boldsymbol{S} = \boldsymbol{s}_6). \tag{423}$$

This section explains how to apply Theorem **T-7** to estimate the counterfactual outcomes that are partially identified following the discussed in Section 9.4. This section focuses on the estimation of the unconditional estimation of $E(Y(t_l)|\boldsymbol{S} = \boldsymbol{s}_4)$. The estimation of the unconditional means

for the remaining counterfactual outcomes in (421)–(423) is obtained in the same fashion as the estimation of $E(Y(t_l)|\boldsymbol{S}=\boldsymbol{s}_4)$.

The counterfactual $E(Y(t_l)|\boldsymbol{S}\in\{\boldsymbol{s}_4,\boldsymbol{s}_6\})$ can be written as $E(Y(t_l)|\boldsymbol{S}\in\Sigma_{t_l}(z_e)\setminus\Sigma_{t_l}(z_8))$ and, according to Theorem **T-6**, it can be estimated by:

$$E(Y(t_l)|\boldsymbol{S}\in\{\boldsymbol{s}_4,\boldsymbol{s}_6\}) \text{ is estimated by } \widehat{\Lambda}_{t_l}(z_e,z_8)=\frac{\Big(\boldsymbol{\lambda}(P_{t_l}(z_e))-\boldsymbol{\lambda}(P_{t_l}(z_8))\Big)'\widehat{\boldsymbol{\beta}}_{t_l}}{\Big(\boldsymbol{\lambda}(P_{t_l}(z_e))-\boldsymbol{\lambda}(P_{t_l}(z_8))\Big)'\widehat{\boldsymbol{\theta}}_{t_l}}, \tag{424}$$

where $P_{t_l}(z_8)\equiv P(T=t_l|Z=z_8)=P(\boldsymbol{S}\in\{\boldsymbol{s}_3,\boldsymbol{s}_5\})\ <\ P_{t_l}(z_e)\equiv P(T=t_l|Z=z_e)=P(\boldsymbol{S}\in\{\boldsymbol{s}_3,\boldsymbol{s}_5,\boldsymbol{s}_4,\boldsymbol{s}_6\})$.

Previous section shows that the estimator $\widehat{\Lambda}_{t_l}(z_e,z_8)$ in (424) can be represented as the output of a 2SLS regression. This representation boils down to a simple substitution. The parameters $\widehat{\boldsymbol{\beta}}_{t_l}$ and $\widehat{\boldsymbol{\theta}}_{t_l}$ are given by:

$$\widehat{\boldsymbol{\beta}}_{t_l}=\Big(\mathbb{M}_{t_l}'\mathbb{M}_{t_l}\Big)^{-1}\mathbb{M}_{t_l}'\Big(\mathbb{D}_{t_l}\odot\mathbb{Y}\Big), \tag{425}$$

$$\widehat{\boldsymbol{\theta}}_{t_l}=\Big(\mathbb{M}_{t_l}'\mathbb{M}_{t_l}\Big)^{-1}\mathbb{M}_{t_l}'\mathbb{D}_{t_l}, \tag{426}$$

where $\mathbb{M}_{t_l}$ denotes an observed $N\times 3$ matrix whose ω-th row is given by the transpose of the vector $\boldsymbol{\lambda}(P_{t_l}(Z_\omega))\equiv[\lambda_1(P_{t_l}(Z_\omega)),\lambda_2(P_{t_l}(Z_\omega)),\lambda_3(P_{t_l}(Z_\omega))]'$. Therefore we have that the estimator $\widehat{\Lambda}_{t_l}(z_e,z_8)$ in (424) can be rewritten as:

$$\widehat{\Lambda}_{t_l}(z_e,z_8)=\frac{\Big(\boldsymbol{\lambda}(P_{t_l}(z_e))-\boldsymbol{\lambda}(P_{t_l}(z_8))\Big)'\Big(\mathbb{M}_{t_l}'\mathbb{M}_{t_l}\Big)^{-1}\mathbb{M}_{t_l}'\Big(\mathbb{D}_{t_l}\odot\mathbb{Y}\Big)}{\Big(\boldsymbol{\lambda}(P_{t_l}(z_e))-\boldsymbol{\lambda}(P_{t_l}(z_8))\Big)'\Big(\mathbb{M}_{t_l}'\mathbb{M}_{t_l}\Big)^{-1}\mathbb{M}_{t_l}'\mathbb{D}_{t_l}}, \tag{427}$$

Thus if we set the instrument to $\widetilde{\mathbb{Z}}_{t_l}(z_e,z_8)\equiv\mathbb{M}_{t_l}\Big(\mathbb{M}_{t_l}'\mathbb{M}_{t_l}\Big)^{-1}\Big(\boldsymbol{\lambda}\big(P_{t_l}(z_e)\big)-\boldsymbol{\lambda}\big(P_{t_l}(z_8)\big)\Big)$ then it is possible to express the estimator $\widehat{\Lambda}_{t_l}(z_e,z_8)$ in (427) as an output of a 2SLS regression:

$$\widehat{\Lambda}_{t_l}(z_e,z_8)=\frac{\widetilde{\mathbb{Z}}'\Big(\mathbb{D}_{t_l}\odot\mathbb{Y}\Big)}{\widetilde{\mathbb{Z}}'\mathbb{D}_{t_l}}, \text{ where } \widetilde{\mathbb{Z}}\equiv\widetilde{\mathbb{Z}}_{t_l}(z_e,z_8) \tag{428}$$

$$=\Big(\widetilde{\mathbb{Z}}'\mathbb{D}_{t_l}\Big)^{-1}\widetilde{\mathbb{Z}}'\Big(\mathbb{D}_{t_l}\odot\mathbb{Y}\Big), \tag{429}$$

$$=\Big(\mathbb{D}_{t_l}'\widetilde{\mathbb{Z}}'(\widetilde{\mathbb{Z}}'\widetilde{\mathbb{Z}})^{-1}\widetilde{\mathbb{Z}}'\mathbb{D}_{t_l}\Big)^{-1}\mathbb{D}_{t_l}'\widetilde{\mathbb{Z}}'(\widetilde{\mathbb{Z}}'\widetilde{\mathbb{Z}})^{-1}\widetilde{\mathbb{Z}}'\Big(\mathbb{D}_{t_l}\odot\mathbb{Y}\Big), \tag{430}$$

$$=\Big(\mathbb{D}_{t_l}'\mathbb{P}_{\widetilde{\mathbb{Z}}}\mathbb{D}_{t_l}\Big)^{-1}\mathbb{D}_{t_l}'\mathbb{P}_{\widetilde{\mathbb{Z}}}\Big(\mathbb{D}_{t_l}\odot\mathbb{Y}\Big). \tag{431}$$

Now consider the estimation of $E(Y(t_l)|\boldsymbol{S}=\boldsymbol{s}_4)$. According to Section 9.2, this counterfactual

mean is evaluated by the following estimator:

$$E(Y(t_l)|\boldsymbol{S}=\boldsymbol{s}_4) \text{ is estimated by } \widehat{\Lambda}_{t_l}(z^*,z_8)=\frac{\Big(\boldsymbol{\lambda}(P_{t_l}(z^*))-\boldsymbol{\lambda}(P_{t_l}(z_8))\Big)'\widehat{\boldsymbol{\beta}}_{t_l}}{\Big(\boldsymbol{\lambda}(P_{t_l}(z^*))-\boldsymbol{\lambda}(P_{t_l}(z_8))\Big)'\widehat{\boldsymbol{\theta}}_{t_l}}, \tag{432}$$

$$\text{where } P_{t_l}(z_8)=P(\boldsymbol{S}\in\{\boldsymbol{s}_3,\boldsymbol{s}_5\}) \;<\; P(\boldsymbol{S}\in\{\boldsymbol{s}_3,\boldsymbol{s}_5,\boldsymbol{s}_4\})\equiv P_{t_l}(z^*). \tag{433}$$

The estimation of the counterfactual mean $E(Y(t_l)|\boldsymbol{S}=\boldsymbol{s}_4)$ in (432) hinges on two probabilities $P_{t_l}(z^*)\equiv P(\boldsymbol{S}\in\{\boldsymbol{s}_3,\boldsymbol{s}_5,\boldsymbol{s}_4\})$ and $P_{t_l}(z_8)\equiv P(T=t_l|Z=z_8)$. The probability $P_{t_l}(z_8)$ is observed and $P_{t_l}(z^*)$ is point identified and can be estimated. Thus we can apply the same rationale of equations (424)–(431) to evaluate the $E(Y(t_l)|\boldsymbol{S}=\boldsymbol{s}_4)$ in (432) as a 2SLS regression. Specifically, $E(Y(t_l)|\boldsymbol{S}=\boldsymbol{s}_4)$ can be estimated as:

$$\widehat{\Lambda}_{t_l}(z^*,z_8)=\Big(\mathbb{D}'_{t_l}\widetilde{\mathbb{Z}}'(\widetilde{\mathbb{Z}}'\widetilde{\mathbb{Z}})^{-1}\widetilde{\mathbb{Z}}'\mathbb{D}_{t_l}\Big)^{-1}\mathbb{D}'_{t_l}\widetilde{\mathbb{Z}}'(\widetilde{\mathbb{Z}}'\widetilde{\mathbb{Z}})^{-1}\widetilde{\mathbb{Z}}'\Big(\mathbb{D}_{t_l}\odot\mathbb{Y}\Big), \tag{434}$$

$$\text{where } \widetilde{\mathbb{Z}} \text{ is given by } \widetilde{\mathbb{Z}}_{t_l}(z^*,z_8)=\Big(\mathbb{M}'_{t_l}\mathbb{M}_{t_l}\Big)^{-1}\Big(\boldsymbol{\lambda}\big(P_{t_l}(z^*)\big)-\boldsymbol{\lambda}\big(P_{t_l}(z_8)\big)\Big), \tag{435}$$

$$\text{such that } P_{t_l}(z^*) \text{ is the estimate of } P(\boldsymbol{S}\in\{\boldsymbol{s}_3,\boldsymbol{s}_5,\boldsymbol{s}_4\}). \tag{436}$$

Counterfactual Outcome Means Controlling for Baseline Variables and Weights

Let $\mathbb{X}_0$ be the $N\times K$ standardized matrix of baseline variable we wish to control for. Each column of $\mathbb{X}_0$ has zero mean. Let $\mathbb{W}$ be the $N\times 1$ vector of positive weights. See Section P.1 for a description of these variables. Let $\text{diag}(\mathbb{W})^{1/2}$ be the $N\times N$ diagonal matrix whose diagonal elements are the squared root of the weights in $\mathbb{W}$. Let $\widetilde{\mathbb{X}}_0\equiv\text{diag}(\mathbb{W})^{1/2}\mathbb{X}_0$ be the weighted version of the standardized baseline variables. As customary, $\mathbb{D}_t$ denotes the $N\times 1$ vector of choice indicator for $t\in\{t_h,t_m,t_l\}$ and $\mathbb{D}_t\odot\mathbb{Y}$ be the $N\times 1$ denotes the vector of element-wise multiplication between the choice indicator vector $\mathbb{D}_t$ and the outcome vector $\mathbb{Y}$. Let $\widetilde{\mathbb{D}}_t\equiv\text{diag}(\mathbb{W})^{1/2}\mathbb{D}_t$ be the weighted version of the vector of choice indicator and $\widetilde{\mathbb{Y}}_t\equiv\text{diag}(\mathbb{W})^{1/2}\big(\mathbb{D}_t\odot\mathbb{Y}\big)$ be the weighted version of the vector $\mathbb{D}_t\odot\mathbb{Y}$ that plays the role of outcome vector in our 2SLS regression. The projection on the space generated by the columns of a given matrix $\widetilde{\mathbb{Z}}$ representing the instrumental variable a matrix $\widetilde{\mathbb{X}}_0$ representing baseline variables is given by:

$$\mathbb{P}_{\widetilde{\mathbb{Z}},\widetilde{\mathbb{X}}_0}=[\widetilde{\mathbb{Z}},\widetilde{\mathbb{X}}_0]\Big([\widetilde{\mathbb{Z}},\widetilde{\mathbb{X}}_0]'[\widetilde{\mathbb{Z}},\widetilde{\mathbb{X}}_0]\Big)^{-1}[\widetilde{\mathbb{Z}},\widetilde{\mathbb{X}}_0]'.$$

Our goal is to estimate the counterfactual outcome mean $E(Y(t_l)|\boldsymbol{S}=\boldsymbol{s}_4)$ accounting for pre-program variables $\mathbb{X}_0$ and weights $\mathbb{W}$. This can be achieved by the estimate of $\boldsymbol{\beta}$ a 2SLS regression:

$$\widetilde{\mathbb{Y}}_{t_l}=\widetilde{\mathbb{D}}_{t_l}\boldsymbol{\beta}+\widetilde{\mathbb{X}}_0\boldsymbol{\gamma}+\boldsymbol{\epsilon}, \tag{437}$$

where the instrumental variable is set to be $[\widetilde{\mathbb{Z}},\widetilde{\mathbb{X}}_0]$ where $\widetilde{\mathbb{Z}}$ is given by:

$$\widetilde{\mathbb{Z}}\equiv\text{diag}(\mathbb{W})^{1/2}\widetilde{\mathbb{Z}}_{t_l}(z^*,z_8)=\text{diag}(\mathbb{W})^{1/2}\Big(\mathbb{M}'_{t_l}\mathbb{M}_{t_l}\Big)^{-1}\Big(\boldsymbol{\lambda}\big(P_{t_l}(z^*)\big)-\boldsymbol{\lambda}\big(P_{t_l}(z_8)\big)\Big), \tag{438}$$

such that $P_{t_l}(z^*)$ is the estimate of $P(\boldsymbol{S} \in \{\boldsymbol{s}_3, \boldsymbol{s}_5, \boldsymbol{s}_4\})$. The 2SLS estimator is therefore given by:

$$\begin{pmatrix} \widehat{\boldsymbol{\beta}} \\ \widehat{\boldsymbol{\gamma}} \end{pmatrix} = \left(\left[\widetilde{\mathbb{D}}_{t_l} \, , \, \widetilde{\mathbb{X}}_0 \right]' \mathbb{P}_{\widetilde{\mathbb{Z}}, \widetilde{\mathbb{X}}_0} \left[\widetilde{\mathbb{D}}_{t_l} \, , \, \widetilde{\mathbb{X}}_0 \right] \right)^{-1} \left(\left[\widetilde{\mathbb{D}}_{t_l} \, , \, \widetilde{\mathbb{X}}_0 \right]' \mathbb{P}_{\widetilde{\mathbb{Z}}, \widetilde{\mathbb{X}}_0} \widetilde{\mathbb{Y}}_{t_l} \right). \tag{439}$$

Inference can be obtained by the method of bootstrap. In practice, the 2SLS in (406) can be obtained by the following steps:

1. Estimate the response-type probabilities controlling for weights $\mathbb{W}$ and standardized baseline variables $\mathbb{X}_0$ as described in Section (O.3).

2. From those probability estimates, set the values for the two main probabilities $P_{t_l}(z_8) = P(\boldsymbol{S} \in \{\boldsymbol{s}_3, \boldsymbol{s}_5\})$ and $P_{t_l}(z^*) = P(\boldsymbol{S} \in \{\boldsymbol{s}_3, \boldsymbol{s}_5, \boldsymbol{s}_4\})$ but also for $P_{t_l}(z_c) = P(\boldsymbol{S} = \boldsymbol{s}_3)$ and $P_{t_l}(z_e) = P(\boldsymbol{S} \in \{\boldsymbol{s}_3, , \boldsymbol{s}_4, \boldsymbol{s}_5, , \boldsymbol{s}_6\})$.

3. Set the polynomial foe the vector-valued function for $\boldsymbol{\lambda}(x)$ and use the estimated probabilities to compute the $N \times 3$ matrix $\mathbb{M}_{t_l}$ and the 3×1 vector difference $\Big(P_{t_l}(z^*)\Big) - \boldsymbol{\lambda}\Big(P_{t_l}(z_8)\Big)\Big)$.

4. Compute the $N \times 1$ instrumental variable given by
$$\widetilde{\mathbb{Z}}_{t_l}(z^*, z_8) = \Big(\mathbb{M}'_{t_l}\mathbb{M}_{t_l}\Big)^{-1}\Big(\boldsymbol{\lambda}\Big(P_{t_l}(z^*)\Big) - \boldsymbol{\lambda}\Big(P_{t_l}(z_8)\Big)\Big).$$

5. Multiply the outcome of interest $\mathbb{Y}$ by the indicator $\mathbb{D}_{t_l}$, that is $\mathbb{Y} \odot \mathbb{D}_{t_l}$.

6. Perform a weighted 2SLS of $\mathbb{Y} \odot \mathbb{D}_{t_l}$ on $\mathbb{D}_{t_l}, \mathbb{X}_0$ using $\widetilde{\mathbb{Z}}_{t_l}(z^*, z_8)$ as instrumental variables and $\mathbb{W}$ as weights.

Estimation of Causal Effects

Consider the estimation of $E(Y(t_h)|\boldsymbol{S} = \boldsymbol{s}_4)$ instead of $E(Y(t_l)|\boldsymbol{S} = \boldsymbol{s}_4)$. According to Table 10 in Section 9.2, this counterfactual mean is evaluated by the following estimator:

$$E(Y(t_h)|\boldsymbol{S} = \boldsymbol{s}_4) \text{ is estimated by } \widehat{\Lambda}_{t_h}(z^*, z_e) = \frac{\Big(\boldsymbol{\lambda}(P_{t_h}(z^*)) - \boldsymbol{\lambda}(P_{t_h}(z_e))\Big)' \widehat{\boldsymbol{\beta}}_{t_h}}{\Big(\boldsymbol{\lambda}(P_{t_h}(z^*)) - \boldsymbol{\lambda}(P_{t_h}(z_e))\Big)' \widehat{\boldsymbol{\theta}}_{t_h}},$$

$$\text{where } P_{t_h}(z_e) = P(\boldsymbol{S} \in \{\boldsymbol{s}_1, \boldsymbol{s}_7\}) \; < \; P(\boldsymbol{S} \in \{\boldsymbol{s}_1, \boldsymbol{s}_7, \boldsymbol{s}_4\}) \equiv P_{t_h}(z^*)$$

$$\text{and } \widehat{\boldsymbol{\beta}}_{t_h} = \Big(\mathbb{M}'_{t_h}\mathbb{M}_{t_h}\Big)^{-1}\mathbb{M}'_{t_h}\Big(\mathbb{D}_{t_h} \odot \mathbb{Y}\Big),$$

$$\widehat{\boldsymbol{\theta}}_{t_h} = \Big(\mathbb{M}'_{t_h}\mathbb{M}_{t_h}\Big)^{-1}\mathbb{M}'_{t_h}\mathbb{D}_{t_h}.$$

We can then follow the same steps that yield the estimation of $E(Y(t_l)|\boldsymbol{S} = \boldsymbol{s}_4)$. In doing so, the estimation of $E(Y(t_h)|\boldsymbol{S} = \boldsymbol{s}_4)$ can be achieved by the estimate of the parameter $\boldsymbol{\beta}$ in the following 2SLS regression:

$$\widetilde{\mathbb{Y}}_{t_h} = \widetilde{\mathbb{D}}_{t_h}\boldsymbol{\beta} + \widetilde{\mathbb{X}}_0\boldsymbol{\gamma} + \boldsymbol{\epsilon}, \tag{440}$$

where $\widetilde{\mathbb{D}}_{t_h} \equiv \text{diag}(\mathbb{W})^{1/2}\mathbb{D}_{t_h}$, $\widetilde{\mathbb{Y}}_{t_h} \equiv \text{diag}(\mathbb{W})^{1/2}\big(\mathbb{D}_{t_h} \odot \mathbb{Y}\big)$, and the instrumental variable is set to be $[\widetilde{\mathbb{Z}}, \widetilde{\mathbb{X}}_0]$ where $\widetilde{\mathbb{Z}}$ is given by:

$$\widetilde{\mathbb{Z}} \equiv \text{diag}(\mathbb{W})^{1/2}\widetilde{\mathbb{Z}}_{t_l}(z^*, z_8) = \text{diag}(\mathbb{W})^{1/2}\Big(\mathbb{M}'_{t_h}\mathbb{M}_{t_h}\Big)^{-1}\Big(\boldsymbol{\lambda}\big(P_{t_h}(z^*)\big) - \boldsymbol{\lambda}\big(P_{t_h}(z_e)\big)\Big), \tag{441}$$

where $P_{t_h}(z^*)$ is the estimate of $P(\boldsymbol{S} \in \{\boldsymbol{s}_1, \boldsymbol{s}_7, \boldsymbol{s}_4\})$. The 2SLS estimator is therefore given by:

$$\begin{pmatrix} \widehat{\boldsymbol{\beta}} \\ \widehat{\boldsymbol{\gamma}} \end{pmatrix} = \left(\left[\widetilde{\mathbb{D}}_{t_h}\,,\, \widetilde{\mathbb{X}}_0\right]' \mathbb{P}_{\widetilde{\mathbb{Z}},\widetilde{\mathbb{X}}_0} \left[\widetilde{\mathbb{D}}_{t_h}\,,\, \widetilde{\mathbb{X}}_0\right]\right)^{-1} \left(\left[\widetilde{\mathbb{D}}_{t_h}\,,\, \widetilde{\mathbb{X}}_0\right]' \mathbb{P}_{\widetilde{\mathbb{Z}},\widetilde{\mathbb{X}}_0} \widetilde{\mathbb{Y}}_{t_h}\right). \tag{442}$$

The estimation the causal effect $E(Y(t_l) - Y(t_h)|\boldsymbol{S} = \boldsymbol{s}_4)$ can be obtained by the difference between the estimates of $E(Y(t_l)|\boldsymbol{S} = \boldsymbol{s}_4)$ and $E(Y(t_h)|\boldsymbol{S} = \boldsymbol{s}_4)$. This estimate can also be evaluated as an output of a 2SLS that stacks the data associated with the 2SLS in (437) and (440). It is useful to define a generic matrix representation of a 2SLS regression by:

$$\mathbb{Y}_g = \mathbb{D}_g\boldsymbol{\beta} + \mathbb{X}_g\boldsymbol{\gamma} + \boldsymbol{\epsilon}, \tag{443}$$

where $\mathbb{Y}_g$ denotes the outcome, $\mathbb{D}_g$ denotes the endogenous variable, $\mathbb{Z}_g$ denotes the instrumental variable, $\mathbb{X}_g$ denotes the pre-program variables, and $\mathbb{W}_g$ denotes weights. In this notation, the parameters $\boldsymbol{\beta}, \boldsymbol{\gamma}$ are estimated by:

$$\begin{pmatrix} \widehat{\boldsymbol{\beta}} \\ \widehat{\boldsymbol{\gamma}} \end{pmatrix} = \left(\left[\widetilde{\mathbb{D}}_g\,,\, \widetilde{\mathbb{X}}_g\right]' \mathbb{P}_{\widetilde{\mathbb{Z}}_g,\widetilde{\mathbb{X}}_g} \left[\widetilde{\mathbb{D}}_g\,,\, \widetilde{\mathbb{X}}_g\right]\right)^{-1} \left(\left[\widetilde{\mathbb{D}}_g\,,\, \widetilde{\mathbb{X}}_g\right]' \mathbb{P}_{\widetilde{\mathbb{Z}}_g,\widetilde{\mathbb{X}}_g} \widetilde{\mathbb{Y}}_g\right), \tag{444}$$

$$\text{where } \widetilde{\mathbb{D}}_g \equiv \text{diag}(\mathbb{W}_g)^{1/2}\mathbb{D}_g,\ \widetilde{\mathbb{Z}}_g \equiv \text{diag}(\mathbb{W}_g)^{1/2}\mathbb{Z}_g,\ \widetilde{\mathbb{X}}_g \equiv \text{diag}(\mathbb{W}_g)^{1/2}\mathbb{X}_g,\ \widetilde{\mathbb{Y}}_g \equiv \text{diag}(\mathbb{W}_g)^{1/2}\mathbb{Y}_g. \tag{445}$$

In this notation, the causal effect $E(Y(t_l) - Y(t_h)|\boldsymbol{S} = \boldsymbol{s}_4)$ can be evaluated by the estimate of $\widehat{\boldsymbol{\beta}}$ in in (444) when matrices $\mathbb{X}_g, \mathbb{D}_g, \mathbb{Z}_g, \mathbb{Y}_g$ are set to:

$$\mathbb{W}_g \equiv \mathbf{1}_{2,1} \otimes \mathbb{W}, \tag{446}$$

$$\mathbb{X}_g \equiv \boldsymbol{I}_2 \otimes \mathbb{X}_0, \tag{447}$$

$$\mathbb{Z}_g = \begin{bmatrix} \widetilde{\mathbb{Z}}_{t_l}(z^*, z_8) & \mathbf{0}_{N,1} \\ \mathbf{0}_{N,1} & \widetilde{\mathbb{Z}}_{t_h}(z^*, z_e) \end{bmatrix}, \tag{448}$$

$$\mathbb{D}_g \equiv \begin{bmatrix} \mathbb{D}_{t_l} & \mathbb{D}_{t_l} \\ \mathbf{0}_{N,1} & \mathbb{D}_{t_h} \end{bmatrix}, \tag{449}$$

$$\mathbb{Y}_g \equiv \begin{bmatrix} \mathbb{Y} \otimes \mathbb{D}_{t_l} \\ \mathbb{Y} \otimes \mathbb{D}_{t_h} \end{bmatrix}, \tag{450}$$

where $\mathbf{1}_{r,c}$ denotes a matrix of dimension $r \times c$, $\boldsymbol{I}_k$ stands for the identity matrix of dimension k and $\otimes$ denotes the Kronecker multiplication. In this settings, the first element of the estimate $\hat{\boldsymbol{\beta}}$ is numerically the same as the difference of counterfactual outcome means estimated in the 2SLS regressions (437) and (440). Note that the parameter associated with pre-program variables $\mathbb{X}_0$ is allowed to differ by treatment choices T_l, t_h. These coefficients could be constrained to be the same across treatment choices. To do so, it suffices to set $\mathbb{X}_g \equiv \mathbf{1}_{2,1} \otimes \mathbb{X}_0$, instead of $\boldsymbol{I}_2 \otimes \mathbb{X}_0$.