Online Appendix for:

Beyond Intention to Treat: Using the Incentives in *Moving to Opportunity* to Identify Neighborhood Effects

A Mathematical Proofs

A.1 Proof of Equation (8)

Let the response vector be $\boldsymbol{S} = [T(z_1), ..., T(z_n)]'$ where $supp(Z) = \{z_1, ...z_n\}$. Note that the treatment T can be expressed as $T = [\mathbf{1}[Z = z_1], ..., \mathbf{1}[Z = z_1]]\boldsymbol{S}$. This implies that T depends only on Z when conditioned on $\boldsymbol{S}$. Moreover, T is deterministic given Z and $\boldsymbol{S}$. The Exogeneity Condition 2 states that $Z \perp\!\!\!\perp (Y(t), T(z), Y(z))$. This assumption implies the following relationships:

$$Z \perp\!\!\!\perp \boldsymbol{S} \tag{65}$$

$$Y(t) \perp\!\!\!\perp (Z, T)|\boldsymbol{S} \tag{66}$$

Relationship (65) is due to $Z \perp\!\!\!\perp T(z)$. Relationship (66) arises from $Y(t) \perp\!\!\!\perp Z|T(z)$ and the fact that T is a function of Z when conditioned on $\boldsymbol{S}$. Finally, the Exclusion Restriction (1) enable us to express the observed outcome as:

$$Y = \sum_{t \in supp(T)} \mathbf{1}[T = t] \cdot Y(t). \tag{67}$$

The derivation of the equation (8) is displayed below:

$$\begin{aligned}
E(Y|Z = z, T = t) &= \sum_{\boldsymbol{s} \in supp(\boldsymbol{S})} E(Y|Z = z, T = t, \boldsymbol{S} = \boldsymbol{s})P(\boldsymbol{S} = \boldsymbol{s}|T = t, Z = z) \\
&= \sum_{\boldsymbol{s} \in supp(\boldsymbol{S})} E(Y|Z = z, T = t, \boldsymbol{S} = \boldsymbol{s})\frac{P(T = t|\boldsymbol{S} = \boldsymbol{s}, Z = z)P(\boldsymbol{S} = \boldsymbol{s}|Z = z)}{P(T = t|Z = z)} \\
\Rightarrow E(Y|Z = z, T = t)P(T = t|Z = z) &= \sum_{\boldsymbol{s} \in supp(\boldsymbol{S})} E(Y|Z = z, T = t, \boldsymbol{S} = \boldsymbol{s})P(T = t|\boldsymbol{S} = \boldsymbol{s}, Z = z)P(\boldsymbol{S} = \boldsymbol{s}|Z = z) \\
&= \sum_{\boldsymbol{s} \in supp(\boldsymbol{S})} E(Y|Z = z, T = t, \boldsymbol{S} = \boldsymbol{s})P(T = t|\boldsymbol{S} = \boldsymbol{s}, Z = z)P(\boldsymbol{S} = \boldsymbol{s}) \\
&= \sum_{\boldsymbol{s} \in supp(\boldsymbol{S})} E(Y|Z = z, \boldsymbol{S} = \boldsymbol{s})\mathbf{1}[T = t|\boldsymbol{S} = \boldsymbol{s}, Z = z]P(\boldsymbol{S} = \boldsymbol{s}) \\
&= \sum_{\boldsymbol{s} \in supp(\boldsymbol{S})} \mathbf{1}[T = t|\boldsymbol{S} = \boldsymbol{s}, Z = z]E(Y|Z = z, T = t, \boldsymbol{S} = \boldsymbol{s})P(\boldsymbol{S} = \boldsymbol{s}) \\
&= \sum_{\boldsymbol{s} \in supp(\boldsymbol{S})} \mathbf{1}[T = t|\boldsymbol{S} = \boldsymbol{s}, Z = z]E(Y(t)|Z = z, T = t, \boldsymbol{S} = \boldsymbol{s})P(\boldsymbol{S} = \boldsymbol{s}) \\
&= \sum_{\boldsymbol{s} \in supp(\boldsymbol{S})} \mathbf{1}[T = t|\boldsymbol{S} = \boldsymbol{s}, Z = z]E(Y(t)|\boldsymbol{S} = \boldsymbol{s})P(\boldsymbol{S} = \boldsymbol{s})
\end{aligned}$$

The first equality applies the law of iterated expectations to the expectation $E(Y|Z = z, T = t)$. The second equality uses the Bayes' theorem. The third equality multiplies both sides of the equation by $P(T = t|Z = z)$. The fourth equality arises from (65). The fifth equality is due the fact that T is deterministic conditioned on $\boldsymbol{S}$ and Z. Thus $P(T=t|\boldsymbol{S}=\boldsymbol{s}, Z=z)$ is either zero or one. The sixth equality simply reorders the terms of the summation. The seventh equality is due to (67). The eight equality is due to (66).

A.2 Proof of Proposition P.1

We seek to obtain a choice rule the stems from WARP and the budget set relationships defined by (15). It is useful to define some basic nomenclature to proof the proposition.

If we fix the instrument to a value $z \in \mathit{supp}(Z)$, then all the bundles $(t,g); g \in \mathcal{B}_i(z,t)$ for any $t \in \mathit{supp}(T)$ are said to be available for family i. If a family prefers a bundle (t,g) instead of (t',g') when both are available, then (t,g) is said to be directly and strictly revealed preferred to (t',g'), that is, $(t,g) \succ_i^d (t',g')$. In particular, if a family i chooses choice t when the IV value is fixed to z, that is, $T_i(z)=t$, then there exists a bundle (t,g^*) for some $g^* \in \mathcal{B}_i(z,t)$ that is strictly revealed preferred to all available bundles, namely, all the bundles $(t',g'); g' \in \mathcal{B}_i(z,t')$ for any choices t' that are different than t. Notationally, we have that $(t,g^*) \succ_i^d (t',g') \, \forall \, g' \in \mathcal{B}_i(z,t'); t' \in \mathit{supp}(T) \setminus \{t\}$.

The WARP criteria of Richter (1971) states that if bundle (t,g) is directly and strictly revealed preferred to (t',g'), that is, $(t,g) \succ_i^d (t',g')$, then (t',g') cannot be revealed preferred to (t,g), namely, $(t,g) \succ_i^d (t',g') \Rightarrow (t',g') \not\succ_i^d (t,g)$.

Each (z,t)-entry of the incentive matrix $\boldsymbol{L}$, that is $\boldsymbol{L}[z,t]$ presents a value that is a relative rank of incentives of z towards choice t. Each t-column of matrix presents the relative rank of incentives for choice t across the IV-values. This ranking renders the budget set relationships in (15), which states that for a given choice t, $\boldsymbol{L}[z,t] \leq \boldsymbol{L}[z',t]$ imply that $\mathcal{B}_i(z,t) \subseteq \mathcal{B}_i(z',t)$ for all $i \in \mathcal{I}$.

It is useful to prove the following Lemma regarding WARP before proving the main proposition:

Lemma L.1. Under budget set relationships (15) and WARP, the following choice rule holds:

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

Proof. Suppose that family i chooses choice t instead of t' when the IV value is fixed to z. Thus there exist a bundle (t,g^*) for some $g^* \in \mathcal{B}_i(z,t)$ such that $(t,g^*) \succ_i^d (t',g')$ for all $g' \in \mathcal{B}_i(z,t')$. Now consider a shift of the IV from z to z'. Suppose that $0 \leq \boldsymbol{L}[z',t] - \boldsymbol{L}[z,t]$ holds. According to (15), the budget set $\mathcal{B}_i(z',t)$ is at least as big as $\mathcal{B}_i(z,t')$. In particular, the bundle (t,g^*) is available. Moreover, suppose that $\boldsymbol{L}[z',t'] - \boldsymbol{L}[z,t'] \leq 0$ also holds. This mean that the budget set $\mathcal{B}_i(z',t')$ is not larger than $\mathcal{B}_i(z,t')$ and, according to WARP, the bundle (t,g^*) is preferred to all the bundles $(t',g'); g' \in \mathcal{B}_i(z',t')$. Thereby, family i will not choose any bundle $(t',g'); g' \in \mathcal{B}_i(z',t')$ over (t,g^*). Consequently, we have that $T_i(z')$ cannot be t', namely, $t_i(z') \neq t'$. □

Note that Lemma **L.1** is consistent with the relative ranking property of the incentive matrix. Any strictly monotonic increasing transformation of the matrix generates the same set of choice restrictions.

The Normal Choice is a statement that compares the relative incentive gain between treatment choices when the instrument changes. Normal choice is defined as:

$$(t \succ_i t')|z \text{ and } \boldsymbol{L}[z',t] - \boldsymbol{L}[z,t] \; = \; \boldsymbol{L}[z',t'] - \boldsymbol{L}[z,t'] \text{ then } (t' \not\succ_i t)|z',$$

We are now equipped to prove the main proposition. Suppose that:

$$T_i(z) = t \text{ and } \boldsymbol{L}[z',t'] - \boldsymbol{L}[z,t'] \leq \boldsymbol{L}[z',t] - \boldsymbol{L}[z,t] \equiv \ell.$$

Note that $\boldsymbol{L}[z',t'] \leq \boldsymbol{L}[z,t'] + \ell$. Now consider an IV-value z^* that offers the same incentives of z' towards t and at least as much incentives than z' towards t'. Namely, let the incentives associated with z^* be: $\boldsymbol{L}[z^*,t] \equiv \boldsymbol{L}[z',t]$ and $\boldsymbol{L}[z^*,t'] = \boldsymbol{L}[z,t'] + \ell \geq \boldsymbol{L}[z',t']$. We can apply normal choice to obtain the following preference restriction:

$$(t \succ_i t')|z \text{ and } \boldsymbol{L}[z^*,t] - \boldsymbol{L}[z,t] = \boldsymbol{L}[z^*,t'] - \boldsymbol{L}[z,t'] = \ell, \text{ then } (t' \not\succ_i t)|z^*.$$

This means that exists a bundle (t,g^*) for some $g^* \in \mathcal{B}_i(z^*,t)$ such that $(t,g^*) \succ_i^d (t',g')$ for all $g' \in \mathcal{B}_i(z^*,t')$. Note that $\mathcal{B}_i(z',t) = \mathcal{B}_i(z^*,t)$ and $\mathcal{B}_i(z',t') \subseteq \mathcal{B}_i(z^*,t)$. Thus, the bundle (t,g^*) is available under z' since $g^* \in \mathcal{B}_i(z',t)$. Moreover, the bundle (t,g^*) remains preferred to all bundles (t',g') for all $g' \in \mathcal{B}_i(z',t')$. Therefore, according to WARP, the i agent does not choose t' under z'.

A.3 Proof of Proposition P.2

Lemma L.1 states that under WARP, the following choice rule holds:

$$T_i(z) = t, \text{ and } \boldsymbol{L}[z',t'] - \boldsymbol{L}[z,t'] \leq 0 \leq \boldsymbol{L}[z',t] - \boldsymbol{L}[z,t], \text{ then } T_i(z') \neq t'. \tag{68}$$

The choice rule above compares two choices and two instrumental variables. There are six possibilities for distinct choices (t,t') such that $t \in \{t_h, t_m, t_l\}$ and $t' \in \{t_h, t_m, t_l\} \setminus \{t\}$. There are also six possibilities for the set of distinct instrumental values (z,z') such that $z \in \{z_c, z_8, z_e\}$ and $z' \in \{z_c, z_8, z_e\} \setminus \{z\}$. Thus, there are a total of 36 choice requirements of the type in (68) that can be checked using MTO data. Only 20 of these 36 possibilities are biding. The resulting 20 choice restrictions are presented in Table A.1. These restrictions are summarized into the eight choice restrictions displayed in Table A.2. The two last choice restrictions of Table A.2 are redundant given the first six restrictions.

In total, the revealed preference analysis generates seven choice restrictions. The seventh choice restriction is due to the Normal choice assumption (28), that is, $T_i(z_c) \neq t_h \Rightarrow T_i(z_8) = T_i(z_c)$.

Table A.1: Choice Restrictions Due to WARP

#	Revealed Choice	Incentive Inequalities					Choice Statement
	$T_i(z)=t$	$\boldsymbol{L}[z',t']-\boldsymbol{L}[z,t']\leq 0 \leq \boldsymbol{L}[z',t]-\boldsymbol{L}[z,t]$					$T(z')\neq t'$
1	$T_i(z_c)=t_h,$	$\boldsymbol{L}[z_e,t_m]-\boldsymbol{L}[z_c,t_m]=0$	$\leq$	0	$\leq$	$0=\boldsymbol{L}[z_e,t_h]-\boldsymbol{L}[z_c,t_h]$	$T_i(z_e)\neq t_m$
2	$T_i(z_c)=t_m,$	$\boldsymbol{L}[z_8,t_h]-\boldsymbol{L}[z_c,t_h]=0$	$\leq$	0	$\leq$	$1=\boldsymbol{L}[z_8,t_m]-\boldsymbol{L}[z_c,t_m]$	$T_i(z_8)\neq t_h$
3	$T_i(z_c)=t_m,$	$\boldsymbol{L}[z_e,t_h]-\boldsymbol{L}[z_c,t_h]=0$	$\leq$	0	$\leq$	$0=\boldsymbol{L}[z_e,t_m]-\boldsymbol{L}[z_c,t_m]$	$T_i(z_e)\neq t_h$
4	$T_i(z_c)=t_l,$	$\boldsymbol{L}[z_8,t_h]-\boldsymbol{L}[z_c,t_h]=0$	$\leq$	0	$\leq$	$1=\boldsymbol{L}[z_8,t_l]-\boldsymbol{L}[z_c,t_l]$	$T_i(z_8)\neq t_h$
5	$T_i(z_c)=t_l,$	$\boldsymbol{L}[z_e,t_h]-\boldsymbol{L}[z_c,t_h]=0$	$\leq$	0	$\leq$	$1=\boldsymbol{L}[z_e,t_l]-\boldsymbol{L}[z_c,t_l]$	$T_i(z_e)\neq t_h$
6	$T_i(z_c)=t_l,$	$\boldsymbol{L}[z_e,t_m]-\boldsymbol{L}[z_c,t_m]=0$	$\leq$	0	$\leq$	$1=\boldsymbol{L}[z_e,t_l]-\boldsymbol{L}[z_c,t_l]$	$T_i(z_e)\neq t_m$
7	$T_i(z_8)=t_h,$	$\boldsymbol{L}[z_c,t_m]-\boldsymbol{L}[z_8,t_m]=-1$	$\leq$	0	$\leq$	$0=\boldsymbol{L}[z_c,t_h]-\boldsymbol{L}[z_8,t_h]$	$T_i(z_c)\neq t_m$
8	$T_i(z_8)=t_h,$	$\boldsymbol{L}[z_e,t_m]-\boldsymbol{L}[z_8,t_m]=-1$	$\leq$	0	$\leq$	$0=\boldsymbol{L}[z_e,t_h]-\boldsymbol{L}[z_8,t_h]$	$T_i(z_e)\neq t_m$
9	$T_i(z_8)=t_h,$	$\boldsymbol{L}[z_c,t_l]-\boldsymbol{L}[z_8,t_l]=-1$	$\leq$	0	$\leq$	$0=\boldsymbol{L}[z_c,t_h]-\boldsymbol{L}[z_8,t_h]$	$T_i(z_c)\neq t_l$
10	$T_i(z_8)=t_h,$	$\boldsymbol{L}[z_e,t_l]-\boldsymbol{L}[z_8,t_l]=0$	$\leq$	0	$\leq$	$0=\boldsymbol{L}[z_e,t_h]-\boldsymbol{L}[z_8,t_h]$	$T_i(z_e)\neq t_l$
11	$T_i(z_8)=t_l,$	$\boldsymbol{L}[z_e,t_h]-\boldsymbol{L}[z_8,t_h]=0$	$\leq$	0	$\leq$	$0=\boldsymbol{L}[z_e,t_l]-\boldsymbol{L}[z_8,t_l]$	$T_i(z_e)\neq t_h$
12	$T_i(z_8)=t_l,$	$\boldsymbol{L}[z_e,t_m]-\boldsymbol{L}[z_8,t_m]=-1$	$\leq$	0	$\leq$	$0=\boldsymbol{L}[z_e,t_l]-\boldsymbol{L}[z_8,t_l]$	$T_i(z_e)\neq t_m$
13	$T_i(z_e)=t_h,$	$\boldsymbol{L}[z_c,t_m]-\boldsymbol{L}[z_e,t_m]=0$	$\leq$	0	$\leq$	$0=\boldsymbol{L}[z_c,t_h]-\boldsymbol{L}[z_e,t_h]$	$T_i(z_c)\neq t_m$
14	$T_i(z_e)=t_h,$	$\boldsymbol{L}[z_c,t_l]-\boldsymbol{L}[z_e,t_l]=-1$	$\leq$	0	$\leq$	$0=\boldsymbol{L}[z_c,t_h]-\boldsymbol{L}[z_e,t_h]$	$T_i(z_c)\neq t_l$
15	$T_i(z_e)=t_h,$	$\boldsymbol{L}[z_8,t_l]-\boldsymbol{L}[z_e,t_l]=0$	$\leq$	0	$\leq$	$0=\boldsymbol{L}[z_8,t_h]-\boldsymbol{L}[z_e,t_h]$	$T_i(z_8)\neq t_l$
16	$T_i(z_e)=t_m,$	$\boldsymbol{L}[z_c,t_h]-\boldsymbol{L}[z_e,t_h]=0$	$\leq$	0	$\leq$	$0=\boldsymbol{L}[z_c,t_m]-\boldsymbol{L}[z_e,t_m]$	$T_i(z_c)\neq t_h$
17	$T_i(z_e)=t_m,$	$\boldsymbol{L}[z_8,t_h]-\boldsymbol{L}[z_e,t_h]=0$	$\leq$	0	$\leq$	$1=\boldsymbol{L}[z_8,t_m]-\boldsymbol{L}[z_e,t_m]$	$T_i(z_8)\neq t_h$
18	$T_i(z_e)=t_m,$	$\boldsymbol{L}[z_c,t_l]-\boldsymbol{L}[z_e,t_l]=-1$	$\leq$	0	$\leq$	$0=\boldsymbol{L}[z_c,t_m]-\boldsymbol{L}[z_e,t_m]$	$T_i(z_c)\neq t_l$
19	$T_i(z_e)=t_m,$	$\boldsymbol{L}[z_8,t_l]-\boldsymbol{L}[z_e,t_l]=0$	$\leq$	0	$\leq$	$1=\boldsymbol{L}[z_8,t_m]-\boldsymbol{L}[z_e,t_m]$	$T_i(z_8)\neq t_l$
20	$T_i(z_e)=t_l,$	$\boldsymbol{L}[z_8,t_h]-\boldsymbol{L}[z_e,t_h]=0$	$\leq$	0	$\leq$	$0=\boldsymbol{L}[z_8,t_l]-\boldsymbol{L}[z_e,t_l]$	$T_i(z_8)\neq t_h$

This table displays the binding choice restrictions generated by the WARP restriction below

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

when applied to the MTO incentive matrix:

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

Table A.2: Summary of Choice Restrictions generated by applying WARP to the MTO Incentive Matrix

#	Choice Restrictions		
4,5,6	$T_i(z_c) = t_l$	$\Rightarrow$	$T_i(z_e) = t_l$ and $T_i(z_8) \neq t_h$
2,3	$T_i(z_c) = t_m$	$\Rightarrow$	$T_i(z_e) \neq t_h$ and $T_i(z_8) \neq t_h$
16,17,18,19	$T_i(z_e) = t_m$	$\Rightarrow$	$T_i(z_c) = t_m$ and $T_i(z_8) = t_m$
13,14,15	$T_i(z_e) = t_h$	$\Rightarrow$	$T_i(z_c) = t_h$ and $T_i(z_8) \neq t_l$
7,8,9,10	$T_i(z_8) = t_h$	$\Rightarrow$	$T_i(z_c) = t_h$ and $T_i(z_e) = t_h$
11,12	$T_i(z_8) = t_l$	$\Rightarrow$	$T_i(z_e) = t_l$
1	$T_i(z_c) = t_l$	$\Rightarrow$	$T_i(z_e) \neq t_m$
20	$T_i(z_e) = t_l$	$\Rightarrow$	$T_i(z_8) \neq t_h$

A.4 WARP L.1 Subsumes Standard Monotonicity Conditions (10)–(12)

This section shows that Lemma **L.1** is able to subsume and outperform the conditions (10)–(12). The monotonicity condition (10), states that $\mathbf{1}[T_i(z_c) = t_l] \leq \mathbf{1}[T_i(z_e) = t_l]$. It comprise two choice restrictions: $T_i(z_c) = t_l \Rightarrow T_i(z_e) \neq t_h$ and $T_i(z_c) = t_l \Rightarrow T_i(z_e) \neq t_m$. These restrictions can be by applying **L.1** to t_l against t_h, t_m when the IV changes from z_c to z_e :

$$T_i(z_c) = t_l \text{ and } \boldsymbol{L}[z_e, t_h] - \boldsymbol{L}[z_c, t_h] = 0 \leq 1 \leq 1 = \boldsymbol{L}[z_e, t_l] - \boldsymbol{L}[z_c, t_l] \Rightarrow T_i(z_e) \neq t_h.$$
$$T_i(z_c) = t_l \text{ and } \boldsymbol{L}[z_e, t_m] - \boldsymbol{L}[z_c, t_m] = 0 \leq 1 \leq 1 = \boldsymbol{L}[z_e, t_l] - \boldsymbol{L}[z_c, t_l] \Rightarrow T_i(z_e) \neq t_m.$$

The monotonicity condition (11), states that $\mathbf{1}[T_i(z_c) \in \{t_m, t_l\}] \leq \mathbf{1}[T_i(z_8) \in \{t_m, t_l\}]$. It comprise two choice restrictions: $T_i(z_c) = t_l \Rightarrow T_i(z_8) \neq t_h$ and $T_i(z_c) = t_m \Rightarrow T_i(z_8) \neq t_h$. These restrictions can be by applying **L.1** to t_l, t_m against t_h when the IV changes from z_c to z_8 :

$$T_i(z_c) = t_l \text{ and } \boldsymbol{L}[z_8, t_h] - \boldsymbol{L}[z_c, t_h] = 0 \leq 0 \leq 1 = \boldsymbol{L}[z_8, t_l] - \boldsymbol{L}[z_c, t_l] \Rightarrow T_i(z_8) \neq t_h.$$
$$T_i(z_c) = t_m \text{ and } \boldsymbol{L}[z_8, t_h] - \boldsymbol{L}[z_c, t_h] = 0 \leq 0 \leq 1 = \boldsymbol{L}[z_8, t_m] - \boldsymbol{L}[z_c, t_m] \Rightarrow T_i(z_8) \neq t_h.$$

The monotonicity condition (12), states that $\mathbf{1}[T_i(z_e) = t_m] \leq \mathbf{1}[T_i(z_8) = t_m]$. It comprise two choice restrictions: $T_i(z_e) = t_m \Rightarrow T_i(z_8) \neq t_h$ and $T_i(z_e) = t_m \Rightarrow T_i(z_8) \neq t_l$. These restrictions can be by applying **L.1** to t_m against t_h, t_e when the IV changes from z_e to z_8 :

$$T_i(z_e) = t_m \text{ and } \boldsymbol{L}[z_8, t_h] - \boldsymbol{L}[z_e, t_h] = 0 \leq 0 \leq 1 = \boldsymbol{L}[z_8, t_m] - \boldsymbol{L}[z_e, t_m] \Rightarrow T_i(z_e) \neq t_h.$$
$$T_i(z_e) = t_m \text{ and } \boldsymbol{L}[z_8, t_l] - \boldsymbol{L}[z_e, t_l] = 0 \leq 0 \leq 1 = \boldsymbol{L}[z_8, t_m] - \boldsymbol{L}[z_e, t_m] \Rightarrow T_i(z_e) \neq t_m.$$

Lemma **L.1** yields additional choice restrictions that are not subsumed by the monotonicity conditions (10)–(12). For example, equation (69) applies **L.1** to t_m against t_h when the IV changes from z_e to z_c :

$$T_i(z_e) = t_m \text{ and } \boldsymbol{L}[z_c, t_h] - \boldsymbol{L}[z_c, t_h] = 0 \leq 0 = \boldsymbol{L}[z_c, t_m] - \boldsymbol{L}[z_e, t_m] \Rightarrow T_i(z_c) \neq t_h. \quad (69)$$

Equation (69) generates the choice restriction $T_i(z_e) = t_m \Rightarrow T_i(z_c) \neq t_h$. This restriction is not implied by the monotonicity conditions (10)–(12). Nevertheless, the choice restriction is intuitive. Note that neither z_e or z_c offers incentives towards choice t_h or t_m. Thus, if a family chooses t_m under z_e, then the family has no incentives to switch its decision towards t_h under z_c.

A.5 Identification of Counterfactual Outcomes in T.1

Heckman and Pinto (2018) show that for any response matrix $\boldsymbol{R}$ and any subset of response types $\mathcal{S} \subset supp(\boldsymbol{S})$, we have that:

$$E(Y(t)|\boldsymbol{S} \in \mathcal{S}) \quad \text{is identified if and only if} \quad \boldsymbol{b}(\mathcal{S})'(\boldsymbol{K}_t)\boldsymbol{b}(\mathcal{S}) = 0, \tag{70}$$

where:

1. $\boldsymbol{I}$ is the identity matrix,
2. $\boldsymbol{B}_t \equiv \mathbf{1}[\boldsymbol{R} = t]$ is a binary matrix that indicates which elements in $\boldsymbol{R}$ are equal to t,
3. $\boldsymbol{B}_t^+$ is the Moore-Penrose pseudo-inverse of $\boldsymbol{B}_t$,
4. $\boldsymbol{K}_t = (\boldsymbol{I}_{7\times 7} - \boldsymbol{B}_t^+\boldsymbol{B}_t)$ is a symmetric $N_S \times N_S$ matrix,
5. $\boldsymbol{b}(\mathcal{S})$ is the binary vector that indicates which response type belongs to $\mathcal{S}$, namely:

$$\boldsymbol{b}(\mathcal{S}) = [\mathbf{1}[\boldsymbol{s}_{ah} \in \mathcal{S}], \mathbf{1}[\boldsymbol{s}_{am} \in \mathcal{S}], \mathbf{1}[\boldsymbol{s}_{al} \in \mathcal{S}], \mathbf{1}[\boldsymbol{s}_{fc} \in \mathcal{S}], \mathbf{1}[\boldsymbol{s}_{pl} \in \mathcal{S}], \mathbf{1}[\boldsymbol{s}_{pm} \in \mathcal{S}], \mathbf{1}[\boldsymbol{s}_{ph} \in \mathcal{S}]]'.$$

The Moore-Penrose matrix is unique and always exists for any real-valued matrix Magnus and Neudecker (1999). If $E(Y(t)|\boldsymbol{S} \in \mathcal{S})$ is identified, then it can be evaluated by the expression

$$E(Y(t)|\boldsymbol{S} \in \mathcal{S}) = \frac{\boldsymbol{b}(\mathcal{S})'\boldsymbol{B}_t^+ \left(\boldsymbol{Q}_Z(t) \odot \boldsymbol{P}_Z(t)\right)}{\boldsymbol{b}(\mathcal{S})'\boldsymbol{B}_t^+ \boldsymbol{P}_Z(t)}.$$

Consider the following matrices for the choice of high-poverty neighborhood t_h :

$$\boldsymbol{B}_{t_h} = \begin{bmatrix} 1 & 0 & 0 & 1 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \Rightarrow \boldsymbol{B}_{t_h}^+ = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0.5 & 0 & -0.5 \\ 0.5 & 0 & -0.5 \\ 0 & 0 & 0 \\ 0 & -1 & 1 \end{bmatrix} \Rightarrow K_{t_h} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.5 & -0.5 & 0 & 0 \\ 0 & 0 & 0 & -0.5 & 0.5 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \tag{71}$$

Equation (73) presents the binary matrix $\boldsymbol{B}_{t_l} = \mathbf{1}[\boldsymbol{R} = t_l]$ which has the same dimension of the response matrix $\boldsymbol{R}$ in (32) and takes value 1 if the respective element in $\boldsymbol{R}$ is t_l or zero otherwise. The seven columns of $\boldsymbol{B}_{t_l}$ are associated with the respective sequence of response types:

$$\boldsymbol{s}_{ah}, \boldsymbol{s}_{am}, \boldsymbol{s}_{al}, \boldsymbol{s}_{fc}, \boldsymbol{s}_{pl}, \boldsymbol{s}_{pm}, \boldsymbol{s}_{ph}.$$

The equation also shows the pseudo-inverse of $\boldsymbol{B}_{t_l}$, that is $\boldsymbol{B}_{t_l}^+$ and the matrix $(\boldsymbol{I}_{7\times 7} - \boldsymbol{B}_{t_h}^+\boldsymbol{B}_{t_h})$. Theorem (**T.1**) states that the counterfactual mean $E(Y(t_h)|\boldsymbol{S} \in \mathcal{S})$ is identified for the following sets of response types $\mathcal{S} \in \{\{\boldsymbol{s}_{ah}\}, \{\boldsymbol{s}_{ph}\}, \{\boldsymbol{s}_{fc}, \boldsymbol{s}_{pl}\}\}$. The indicator vectors for each subset of response types are:

$$\boldsymbol{b}(\{\boldsymbol{s}_{ah}\}) = [1, 0, 0, 0, 0, 0, 0]',$$
$$\boldsymbol{b}(\{\boldsymbol{s}_{ph}\}) = [0, 0, 0, 0, 0, 0, 1]',$$
$$\boldsymbol{b}(\{\boldsymbol{s}_{fc}, \boldsymbol{s}_{pl}\}) = [0, 0, 0, 1, 1, 0, 0]'.$$

Note that $\boldsymbol{K}_{t_h}$ is symmetric. Its first row/column are zero which implies that $\boldsymbol{b}(\{\boldsymbol{s}_{ah}\})'\boldsymbol{K}_{\boldsymbol{t_h}}\boldsymbol{b}(\{\boldsymbol{s}_{ah}\}) = 0$, thus by (70), $E(Y(t_h)|\boldsymbol{S} = \boldsymbol{s_{ah}})$ is identified. The last row/column of $\boldsymbol{K}_{t_h}$ are zero, which implies that $\boldsymbol{b}(\{\boldsymbol{s}_{ph}\})'\boldsymbol{K}_{\boldsymbol{t_h}}\boldsymbol{b}(\{\boldsymbol{s}_{ph}\}) = 0$, thus by (70), $E(Y(t_h)|\boldsymbol{S} = \boldsymbol{s_{ph}})$ is identified. Lastly, it is easy to see that the sum of the fourth and fifth rows/columns of $\boldsymbol{K}_{t_h}$ are zero, which implies that $\boldsymbol{b}(\{\boldsymbol{s}_{fc}, \boldsymbol{s}_{pl}\})'\boldsymbol{K}_{\boldsymbol{t_h}}\boldsymbol{b}(\{\boldsymbol{s}_{fc}, \boldsymbol{s}_{pl}\}) = 0$, and thereby $E(Y(t_h)|\boldsymbol{S} \in \{\boldsymbol{s}_{fc}, \boldsymbol{s}_{pl})$ is identified.

The matrices for medium-poverty neighborhood t_m are displayed below:

$$\boldsymbol{B}_{t_m} = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \Rightarrow \boldsymbol{B}^+_{t_m} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \\ -0.5 & 0.5 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & -1 \\ -0.5 & 0.5 & 0 \end{bmatrix} \Rightarrow \boldsymbol{K}_{t_m} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.5 & 0 & 0 & -0.5 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -0.5 & 0 & 0 & 0.5 \end{bmatrix} \tag{72}$$

Applying the same analysis of the choice t_h to choice t_m we have that:

$$\begin{aligned} \boldsymbol{b}(\{\boldsymbol{s}_{am}\}) = [0,1,0,0,0,0,0]' &\Rightarrow \boldsymbol{b}(\{\boldsymbol{s}_{am}\})'\boldsymbol{K}_{t_m}\boldsymbol{b}(\{\boldsymbol{s}_{am}\}) = 0 \\ \boldsymbol{b}(\{\boldsymbol{s}_{pm}\}) = [0,0,0,0,0,1,0]' &\Rightarrow \boldsymbol{b}(\{\boldsymbol{s}_{pm}\})'\boldsymbol{K}_{t_m}\boldsymbol{b}(\{\boldsymbol{s}_{pm}\}) = 0 \\ \boldsymbol{b}(\{\boldsymbol{s}_{fc}, \boldsymbol{s}_{ph}\}) = [0,0,0,1,0,0,1]' &\Rightarrow \boldsymbol{b}(\{\boldsymbol{s}_{fc}, \boldsymbol{s}_{ph}\})'\boldsymbol{K}_{t_m}\boldsymbol{b}(\{\boldsymbol{s}_{fc}, \boldsymbol{s}_{ph}\}) = 0. \end{aligned}$$

According to equation (70), we have that $E(Y(t_m)|\boldsymbol{S} = \boldsymbol{s}_{am}), E(Y(t_m)|\boldsymbol{S} = \boldsymbol{s}_{pm})$ and $E(Y(t_m)|\boldsymbol{S} \in \{\boldsymbol{s}_{fc}, \boldsymbol{s}_{ph}\})$ are identified.

The matrices for low-poverty neighborhood t_l are displayed below:

$$\boldsymbol{B}_{t_l} = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 1 & 0 \end{bmatrix} \Rightarrow \boldsymbol{B}^+_{t_l} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & -0.5 & 0.5 \\ -1 & 1 & 0 \\ 0 & -0.5 & 0.5 \\ 0 & 0 & 0 \end{bmatrix} \Rightarrow \boldsymbol{K}_{t_l} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.5 & 0 & -0.5 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -0.5 & 0 & 0.5 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \tag{73}$$

Applying the same analysis of the choice t_h to choice t_m we have that:

$$\begin{aligned} \boldsymbol{b}(\{\boldsymbol{s}_{al}\}) = [0,0,1,0,0,0,0]' &\Rightarrow \boldsymbol{b}(\{\boldsymbol{s}_{al}\})'\boldsymbol{K}_{t_l}\boldsymbol{b}(\{\boldsymbol{s}_{al}\}) = 0 \\ \boldsymbol{b}(\{\boldsymbol{s}_{pl}\}) = [0,0,0,0,1,0,0]' &\Rightarrow \boldsymbol{b}(\{\boldsymbol{s}_{pl}\})'\boldsymbol{K}_{t_l}\boldsymbol{b}(\{\boldsymbol{s}_{pl}\}) = 0 \\ \boldsymbol{b}(\{\boldsymbol{s}_{fc}, \boldsymbol{s}_{pm}\}) = [0,0,0,1,0,1,0]' &\Rightarrow \boldsymbol{b}(\{\boldsymbol{s}_{fc}, \boldsymbol{s}_{pm}\})'\boldsymbol{K}_{t_l}\boldsymbol{b}(\{\boldsymbol{s}_{fc}, \boldsymbol{s}_{pm}\}) = 0. \end{aligned}$$

According to equation (70), we have that $E(Y(t_l)|\boldsymbol{S} = \boldsymbol{s}_{al}), E(Y(t_l)|\boldsymbol{S} = \boldsymbol{s}_{pl})$ and $E(Y(t_l)|\boldsymbol{S} \in \{\boldsymbol{s}_{fc}, \boldsymbol{s}_{pm}\})$ are identified.

A.6 Proof of Theorem T.2 (Identification of Response type Probabilities)

Item (i) of the theorem states that all response type probabilities are identified. This proof stems from equation (9). The matrix version of the equation is given by:

$$\boldsymbol{P}_Z(t) = \boldsymbol{B}_t \cdot \boldsymbol{P}_S; \quad t \in \{t_h, t_m, t_l\}, \tag{74}$$

where $\boldsymbol{P}_Z(t) = [P(T = t|Z = z_c), P(T = t|Z = z_8), P(T = t|Z = z_e)]'$ is the 3×1 vector of propensity scores. It is useful to represent this vector using the expectation of the treatment indicator

$D_t = \mathbf{1}[T = t]$ conditioned on the instrument, that is:

$$\boldsymbol{P}_Z(t) = [E(D_t|Z = z_c), E(D_t|Z = z_8), E(D_t|Z = z_e)]' \tag{75}$$
$$= [P(T = t|Z = z_c), P(T = t|Z = z_8), P(T = t|Z = z_e)]' . \tag{76}$$

The entity $\boldsymbol{P}_S$ in the left-hand side of equation (74) is the vector of response type probabilities defined as:

$$\boldsymbol{P}_S = \begin{bmatrix} P(\boldsymbol{S} = \boldsymbol{s}_{ah}) \\ P(\boldsymbol{S} = \boldsymbol{s}_{am}) \\ P(\boldsymbol{S} = \boldsymbol{s}_{al}) \\ P(\boldsymbol{S} = \boldsymbol{s}_{fc}) \\ P(\boldsymbol{S} = \boldsymbol{s}_{pl}) \\ P(\boldsymbol{S} = \boldsymbol{s}_{pm}) \\ P(\boldsymbol{S} = \boldsymbol{s}_{ph}) \end{bmatrix}. \tag{77}$$

The matrix $\boldsymbol{B}_t$ in (74) is defined as $\boldsymbol{B}_t = \mathbf{1}[\boldsymbol{R} = t]$ which is a binary 3×7 matrix that indicates which elements in $\boldsymbol{R}$ are equal to $t \in \{t_l, t_m, t_h\}$. We can stack equation (74) across neighborhood choices to generate the following equation:

$$\begin{bmatrix} \boldsymbol{P}_Z(t_h) \\ \boldsymbol{P}_Z(t_l) \\ \boldsymbol{P}_Z(t_m) \end{bmatrix} = \boldsymbol{B}_T \cdot \boldsymbol{P}_S, \text{ where } \boldsymbol{B}_T \equiv \begin{bmatrix} \boldsymbol{B}_{t_h} \\ \boldsymbol{B}_{t_m} \\ \boldsymbol{B}_{t_l} \end{bmatrix}. \tag{78}$$

Heckman and Pinto (2018) show that the response type probabilities are point-identified if and only if the column rank of $\boldsymbol{B}_T$ is equal to the number of response types, namely $\text{rank}(\boldsymbol{B}_T) = 7$. The matrix $\boldsymbol{B}_T$ is presented below:

$$\boldsymbol{B}_T \equiv \begin{bmatrix} \boldsymbol{B}_{t_h} \\ \boldsymbol{B}_{t_m} \\ \boldsymbol{B}_{t_l} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 1 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 1 & 0 \end{bmatrix} \tag{79}$$

It is easy to check that the columns are linear independent, which implies that $\text{rank}(\boldsymbol{B}_T) = 7$. The response type probabilities are identified as $\boldsymbol{P}_S = \boldsymbol{B}_T^+[\boldsymbol{P}_Z(t_h)', \boldsymbol{P}_Z(t_l)', \boldsymbol{P}_Z(t_m)']'$, where $\boldsymbol{B}_T^+$ is the Moore-Penrose pseudo-inverse of $\boldsymbol{B}_T$. We can use the fact that $P(T = t_h|Z = z) + P(T = t_m|Z = z) + P(T = t_l|Z = z) = 1$ for each $z \in \{z_c, z_8, z_e\}$ in order to write the vector of response type probabilities as:

$$\begin{bmatrix} P(\boldsymbol{S} = \boldsymbol{s}_{ah}) \\ P(\boldsymbol{S} = \boldsymbol{s}_{am}) \\ P(\boldsymbol{S} = \boldsymbol{s}_{al}) \\ P(\boldsymbol{S} = \boldsymbol{s}_{fc}) \\ P(\boldsymbol{S} = \boldsymbol{s}_{pl}) \\ P(\boldsymbol{S} = \boldsymbol{s}_{pm}) \\ P(\boldsymbol{S} = \boldsymbol{s}_{ph}) \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 & 1 & 0 & -1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & -1 & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} E(D_{t_h}|Z = z_c) \\ E(D_{t_h}|Z = z_8) \\ E(D_{t_h}|Z = z_e) \\ E(D_{t_m}|Z = z_c) \\ E(D_{t_m}|Z = z_8) \\ E(D_{t_m}|Z = z_e) \\ E(D_{t_l}|Z = z_c) \\ E(D_{t_l}|Z = z_8) \\ E(D_{t_l}|Z = z_e) \end{bmatrix}.$$

Item (ii) of the theorem states that all the expected values of baseline variables conditioned on response types are identified. Notationally, we have that $E(\boldsymbol{X}|\boldsymbol{S} = \boldsymbol{s}); \boldsymbol{s} \in \{\boldsymbol{s}_{ah}, \boldsymbol{s}_{am}, \boldsymbol{s}_{al}, \boldsymbol{s}_{fc}, \boldsymbol{s}_{pl}, \boldsymbol{s}_{pm}, \boldsymbol{s}_{ph}\}$

are identified. The proof of this statement stems from equation (8) and the fact that the the treatment T does not cause baseline variables X, therefore $X(t) = X(t')$ for all $t, t' \in \{t_h, t_m, t_l\}$. Note that X play the role of special outcomes for which the IV properties (1)–(3) hold. These baseline variables do not include pre-intervention variables that were used in the randomization protocol of the instrumental variable, such as the intervention sites. Now consider the following derivation:

$$\begin{aligned}
E(X|T=t, Z=z)\, P(T=t|Z=z) &= \sum_{\boldsymbol{s} \in supp(\boldsymbol{S})} \mathbf{1}[T=t|\boldsymbol{S}=\boldsymbol{s}, Z=z]\, E(X(t)|\boldsymbol{S}=\boldsymbol{s})\, P(\boldsymbol{S}=\boldsymbol{s}) \\
E(X|T=t, Z=z)\, P(T=t|Z=z) &= \sum_{\boldsymbol{s} \in supp(\boldsymbol{S})} \mathbf{1}[T=t|\boldsymbol{S}=\boldsymbol{s}, Z=z]\, E(X|\boldsymbol{S}=\boldsymbol{s})\, P(\boldsymbol{S}=\boldsymbol{s}), \\
E(X \cdot D_t|Z=z) &= \sum_{\boldsymbol{s} \in supp(\boldsymbol{S})} \mathbf{1}[T=t|\boldsymbol{S}=\boldsymbol{s}, Z=z]\, E(X|\boldsymbol{S}=\boldsymbol{s})\, P(\boldsymbol{S}=\boldsymbol{s}), \qquad (80)
\end{aligned}$$

where the first equality is due to equation (8), the second equality is due to $X(t) = X(t')$ for all $t, t' \in \{t_h, t_m, t_l\}$, and the last equation rewrites is due to $E(X|T=t, Z=z)\, P(T=t|Z=z) = E(X \cdot D_t|Z=z)$ such that $D_t = \mathbf{1}[T=t]$ denotes the treatment indicator.

The matrix version of equation (80) is given by:

$$\boldsymbol{E}_{XZ}(t) = \boldsymbol{B}_t \cdot \boldsymbol{E}_{XS}; \quad t \in \{t_h, t_m, t_l\}, \qquad (81)$$

where $\boldsymbol{E}_{XZ}(t) = [E(XD_t|Z=z_c), E(XD_t|Z=z_8), E(XD_t|Z=z_e)]'$ is the vector of expectations of the baseline variable conditioned on the instrument and $\boldsymbol{E}_{XS}$ is the vector of expectations of the baseline variable conditioned on the response types times the response type probabilities:

$$\boldsymbol{E}_{XS} = \begin{bmatrix} E(X|\boldsymbol{S}=\boldsymbol{s}_{ah})\, P(\boldsymbol{S}=\boldsymbol{s}_{ah}) \\ E(X|\boldsymbol{S}=\boldsymbol{s}_{am})\, P(\boldsymbol{S}=\boldsymbol{s}_{am}) \\ E(X|\boldsymbol{S}=\boldsymbol{s}_{al})\, P(\boldsymbol{S}=\boldsymbol{s}_{al}) \\ E(X|\boldsymbol{S}=\boldsymbol{s}_{fc})\, P(\boldsymbol{S}=\boldsymbol{s}_{fc}) \\ E(X|\boldsymbol{S}=\boldsymbol{s}_{pl})\, P(\boldsymbol{S}=\boldsymbol{s}_{pl}) \\ E(X|\boldsymbol{S}=\boldsymbol{s}_{pm})\, P(\boldsymbol{S}=\boldsymbol{s}_{pm}) \\ E(X|\boldsymbol{S}=\boldsymbol{s}_{ph})\, P(\boldsymbol{S}=\boldsymbol{s}_{ph}) \end{bmatrix}.$$

Note that the equation (81), $\boldsymbol{E}_{XZ}(t) = \boldsymbol{B}_t \cdot \boldsymbol{E}_{XS}$ is closely related to equation (74), $\boldsymbol{P}_Z(t) = \boldsymbol{B}_t \cdot \boldsymbol{P}_S$, Indeed, equation (81) can be obtained by replacing each entry $E(D_t|Z=z)$ in $\boldsymbol{P}_Z(t)$ by $E(XD_t|Z=z)$ and each entry $P(\boldsymbol{S}=\boldsymbol{s})$ in $\boldsymbol{P}_S$ by $E(X|\boldsymbol{S}=\boldsymbol{s})\, P(\boldsymbol{S}=\boldsymbol{s})$. If we follow the same argument of item (i) of this proof, we have that $E(X|\boldsymbol{S}=\boldsymbol{s})\, P(\boldsymbol{S}=\boldsymbol{s})$ for each response type are identified by:

$$\begin{bmatrix} E(X|\boldsymbol{S}=\boldsymbol{s}_{ah})\, P(\boldsymbol{S}=\boldsymbol{s}_{ah}) \\ E(X|\boldsymbol{S}=\boldsymbol{s}_{am})\, P(\boldsymbol{S}=\boldsymbol{s}_{am}) \\ E(X|\boldsymbol{S}=\boldsymbol{s}_{al})\, P(\boldsymbol{S}=\boldsymbol{s}_{al}) \\ E(X|\boldsymbol{S}=\boldsymbol{s}_{fc})\, P(\boldsymbol{S}=\boldsymbol{s}_{fc}) \\ E(X|\boldsymbol{S}=\boldsymbol{s}_{pl})\, P(\boldsymbol{S}=\boldsymbol{s}_{pl}) \\ E(X|\boldsymbol{S}=\boldsymbol{s}_{pm})\, P(\boldsymbol{S}=\boldsymbol{s}_{pm}) \\ E(X|\boldsymbol{S}=\boldsymbol{s}_{ph})\, P(\boldsymbol{S}=\boldsymbol{s}_{ph}) \end{bmatrix} = \underbrace{\frac{1}{9}\begin{bmatrix} 1 & 7 & 1 & 1 & -2 & 1 & 1 & -2 & 1 \\ 1 & 1 & -2 & 1 & 1 & 7 & 1 & 1 & -2 \\ -2 & 1 & 1 & -2 & 1 & 1 & 7 & 1 & 1 \\ 3 & 3 & -6 & -6 & 3 & 3 & 3 & -6 & 3 \\ 3 & -3 & 0 & 3 & -3 & 0 & -6 & 6 & 0 \\ -3 & 0 & 3 & 6 & 0 & -6 & -3 & 0 & 3 \\ 0 & -6 & 6 & 0 & 3 & -3 & 0 & 3 & -3 \end{bmatrix}}_{\boldsymbol{B}_T^+} \begin{bmatrix} E(XD_{t_h}|Z=z_c) \\ E(XD_{t_h}|Z=z_8) \\ E(XD_{t_h}|Z=z_e) \\ E(XD_{t_m}|Z=z_c) \\ E(XD_{t_m}|Z=z_8) \\ E(XD_{t_m}|Z=z_e) \\ E(XD_{t_l}|Z=z_c) \\ E(XD_{t_l}|Z=z_8) \\ E(XD_{t_l}|Z=z_e) \end{bmatrix}.$$

The equation above identifies $E(X|\boldsymbol{S}=\boldsymbol{s})\, P(\boldsymbol{S}=\boldsymbol{s})$ for all response types. Recall that the response type probabilities are identified in item (i). Therefore, $E(X|\boldsymbol{S}=\boldsymbol{s})$ are identified for all response types.

A.7 Proof of Proposition P.4

According to equation 6, the TOT parameter that compares the experimental and control groups is given by:

$$TOT_e = \frac{E(Y|Z = z_e) - E(Y|Z = z_c)}{P(C_e = 1|Z = z_e)},$$

where C_e indicates if the experimental voucher is used. We can rewrite the terms in the in the numerator of TOT_e as:

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

We can then use equation (8) and the response matrix (29) to express each term $E(Y|T = t, Z = z)P(T = t|Z = z)$ as a sum of counterfactual outcomes conditioned on response types:

$$\begin{aligned} E(Y|Z = z_e) = {} & E(Y(t_h)|\boldsymbol{S} = \boldsymbol{s}_{ah})P(\boldsymbol{S} = \boldsymbol{s}_{ah}) + E(Y(t_m)|\boldsymbol{S} = \boldsymbol{s}_{am})P(\boldsymbol{S} = \boldsymbol{s}_{am}) \\ & + E(Y(t_l)|\boldsymbol{S} = \boldsymbol{s}_{al})P(\boldsymbol{S} = \boldsymbol{s}_{al}) + E(Y(t_l)|\boldsymbol{S} = \boldsymbol{s}_{fc})P(\boldsymbol{S} = \boldsymbol{s}_{fc}) \\ & + E(Y(t_l)|\boldsymbol{S} = \boldsymbol{s}_{pl})P(\boldsymbol{S} = \boldsymbol{s}_{pl}) + E(Y(t_l)|\boldsymbol{S} = \boldsymbol{s}_{pm})P(\boldsymbol{S} = \boldsymbol{s}_{pm}) + E(Y(t_h)|\boldsymbol{S} = \boldsymbol{s}_{ph})P(\boldsymbol{S} = \boldsymbol{s}_{ph}), \end{aligned}$$

$$\begin{aligned} E(Y|Z = z_c) = {} & E(Y(t_h)|\boldsymbol{S} = \boldsymbol{s}_{ah})P(\boldsymbol{S} = \boldsymbol{s}_{ah}) + E(Y(t_m)|\boldsymbol{S} = \boldsymbol{s}_{am})P(\boldsymbol{S} = \boldsymbol{s}_{am}) \\ & + E(Y(t_l)|\boldsymbol{S} = \boldsymbol{s}_{al})P(\boldsymbol{S} = \boldsymbol{s}_{al}) + E(Y(t_h)|\boldsymbol{S} = \boldsymbol{s}_{fc})P(\boldsymbol{S} = \boldsymbol{s}_{fc}) \\ & + E(Y(t_h)|\boldsymbol{S} = \boldsymbol{s}_{pl})P(\boldsymbol{S} = \boldsymbol{s}_{pl}) + E(Y(t_m)|\boldsymbol{S} = \boldsymbol{s}_{pm})P(\boldsymbol{S} = \boldsymbol{s}_{pm}) + E(Y(t_l)|\boldsymbol{S} = \boldsymbol{s}_{ph})P(\boldsymbol{S} = \boldsymbol{s}_{ph}). \end{aligned}$$

The voucher effect $E(Y|Z = z_e) - E(Y|Z = z_c)$ can be rewritten as:

$$\begin{aligned} E(Y|Z = z_e) - E(Y|Z = z_c) = {} & E(Y(t_l) - Y(t_h)|\boldsymbol{S} = \boldsymbol{s}_{fc})P(\boldsymbol{S} = \boldsymbol{s}_{fc}) \\ & + E(Y(t_l) - Y(t_h)|\boldsymbol{S} = \boldsymbol{s}_{pl})P(\boldsymbol{S} = \boldsymbol{s}_{pl}) + E(Y(t_l) - Y(t_m)|\boldsymbol{S} = \boldsymbol{s}_{pm})P(\boldsymbol{S} = \boldsymbol{s}_{pm}), \\ = {} & E(Y(t_l) - Y(t_h)|\boldsymbol{S} \in \{\boldsymbol{s}_{fc}, \boldsymbol{s}_{pl}\})P(\boldsymbol{S} \in \{\boldsymbol{s}_{fc}, \boldsymbol{s}_{pl}\}) + E(Y(t_l) - Y(t_m)|\boldsymbol{S} = \boldsymbol{s}_{pm})P(\boldsymbol{S} = \boldsymbol{s}_{pm}). \end{aligned}$$

Thus TOT effect that compares experimental versus control groups is given by:

$$\begin{aligned} TOT_e &= \frac{E(Y(t_l) - Y(t_h)|\boldsymbol{S} \in \{\boldsymbol{s}_{fc}, \boldsymbol{s}_{pl}\})P(\boldsymbol{S} \in \{\boldsymbol{s}_{fc}, \boldsymbol{s}_{pl}\}) + E(Y(t_l) - Y(t_m)|\boldsymbol{S} = \boldsymbol{s}_{pm})P(\boldsymbol{S} = \boldsymbol{s}_{pm})}{P(\boldsymbol{S} \in \{\boldsymbol{s}_{fc}, \boldsymbol{s}_{pl}, \boldsymbol{s}_{pm}\})} \cdot \xi_e, \\ s.t. \xi_e &= \frac{P(\boldsymbol{S} \in \{\boldsymbol{s}_{fc}, \boldsymbol{s}_{pl}, \boldsymbol{s}_{pm}\})}{P(C_e = 1|Z = z_e)}. \end{aligned}$$

We can apply analogous arguments to examine the causal content of the the TOT parameter that compares the Section 8 and control groups. The TOT parameter is given by:

$$TOT_8 = \frac{E(Y|Z = z_8) - E(Y|Z = z_c)}{P(C_8 = 1|Z = z_8)},$$

where C_8 indicates if the Section 8 voucher is used. The expectation $E(Y|Z = z_8)$ can be expressed as:

$$
\begin{aligned}
E(Y|Z=z_8) = &\, E(Y(t_h)|\boldsymbol{S}=\boldsymbol{s}_{ah})P(\boldsymbol{S}=\boldsymbol{s}_{ah}) + E(Y(t_m)|\boldsymbol{S}=\boldsymbol{s}_{am})P(\boldsymbol{S}=\boldsymbol{s}_{am}) \\
&+ E(Y(t_l)|\boldsymbol{S}=\boldsymbol{s}_{al})P(\boldsymbol{S}=\boldsymbol{s}_{al}) + E(Y(t_m)|\boldsymbol{S}=\boldsymbol{s}_{fc})P(\boldsymbol{S}=\boldsymbol{s}_{fc}) \\
&+ E(Y(t_l)|\boldsymbol{S}=\boldsymbol{s}_{pl})P(\boldsymbol{S}=\boldsymbol{s}_{pl}) + E(Y(t_m)|\boldsymbol{S}=\boldsymbol{s}_{pm})P(\boldsymbol{S}=\boldsymbol{s}_{pm}) + E(Y(t_m)|\boldsymbol{S}=\boldsymbol{s}_{ph})P(\boldsymbol{S}=\boldsymbol{s}_{ph}).
\end{aligned}
$$

The voucher effect $E(Y|Z=z_8) - E(Y|Z=z_c)$ can be rewritten as:

$$
\begin{aligned}
E(Y|Z=z_8) - E(Y|Z=z_c) = &\, E(Y(t_m)-Y(t_h)|\boldsymbol{S}=\boldsymbol{s}_{fc})P(\boldsymbol{S}=\boldsymbol{s}_{fc}) \\
&+ E(Y(t_l)-Y(t_h)|\boldsymbol{S}=\boldsymbol{s}_{pl})P(\boldsymbol{S}=\boldsymbol{s}_{pl}) + E(Y(t_m)-Y(t_h)|\boldsymbol{S}=\boldsymbol{s}_{ph})P(\boldsymbol{S}=\boldsymbol{s}_{ph}), \\
&= E(Y(t_l)-Y(t_h)|\boldsymbol{S}=\boldsymbol{s}_{pl})P(\boldsymbol{S}=\boldsymbol{s}_{pl}) + E(Y(t_m)-Y(t_h)|\boldsymbol{S}\in\{\boldsymbol{s}_{fc},\boldsymbol{s}_{ph}\})P(\boldsymbol{S}\in\{\boldsymbol{s}_{fc},\boldsymbol{s}_{ph}\}).
\end{aligned}
$$

Thus TOT effect that compares Section 8 versus control groups is given by:

$$
\begin{aligned}
TOT_8 &= \frac{E(Y(t_l)-Y(t_h)|\boldsymbol{S}=\boldsymbol{s}_{pl})P(\boldsymbol{S}=\boldsymbol{s}_{pl}) + E(Y(t_m)-Y(t_h)|\boldsymbol{S}\in\{\boldsymbol{s}_{fc},\boldsymbol{s}_{ph}\})P(\boldsymbol{S}\in\{\boldsymbol{s}_{fc},\boldsymbol{s}_{ph}\}}{P(\boldsymbol{S}\in\{\boldsymbol{s}_{fc},\boldsymbol{s}_{pl},\boldsymbol{s}_{ph}\})}\cdot\xi_8, \\
s.t.\xi_8 &= \frac{P(\boldsymbol{S}\in\{\boldsymbol{s}_{fc},\boldsymbol{s}_{pl},\boldsymbol{s}_{ph}\})}{P(C_8=1|Z=z_8)}.
\end{aligned}
$$

A.8 Proof of Theorem T.3

Table A.3 presents which response types are eliminated by the monotonicity conditions in **T.3**. The response types that survive the elimination process are precisely those displayed in the response matrix 29. It remains to prove that no other set of monotonicity conditions of the type described by unordered monotonicity 45 generates the same response matrix. To do so, it suffices to show that a change the direction of each the monotonicity conditions violates a pattern of counterfactual choices displayed in the response matrix. For convenience, the response matrix is presented below.

$$
\boldsymbol{R} = \begin{array}{c} \begin{matrix} \boldsymbol{s}_{ah} & \boldsymbol{s}_{am} & \boldsymbol{s}_{al} & \boldsymbol{s}_{fc} & \boldsymbol{s}_{pl} & \boldsymbol{s}_{pm} & \boldsymbol{s}_{ph} \end{matrix} \\ \begin{bmatrix} t_h & t_m & t_l & t_h & t_h & t_m & t_h \\ t_h & t_m & t_l & t_m & t_l & t_m & t_m \\ t_h & t_m & t_l & t_l & t_l & t_l & t_h \end{bmatrix} \end{array} \quad \begin{matrix} T_i(z_c) \\ T_i(z_8) \\ T_i(z_e) \end{matrix}
$$

The items below indicate which counterfactual choice pattern is violated if we reverse the direction of each of the monotonicity conditions in **T.3**.

1. $\mathbf{1}[T_i(z_c)=t_h] \leq \mathbf{1}[T_i(z_8)=t_h]$ violates the choice pattern in response type $\boldsymbol{s}_{ph}$ when the instrument switches from z_c to z_8.

2. $\mathbf{1}[T_i(z_8)=t_h] \geq \mathbf{1}[T_i(z_e)=t_h]$ violates the choice pattern in response type $\boldsymbol{s}_{ph}$ when the instrument switches from z_e to z_8.

3. $\mathbf{1}[T_i(z_e)=t_h] \geq \mathbf{1}[T_i(z_c)=t_h]$ violates the choice pattern in response type $\boldsymbol{s}_{fc}$ when the instrument switches from z_c to z_e.

4. $\mathbf{1}[T_i(z_c)=t_m] \geq \mathbf{1}[T_i(z_8)=t_m]$ violates the choice pattern in response type $\boldsymbol{s}_{fc}$ when the instrument switches from z_8 to z_c.

5. $\mathbf{1}[T_i(z_8)=t_m] \leq \mathbf{1}[T_i(z_e)=t_m]$ violates the choice pattern in response type $\boldsymbol{s}_{pm}$ when the instrument switches from z_8 to z_e.

6. $\mathbf{1}[T_i(z_e) = t_m] \geq \mathbf{1}[T_i(z_c) = t_m]$ violates the choice pattern in response type $\boldsymbol{s}_{pm}$ when the instrument switches from z_c to z_e.

7. $\mathbf{1}[T_i(z_c) = t_l] \geq \mathbf{1}[T_i(z_8) = t_l]$ violates the choice pattern in response type $\boldsymbol{s}_{pl}$ when the instrument switches from z_8 to z_c.

8. $\mathbf{1}[T_i(z_8) = t_l] \geq \mathbf{1}[T_i(z_e) = t_l]$ violates the choice pattern in response type $\boldsymbol{s}_{pm}$ when the instrument switches from z_e to z_8.

9. $\mathbf{1}[T_i(z_e) = t_l] \leq \mathbf{1}[T_i(z_c) = t_l]$ violates the choice pattern in response type $\boldsymbol{s}_{pl}$ when the instrument switches from z_e to z_c.

A.9 Proof of Theorem T.4

Item (i) of the theorem states that the choice indicator $D_t = \mathbf{1}[T = t]$ can be expressed as the separable equation $D_t = \mathbf{1}[P_t(Z) \geq U_t]$ where U_t is an unobserved variables that is uniformity distributed in $[0, 1]$. The proof consisting in constructing a variable U_t such that $D_t = \mathbf{1}[P_t(Z) \geq U_t]$ w.p.1. and to show that the constructed variable has uniform distribution.

Remark 1.1. The theorem describes the IV model using the Rubin-Holland causal model, which employs the language of potential outcomes in (1)–(3) *to define IV model. The main advantage of using the Rubin-Holland causal model is its simplicity. However, this causal framework has a major drawback. The language of potential outcomes severely limits the interpretation of the IV model, and, in particular, the interpretation of variable U_t. Appendix D uses structural equation to equivalently describe the IV model. I refer to Appendix D for the interpretation of the unobserved variables in the IV model.*

The first step to proving item (i) is to show the choice indicator D_t can be expressed as a threshold crossing indicator. This fact stems from the triangular property of the MTO response matrix, namely, each binary matrix $\boldsymbol{B}_t = \mathbf{1}[\boldsymbol{R} = t]; t \in \{t_h, t_m, t_l\}$ can be written as a lower triangular matrix as displayed in equations (35), (42), and (43) of Section 5.

Consider the case of low poverty neighborhoods t_l as our leading example. The triangular response matrix for t_l in (35) is displayed bellow for our convenience:

$$\boldsymbol{R}_l = \begin{array}{c} \begin{matrix} \boldsymbol{s}_{al} & \boldsymbol{s}_{pl} & \boldsymbol{s}_{fc} & \boldsymbol{s}_{pm} & \boldsymbol{s}_{ah} & \boldsymbol{s}_{am} & \boldsymbol{s}_{pl} \end{matrix} \\ \begin{bmatrix} t_l & t_h & t_h & t_m & t_h & t_m & t_h \\ t_l & t_l & t_m & t_m & t_h & t_m & t_m \\ t_l & t_l & t_l & t_l & t_h & t_m & t_h \end{bmatrix} \end{array} \begin{matrix} z_c \\ z_8 \\ z_e \end{matrix}$$

Let $\boldsymbol{B}_{t_l} \equiv \mathbf{1}[\boldsymbol{R}_l = t_l]$ be the binary matrix that takes 1 if the respective element in $\boldsymbol{R}_l$ is t_l, that is:

$$\boldsymbol{B}_{t_l} \equiv \mathbf{1}[\boldsymbol{R}_l = t_l] = \begin{array}{c} \begin{matrix} \boldsymbol{s}_{al} & \boldsymbol{s}_{pl} & \boldsymbol{s}_{fc} & \boldsymbol{s}_{pm} & \boldsymbol{s}_{ah} & \boldsymbol{s}_{am} & \boldsymbol{s}_{pl} \end{matrix} \\ \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 \end{bmatrix} \end{array} \begin{matrix} z_c \\ z_8 \\ z_e \end{matrix}$$

It is useful to relabel the indexes of the z-values according to increasing values of row-sums and

Table A.3: Elimination of MTO Response types by the Unordered Monotonicity Condition

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
$T_i(z_c)$	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
$T_i(z_8)$	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
$T_i(z_e)$	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 $S_i = [T_i(z_c)\ ,\ T_i(z_8)\ ,\ T_i(z_e)]$ 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 $T_i(z_c), T_i(z_8)$ and $T_i(z_e)$ respectively. Columns present all the values of response type as choices range over $supp(T) = \{t_h, t_m, t_l\}$. 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	$\mathbf{1}[T_i(z_c) = t_h]$	$\geq$	$\mathbf{1}[T_i(z_8) = t_h]$
Monotonicity Relation 2	(z_8, z_e)	t_h	$\mathbf{1}[T_i(z_8) = t_h]$	$\leq$	$\mathbf{1}[T_i(z_e) = t_h]$
Monotonicity Relation 3	(z_e, z_c)	t_h	$\mathbf{1}[T_i(z_c) = t_h]$	$\geq$	$\mathbf{1}[T_i(z_e) = t_h]$
Monotonicity Relation 4	(z_c, z_8)	t_m	$\mathbf{1}[T_i(z_c) = t_m]$	$\leq$	$\mathbf{1}[T_i(z_8) = t_m]$
Monotonicity Relation 5	(z_8, z_e)	t_m	$\mathbf{1}[T_i(z_8) = t_m]$	$\geq$	$\mathbf{1}[T_i(z_e) = t_m]$
Monotonicity Relation 6	(z_e, z_c)	t_m	$\mathbf{1}[T_i(z_e) = t_m]$	$\leq$	$\mathbf{1}[T_i(z_c) = t_m]$
Monotonicity Relation 7	(z_c, z_8)	t_l	$\mathbf{1}[T_i(z_c) = t_l]$	$\leq$	$\mathbf{1}[T_i(z_8) = t_l]$
Monotonicity Relation 8	(z_8, z_e)	t_l	$\mathbf{1}[T_i(z_8) = t_l]$	$\leq$	$\mathbf{1}[T_i(z_e) = t_l]$
Monotonicity Relation 9	(z_e, z_c)	t_l	$\mathbf{1}[T_i(z_e) = t_l]$	$\geq$	$\mathbf{1}[T_i(z_c) = t_l]$

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.

relabel the indexes of the response types according to decreasing values of the columns-sums:

$$\boldsymbol{B}_{t_l} \equiv \begin{array}{c} \\ \end{array}\overset{\begin{array}{ccccccc} \boldsymbol{s}_1 & \boldsymbol{s}_2 & \boldsymbol{s}_3 & \boldsymbol{s}_4 & \boldsymbol{s}_5 & \boldsymbol{s}_6 & \boldsymbol{s}_7 \end{array}}{\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 \end{bmatrix}} \begin{array}{c} z_1 \\ z_2 \\ z_3 \end{array} \tag{82}$$

Equation (8) enable us to relate propensity scores and response type probabilities by the following equation:

$$P(T=t|Z=z) = \sum_{\boldsymbol{s} \in supp(\boldsymbol{S})} \mathbf{1}[T=t|\boldsymbol{S}=\boldsymbol{s}, Z=z]\, P(\boldsymbol{S}=\boldsymbol{s}). \tag{83}$$

We can use the matrix version of equation (83) to relate propensity scores and response type probabilities as following:

$$\begin{bmatrix} P_{t_l}(z_1) \\ P_{t_l}(z_2) \\ P_{t_l}(z_3) \end{bmatrix} = \underbrace{\overset{\begin{array}{ccccccc} \boldsymbol{s}_1 & \boldsymbol{s}_2 & \boldsymbol{s}_3 & \boldsymbol{s}_4 & \boldsymbol{s}_5 & \boldsymbol{s}_6 & \boldsymbol{s}_7 \end{array}}{\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 \end{bmatrix}}}_{\text{Binary matrix } B_{t_l} = \mathbf{1}[\boldsymbol{R}_l = t_l]} \begin{bmatrix} P(\boldsymbol{S}=\boldsymbol{s}_1) \\ P(\boldsymbol{S}=\boldsymbol{s}_2) \\ P(\boldsymbol{S}=\boldsymbol{s}_3) \\ P(\boldsymbol{S}=\boldsymbol{s}_4) \\ P(\boldsymbol{S}=\boldsymbol{s}_5) \\ P(\boldsymbol{S}=\boldsymbol{s}_6) \\ P(\boldsymbol{S}=\boldsymbol{s}_7) \end{bmatrix}. \tag{84}$$

Equation (84) generates the following relationships between the relabeled propensity scores and response type probabilities:

$$P_{t_l}(z_1) = \sum_{j=1}^{1} P(\boldsymbol{S}=\boldsymbol{s}_j) \tag{85}$$

$$P_{t_l}(z_2) = \sum_{j=1}^{2} P(\boldsymbol{S}=\boldsymbol{s}_j) \tag{86}$$

$$P_{t_l}(z_3) = \sum_{j=1}^{4} P(\boldsymbol{S}=\boldsymbol{s}_j) \tag{87}$$

The key property that arises from the triangular property of matrix (82) is that we can express each of the elements of B_{t_l} in (82) as an indicator of an inequality between the propensity score and the sum of the response type probabilities. To see this property, let $\boldsymbol{B}_{t_l}[z_i, \boldsymbol{s}_j]; (i,j) \in \{1,2,3\} \times \{1,...,7\}$ denotes the elements of matrix $\boldsymbol{B}_{t_l}$. The matrix equation (84) renders the following properties of these elements:

$$\boldsymbol{B}_{t_l}[z_i, \boldsymbol{s}_j] = 1 \quad \Leftrightarrow \quad P_l(z_i) \geq \sum_{j'=1}^{j} P(\boldsymbol{S}=\boldsymbol{s}_{j'}), \tag{88}$$

$$\text{or} \quad \boldsymbol{B}_{t_l}[z_i, \boldsymbol{s}_j] = 0 \quad \Leftrightarrow \quad P_l(z_i) < \sum_{j'=1}^{j} P(\boldsymbol{S}=\boldsymbol{s}_{j'}). \tag{89}$$

Note that $\boldsymbol{B}_{t_l}[z_i, \boldsymbol{s}_j] = \mathbf{1}[T = t_l | Z = z_i, \boldsymbol{S} = \boldsymbol{s}_j]$, thus, we can express the choice indicator as:

$$\mathbf{1}[T = t_l | Z = z_i, \boldsymbol{S} = \boldsymbol{s}_j] = \mathbf{1}\left[P_l(z_i) \geq \sum_{j'=1}^{j} P(\boldsymbol{S} = \boldsymbol{s}_{j'})\right] \text{ for } (i,j) \in \{1,2,3\} \times \{1,...,7\}. \tag{90}$$

Equation 90 shows that the choice indicator (right-hand side) can be expressed as a the indicator of a separable inequality that compares the propensity score $P_l(z_i)$ with the sum of response type probabilities.

There are several ways to construct a variable $U_{t_l} \sim Unif[0,1]$ such that $D_{t_l} = \mathbf{1}[P_t(Z) \geq U_{t_l}]$. For instance, let $U_1, ..., U_7$ be i.i.d. random variables uniformly distributed in $[0,1]$. Let U_{t_l} be defined as:

$$U_{t_l} = \sum_{j=1}^{7} \mathbf{1}[\boldsymbol{S} = \boldsymbol{s}_j] \cdot \left(\sum_{j'=0}^{j-1} P(\boldsymbol{S} = \boldsymbol{s}_{j-1}) + U_j \cdot P(\boldsymbol{S} = \boldsymbol{s}_j)\right), \text{ where } P(\boldsymbol{S} = \boldsymbol{s}_0) \equiv 0. \tag{91}$$

Variable U_{t_l} has a uniform distribution in $[\sum_{j'=0}^{j-1} P(\boldsymbol{S} = \boldsymbol{s}_{j'}), \sum_{j'=0}^{j} P(\boldsymbol{S} = \boldsymbol{s}_{j'})]$ conditional on $\boldsymbol{S} = \boldsymbol{s}_j$. Unconditionally, variable U_{t_l} has a uniform distribution in $[0,1]$. Let the indicator variable be defined as:

$$\tilde{D}_{t_l} = \mathbf{1}[P_{t_l}(Z) \geq U_{t_l}]. \tag{92}$$

Note that:

$$\left(\tilde{D}_{t_l} | Z = z_i, \boldsymbol{S} = \boldsymbol{s}_i\right) = \mathbf{1}[P_{t_l}(z_i) \geq U_{t_l,j}],$$
$$\text{where} \quad U_{t_l,j} \sim Unif\left[\sum_{j'=0}^{j-1} P(\boldsymbol{S} = \boldsymbol{s}_{j'}), \sum_{j'=0}^{j} P(\boldsymbol{S} = \boldsymbol{s}_{j'})\right].$$

According to (88)–(89), we have that $(\tilde{D}_{t_l} | Z = z_i, \boldsymbol{S} = \boldsymbol{s}_j) = \boldsymbol{B}_{t_l}[z_i, \boldsymbol{s}_j]$ for all (i,j) $\in \{1,2,3\} \times \{1,...,7\}$. Moreover, we have that $D_{t_l} \equiv \mathbf{1}[T = t_l]$, thus we can combine all results into:

$$(D_{t_l} | Z = z_i, \boldsymbol{S} = \boldsymbol{s}_j) = \mathbf{1}[T = t_l | Z = z_i, \boldsymbol{S} = \boldsymbol{s}_j] = \boldsymbol{B}_{t_l}[z_i, \boldsymbol{s}_j] = (\tilde{D}_{t_l} | Z = z_i, \boldsymbol{S} = \boldsymbol{s}_j).$$

In particular, we have that:

$$D_{t_l} = D_{t_l}(Z, \boldsymbol{S}) = (\tilde{D}_{t_l} | Z, \boldsymbol{S}) = \mathbf{1}[P_{t_l}(Z) \geq U_{t_l}].$$

As mentioned, the rationale for establishing that $D_{t_l} = \mathbf{1}[P_{t_l}(Z) \geq U_{t_l}]$ holds stems from the triangular property of the MTO matrix for choice t_l. The variables U_{t_h}, U_{t_m} can be constructed in the same fashion since the triangular property of the MTO matrix holds for t_h and t_m.

To prove the item (ii) of the theorem, first note that the exogeneity condition (2), $Z \perp\!\!\!\perp (Y(t), T(z))$, implies that $Z \perp\!\!\!\perp \boldsymbol{S}$, but U_t is a function of only $\boldsymbol{S}$, which implies that $Z \perp\!\!\!\perp (U_t, Y(t))$, and thereby $Z \perp\!\!\!\perp Y(t)|U_t$, holds. Thus, we have that:

$$
\begin{aligned}
E(Y \cdot \mathbf{1}[T=t]|Z=z) &= E(Y(t)\cdot D_t|Z=z) && (93)\\
&= E(Y(t)\cdot \mathbf{1}[P_{t_l}(Z) \geq U_{t_l}]|Z=z) && (94)\\
&= E(Y(t)\cdot \mathbf{1}[P_{t_l}(z) \geq U_{t_l}]) && (95)\\
&= \int_0^{P_{t_l}(z)} E(Y(t)|U_{t_l}=u)du, && (96)
\end{aligned}
$$

where the second equality uses $D_t = \mathbf{1}[P_t(Z) \geq U_{t_l}]$, the third equality is due to $Z \perp\!\!\!\perp (Y(t), U_t)$, and the fourth equality is due to $U_{t_l} \sim Unif[0,1]$.

Let $z, z' \in supp(Z)$ such that $P_t(z') > P_t(z)$. Equation (96) enable us to write:

$$
\begin{aligned}
E(YD_t|Z=z') - E(YD_t|Z=z) &= E(Y(t)\mathbf{1}[P_{t_l}(z') \geq U_{t_l}]) - E(Y\mathbf{1}[P_{t_l}(z) \geq U_{t_l}])\\
&= E(Y(t)\cdot(\mathbf{1}[P_{t_l}(z') \geq U_{t_l}] - \mathbf{1}[P_{t_l}(z) \geq U_{t_l}]))\\
&= E(Y(t)\cdot(\mathbf{1}[P_{t_l}(z') \geq U_{t_l} \geq P_{t_l}(z)]))\\
&= \int_{P_{t_l}(z)}^{P_{t_l}(z')} E(Y(t)|U_t=u)du
\end{aligned}
$$

Therefore we have that:

$$
\frac{E(YD_t|Z=z') - E(YD_t|Z=z)}{P_t(z') - P_t(z)} = \frac{\int_{P_t(z)}^{P_t(z')} E(Y(t)|U_t=u)du}{P_t(z') - P_t(z)}.
$$

B Defining Neighborhood Choices

The neighborhood choices are defined according to the eligibility criteria of MTO vouchers:

- Low poverty neighborhood (t_l) are the neighborhoods whose poverty level is below 10% according to the 1990 U.S. Census.
- High poverty neighborhood (t_h) are the housing projects targeted by the MTO experiment.
- Medium poverty neighborhood (t_m) are the remaining neighborhoods.

Each choice refers to the neighborhood decision at the beginning of the intervention. Thus, each neighborhood choice indicates the initial family decision of neighborhood relocation but also eventual subsequent moves made by the family.

Families using the vouchers were supposed to move from housing projects within six months of the voucher assignment. However, this rule was not strictly enforced: 17% of the families that used the Section 8 voucher and 36% of families that used the experimental voucher took more than 6 months to move. Thus the neighborhood choice depends on the voucher utilization, the neighborhood poverty level and also on the time that the family took to relocate.

It is useful to classify the families into three groups: stayers, compliers, and self-movers. *Stayers* are families that had not moved from their original housing projects since the intervention onset until the time of the interim evaluation in 2002. *Compliers* are families that use the experimental or Section 8 vouchers to relocate. *Self-movers* are families that had moved at the time of the interim evaluation without using the voucher. Table A.4 presents the distribution of these family types across sites. Around 20% of families that receive vouchers and 30% of the control families stayed in their original dwellings by the time of the interim evaluation. Self-movers totals 36% of experimental families and 24% of Section 8 families in 2002.

Table A.4: Relocation Rates by Site at the Time of the Interim Evaluation in 2002

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 %
			Compliers	818	47%	146	58%	168	46%	155	34%	167	67%	182	45%
Experimental	1729	41%	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%
			Compliers	716	59%	135	72%	129	48%	134	66%	130	77%	188	49%
Section 8	1209	28%	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 relocation of families by voucher assignment and site in 2002. MTO families 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 interim evaluation in 2002; (3) stayers – families that had not moved since intervention onset in 1994–1998 until the interim evaluation in 2002.

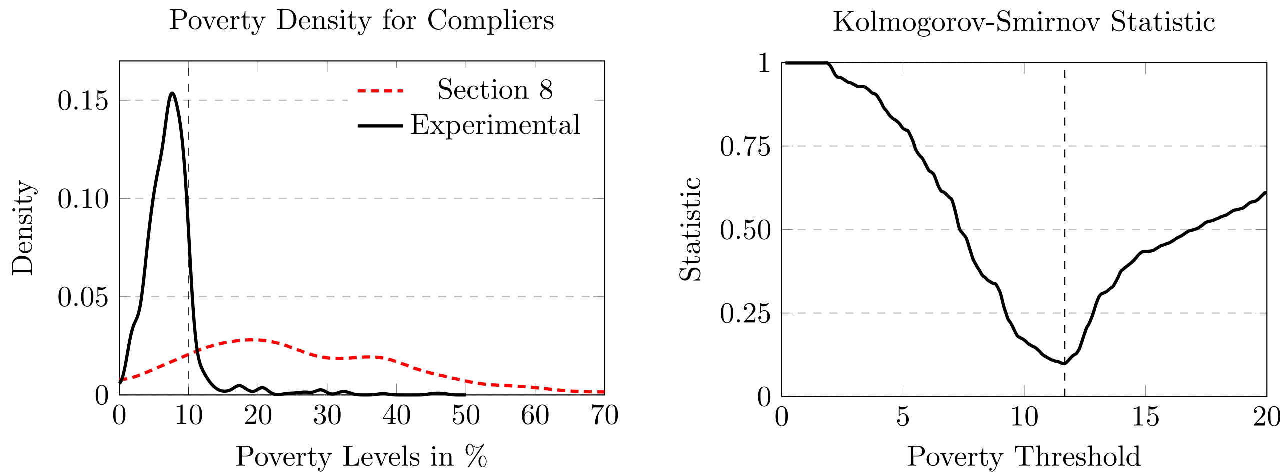
The neighborhood choices of stayers and compliers are easily characterized. The neighborhood choice of families who stay in their original dwellings is t_h. The experimental voucher can only be used to relocate to low poverty neighborhoods. Thus the neighborhood choice of experimental families that use the voucher is t_l. Families that decide to use the Section 8 voucher choose between low (t_l) or medium-poverty (t_m) neighborhoods. This ambiguity is resolved by assessing the poverty levels of the chosen neighborhoods.

The experimental voucher defines low-poverty neighborhoods as those whose poverty level is below a soft target of 10%.[64] In practice, 11% of neighborhoods classified as low-poverty were slightly above the nominal threshold (first graph of Figure A.1). I employ a simple approach to address for this fact. I use the poverty distribution of Section 8 compliers to estimate a threshold that best conforms with the poverty distribution of low-poverty neighborhoods. Specifically, I estimate the threshold that minimizes the Kolmogorov-Smirnov statistic between the poverty distribution of Section 8 compliers and the poverty distribution of experimental compliers. The empirical threshold is 11.67% (second graph of Figure A.1).

It remains to determine the neighborhood choice for the self-movers, which comprise all families that have relocated between surveys. The goal is to identify families who decided to move by the time of the onset of the intervention. To do so, I explore the available information on the time spell from voucher assignment until the first relocation. I account for this fact using the same procedure that yields the poverty threshold. I estimate the threshold the minimizes the difference on the distribution on relocation time between compliers and self-movers. The first graph of Figure A.2 presents the distribution of relocation time for compliers while the second graph presents the Kolmogorov-Smirnov statistics for the difference on relocation time between compliers and self-movers. The corrected thresholds for relocation time are 8.6 months for medium-poverty neighborhoods and 10.6 months for low-poverty neighborhoods. The neighborhood choice of self-movers that relocate before these thresholds is set at either low or medium-poverty neighborhoods.

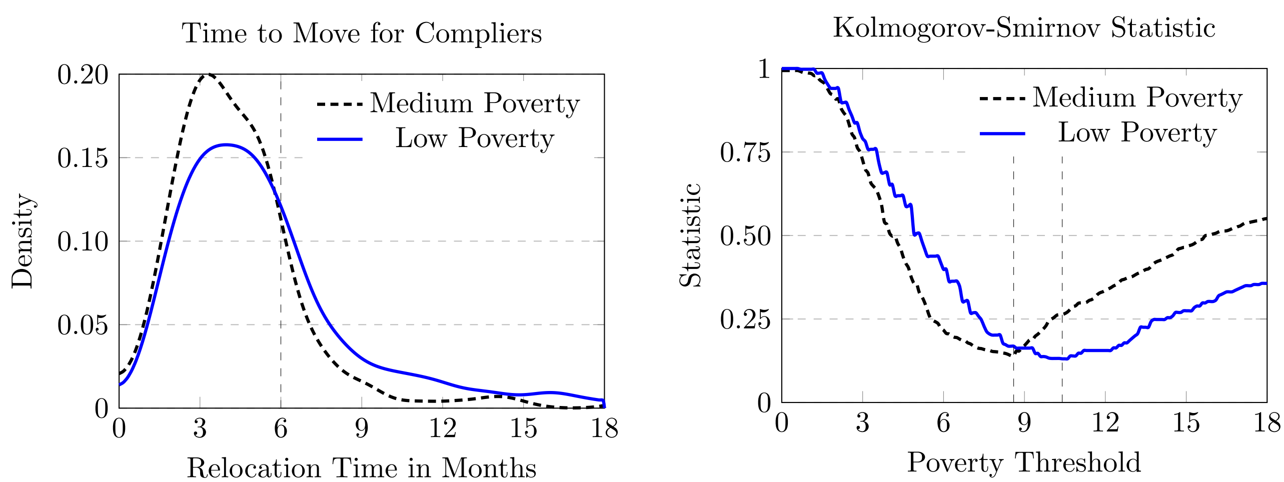
[64] Using the 1990 US Census.

Figure A.1: Poverty Densities and Threshold Investigation



The first graph presents the poverty density of chosen neighborhoods for families who comply with the Experimental and Section 8 vouchers. The second graph presents the Kolmogorov-Smirnov statistics (y-axis) between the poverty distribution of experimental compliers and the poverty distribution of Section 8 compliers that is right-bounded by a threshold (x-axis).

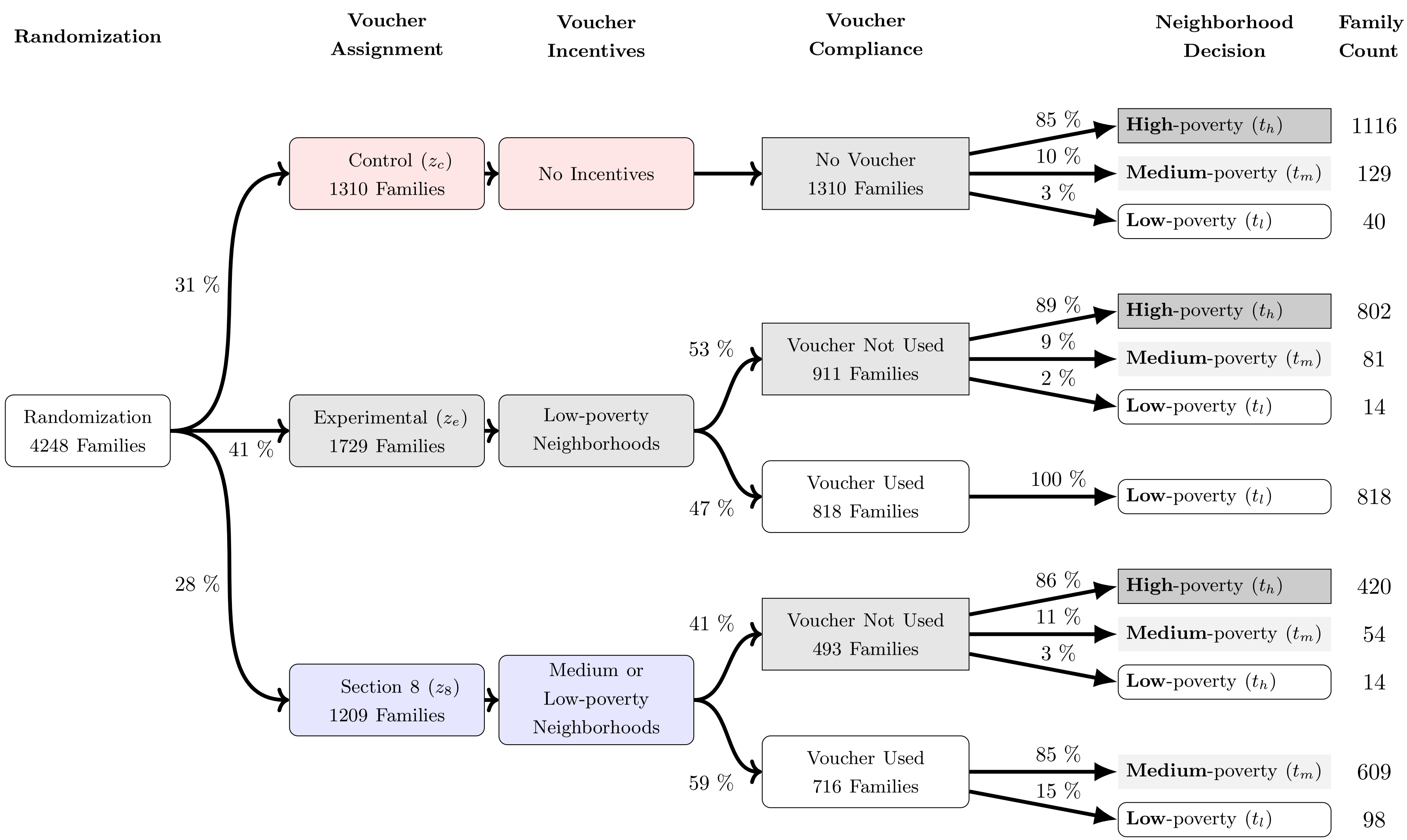
Figure A.2: Time to Relocate Densities and Threshold Investigation



The first graph presents the density of the time to relocate into low and medium poverty neighborhoods since voucher assignment for families who comply with the vouchers. Density estimates use the Gaussian Kernel with optimal bandwidth. The second graph presents the Kolmogorov-Smirnov statistics (y-axis) between the distribution time to relocate of voucher compliers and the distribution of time to relocate for self-movers that are right-bounded by a threshold (x-axis).

Figure A.3 summarizes the neighborhood decision of the MTO families by voucher assignment. Nearly 85% the control families choose high-poverty neighborhoods, 10% choose medium poverty neighborhoods and 3% choose low-poverty neighborhoods. Families that do not use the voucher share a similar composition of neighborhood choices. Around 15% of families that use the Section 8 voucher decide for low-poverty neighborhoods while 85% of Section 8 compliers choose medium-poverty neighborhoods. Neighborhood choices are robust across variations of the assignment procedure. For instance, we can generate alternative values for the neighborhood choices by setting the poverty threshold to its nominal value of 10% and the relocation time to 6 months generates. These values agree with the neighborhood choices described by the procedure above in 97% of the cases.

Figure A.3: **Neighborhood Relocation by Voucher Assignment and Compliance**



This figure describes the possible decision patterns of families in MTO that resulted from voucher assignment and family compliance. Low-poverty (t_l) neighborhoods are defined as those whose share of poor residents is below 10% according to the 1990 sensus (Orr et al., 2003). High poverty (t_h) neighborhoods are the housing projects originally targeted by the intervention. Medium poverty (t_m) neighborhoods are neither the high poverty nor the ones classified as low poverty. Families who stay in their baseline housing live in high poverty neighborhoods (t_h). Families who use the experimental voucher (z_e) relocate to low poverty neighborhoods (t_l). Families who use Section 8 voucher (z_8) can decide between low (t_l) or medium (t_m) poverty neighborhoods. Control families (z_c) and families that do not use the vouchers may choose freely among all three neighborhoods: high (t_h), medium (t_m) or low (t_l).

C TOT parameter of a Simplified Intervention

Consider a simpler intervention than MTO, which randomly assigns families living in high poverty neighborhoods to either a control group z_c or an experimental group z_c. The experimental group receives a voucher that incentivizes families to move to low poverty neighborhoods, while the control group receives no incentives. Families decide between two choices: remain in a high poverty neighborhood t_h or move to a low poverty neighborhood t_l.

Suppose that the investigators can prevent families from moving. Families assigned to the control group remain in the high poverty area. Families assigned to the experimental voucher that decide to not use the voucher do not move either. The only families that relocate are those assigned to the experimental voucher that agree to use the voucher.

This model admits two latent family types. *Compliers* are families who intend to use the experimental voucher in case they are assigned to it. *Non-compliers* are families that do not intend to use the experimental voucher. Notationally, $T_i(z) \in \{t_h, t_l\}$ denotes the potential choice of family i that is assigned to the group $z \in \{z_c, z_e\}$ and $Y_i(t)$ denotes the potential outcome of family i when the neighborhood choice is fixed at $t \in \{t_h, t_l\}$.

The response vector $\boldsymbol{S}_i = [T_i(z_c), T_i(z_e)]'$ lists the potential choices of a family i if it were assigned to the control and experimental groups respectively. If family i is a complier, then $\boldsymbol{S}_i = [t_h, t_l]'$, otherwise family i is a non-complier and its response vector is given by $\boldsymbol{S}_i = [t_h, t_h]'$. For notational simplicity, let $\boldsymbol{s}_n = [t_h, t_h]'$ denote the potential choices for non-compliers and let $\boldsymbol{s}_c = [t_h, t_l]'$ denote the potential choices for compliers.

IV assumptions apply, namely, $Z \perp\!\!\!\perp (T(z), Y(t))$ for all $(z, t) \in \{t_h, t_l\} \times \{z_c, z_e\}$. IV assumptions imply that $Z \perp\!\!\!\perp \boldsymbol{S}$. Otherwise stated, the IV randomization ensures that the share of family types in the experimental and control groups is the same. Consequently, the voucher take-up rate identifies the share of compliers, that is $P(\boldsymbol{S} = \boldsymbol{s}_c)$.

The intention-to-treat (ITT) parameter is given by $ITT = E(Y|Z = z_e) - E(Y|Z = z_c)$. It identifies the causal effect of being offered the experimental voucher. It is useful to express the ITT parameter as the weighted average of the voucher effects for compliers and non-compliers multiplied by their respective shares:

$$ITT = ITT_e(\boldsymbol{s}_n)P(\boldsymbol{S} = \boldsymbol{s}_{nc}) + ITT(\boldsymbol{s}_c)P(\boldsymbol{S} = \boldsymbol{s}_c), \tag{97}$$

$$ITT(\boldsymbol{s}_n) \equiv E(Y|Z = z_e, \boldsymbol{S} = \boldsymbol{s}_n) - E(Y|Z = z_c, \boldsymbol{S} = \boldsymbol{s}_n) = E(Y(t_h) - Y(t_h)|\boldsymbol{S} = \boldsymbol{s}_n) = 0, \tag{98}$$

$$ITT(\boldsymbol{s}_c) \equiv E(Y|Z = z_e, \boldsymbol{S} = \boldsymbol{s}_c) - E(Y|Z = z_c, \boldsymbol{S} = \boldsymbol{s}_c) = E(Y(t_l) - Y(t_h)|\boldsymbol{S} = \boldsymbol{s}_c), \tag{99}$$

where $ITT(\boldsymbol{s}_c), ITT(\boldsymbol{s}_n)$ denotes the the ITT parameter for families of type $\boldsymbol{s}_c$ and $\boldsymbol{s}_n$ respectively: The voucher effect for non-compliers is given by $ITT(\boldsymbol{s}_n)$ in (98), and it is zero as these families do not relocate. The voucher effect for compliers is given by $ITT(\boldsymbol{s}_c)$ in (99). Note that the compliers always choose t_h if assigned to z_c and choose t_l if assigned to z_e. Thus $ITT(\boldsymbol{s}_c)$ gives the causal effect of low versus high poverty neighborhoods on the outcome.

TOT in (6) is the ITT effect divided by the voucher take-up rate. In this setup, the TOT identifies the causal effect of the low versus high poverty neighborhoods for compliers:

$$TOT = \frac{ITT}{P(\boldsymbol{S} = \boldsymbol{s}_c)} \tag{100}$$

$$= \frac{ITT(\boldsymbol{s}_c)P(\boldsymbol{S} = \boldsymbol{s}_c)}{P(\boldsymbol{S} = \boldsymbol{s}_c)} \tag{101}$$

$$= E(Y(t_l) - Y(t_h)|\boldsymbol{S} = \boldsymbol{s}_c), \tag{102}$$

where the first equality is the definition of TOT, the second equality is due to (97)–(98). The third equation is due to (99).

D How IV Controls for Unobserved Characteristics

The main paper uses the language of potential outcomes to examine the identification in MTO. The primary advantage of the potential outcomes framework is its simplicity. The framework does not employ structural equations nor explicitly display unobserved variables. Its simplicity comes at a cost. It harms the interpretation of the causal model that generates the data. In particular, it is difficult to understand that the identification of causal parameters hinges on controlling for family unobserved characteristics. This section describes the causal model of MTO using structural equations. It clarifies the causal concepts underlying causality and the identification of causal parameters. See Heckman and Pinto (2022) for a recent discussion on causality, structural models and the limitations of the potential outcome framework.

The observed variables in MTO are: (1) voucher assignment $Z \in \{z_c, z_8, z_e\}$; (2) neighborhood choice $T \in \{t_h, t_m, t_l\}$; (3) outcome $Y \in \mathbb{R}$; and (4) baseline characteristics $\boldsymbol{X} \in \mathbb{R}^{|\boldsymbol{x}|}$. The MTO model is characterized by the following system of causal relations:

$$\begin{aligned} \text{Choice Equation}: &\quad T = f_T(Z, \boldsymbol{V}, \boldsymbol{X}), && (103)\\ \text{Outcome Equation}: &\quad Y = f_Y(T, \boldsymbol{V}, \boldsymbol{X}, \epsilon), && (104)\\ \text{Conditional Independence}: &\quad Z \perp\!\!\!\perp \boldsymbol{V} | \boldsymbol{X}, && (105)\end{aligned}$$

where $\boldsymbol{V}$ denotes the vector of family unobserved characteristics and ϵ is an unobserved variable satisfying $(Z, T, \boldsymbol{X}, \boldsymbol{V}) \perp\!\!\!\perp \epsilon$.[65] $\boldsymbol{V}$ is a confounding random vector that generates selection bias by causing both choice T and the outcome Y. Baseline variables $\boldsymbol{X}$ are family observed characteristics that cause T and Y. The experiment generates two required properties for Z to be an instrument: (104) implies that Z only affects Y through its impact on T (exclusion restriction); and (105) implies that Z is statistically independent of unobserved characteristics $\boldsymbol{V}$ given baseline variables $\boldsymbol{X}$.

The potential (counterfactual) outcome of family $i \in \mathcal{I}$ placed in neighborhood t is given by $Y_i(t) \equiv f_Y(t, \boldsymbol{V}_i, \boldsymbol{X}_i, \epsilon_i)$. It is the hypothetical outcome that would occur if the neighborhood choice of family i were exogenously set to $t \in \{t_h, t_m, t_l\}$. The potential choice of family i is given by $T_i(z) \equiv f_T(z, \boldsymbol{V}_i, \boldsymbol{X}_i)$. It is the choice that family i would take if it were exogenously assigned to voucher $z \in \{z_c, z_8, z_e\}$. Conditional independence (105) implies the IV exogeneity condition $\big(Y(t), T(z)\big) \perp\!\!\!\perp Z|\boldsymbol{X}$ for $(z,t) \in \{z_c, z_8, z_e\} \times \{t_h, t_m, t_l\}$. Outcome Y and choice T can be written in terms of potential variables as:

$$Y = \sum_{t \in \{t_l, t_m, t_h\}} D_t \cdot Y(t) = Y(T), \quad \text{and} \quad T = \sum_{z \in \{z_c, z_8, z_e\}} D_z \cdot T(z) = T(Z), \tag{106}$$

where $D_t = \mathbf{1}[T = t]; t \in \{t_h, t_m, t_l\}$ indicates neighborhood choices, $D_z = \mathbf{1}[Z = z]; z \in \{z_c, z_8, z_e\}$ indicates voucher assignment and $\mathbf{1}[A]$ is the indicator function that takes value 1 if event A is true and zero otherwise.

[65] Measurement error and misspecification are possible sources of the unobserved error term ϵ.

The causal effect of living in a low versus high-poverty neighborhood for family i is defined as $Y_i(t_l) - Y_i(t_h)$. It is the difference in the potential outcome of family i if it were to reside in each of these two neighborhood types. If responses are heterogeneous, this individual effect is not identified since we only observe the potential outcome corresponding to the neighborhood chosen by the family. A mean neighborhood treatment effect is the expectation of individual effects, such as $Y_i(t_l) - Y_i(t_h)$, for subsets of families $i \in \mathcal{I}$. To gain intuition, it is useful to write the observed outcome of families $i \in \mathcal{I}$ that choose t_l or t_h as:

$$Y_i = \beta_0 + \beta_i D_{t_l,i} + \epsilon_i, \tag{107}$$

where $\beta_i = Y_i(t_l) - Y_i(t_h),\ \beta_0 = E(Y(t_h)),\ \epsilon_i = Y_i(t_h) - E(Y(t_h))$, and $D_{t,i} \equiv \mathbf{1}[T_i = t]$ is the choice indicator for a family i such that $T_i \in \{t_l, t_h\}$.

Equation (107) is a random coefficient model where β_i varies across $i \in \mathcal{I}$.[66] If $Y(t)$ and T were statistically independent, $Y(t) \perp\!\!\!\perp T$, then we could evaluate the average neighborhood effect $E(Y(t_h) - Y(t_h))$ by least squares taking mean differences. Selection bias induces a correlation between $Y(t)$ and T via $\boldsymbol{V}$. As a consequence, the regressor $D_{t_l,i}$ in (107) correlates with both the error term $\epsilon_i = Y_i(t_h) - E(Y(t_h))$ and the random coefficient $\beta_i = Y_i(t_l) - Y_i(t_h)$. Without further assumptions, neither least squares nor two-stage least squares identifies $E(Y(t_h) - Y(t_h))$.[67]

A popular identification strategy invokes a *matching* condition which assumes that T and $Y(t)$ are independent conditioned on $\boldsymbol{X}$, $Y(t) \perp\!\!\!\perp T|\boldsymbol{X}$. This assumption enables the analyst to identify counterfactual outcomes by controlling for $\boldsymbol{X} : E(Y|T = t, \boldsymbol{X}) = E(Y(t)|T = t, \boldsymbol{X}) = E(Y(t)|\boldsymbol{X})$, where the first equality is due to (106) and the second is due to $Y(t) \perp\!\!\!\perp T|\boldsymbol{X}$. The average neighborhood effect across all families in $i \in \mathcal{I}$ is obtained by integrating out $\boldsymbol{X}$:

$$E(Y(t_l) - Y(t_h)) = \int \left(E(Y|T = t_l, \boldsymbol{X} = \boldsymbol{x}) - E(Y|T = t_h, \boldsymbol{X} = \boldsymbol{x})\right) dF_{\boldsymbol{X}}(\boldsymbol{x}), \tag{108}$$

where $F_{\boldsymbol{X}}(\cdot)$ is the cumulative distribution function (CDF) of $\boldsymbol{X}$.

A matching assumption is not valid if there is selection bias on unobservables that are not in $\boldsymbol{X}$. However, it is always true that $Y(t) \perp\!\!\!\perp T|(\boldsymbol{X}, \boldsymbol{V})$ holds. The identification of causal effects hinges on controlling for $\boldsymbol{X}$ as well as for the unobservables $\boldsymbol{V}$. This paper presents a nonparametric method to control for $\boldsymbol{V}$. I suppress $\boldsymbol{X}$ henceforward to simplify notation. The analysis should be understood as conditioned on $\boldsymbol{X}$.

One identification strategy invokes a parametric model that uses Z to control for $\boldsymbol{V}$. Examples of such approach in the MTO literature are Aliprantis and Richter (2020); Chesher et al. (2020); Galiani et al. (2015). This paper takes a different approach. I exploit the instrument Z and the incentives in MTO to *nonparametrically* control for $\boldsymbol{V}$. The approach does not rely on any functional form assumptions, nor does it require intensive computational effort.

It is possible to control for $\boldsymbol{V}$ by partitioning families based on choice behavior described by *response types* or *principal strata*, namely, the counterfactual choices that the family would take across the instrumental values.[68]

Let the *Response vector* $\boldsymbol{S}_i = [T_i(z_c), T_i(z_8), T_i(z_e)]'$ be the neighborhood choices made by family

[66] Also called Switching Regression Model (Quandt, 1958, 1972).

[67] Later in this paper I present assumptions that enable the analyst to use a modified version of the two-stage least square regression to identify counterfactual outcomes.

[68] The use of response types dates back to Balke and Pearl (1994) and Frangakis and Rubin (2002). See Pinto (2016) or Heckman and Pinto (2018) for a discussion.

i when assigned to each of the instrumental values z_c, z_8, z_e. A response type consists of a vector of choice values that $\boldsymbol{S}$ may take. For instance, family i that has *response type* $\boldsymbol{S}_i = [t_h, t_m, t_l]'$ chooses a high-poverty neighborhood when offered z_c ($T_i(z_c) = t_h$), a medium-poverty neighborhood when offered z_8 ($T_i(z_8) = t_m$), and a low-poverty neighborhood when offered z_e ($T_i(z_e) = t_l$).

Choice T is determined by Z and $\boldsymbol{S}$. Given a response type, choice T depends only on assignment Z, which is independent of its potential outcome $Y(t)$. Therefore $Y(t) \perp\!\!\!\perp T|\boldsymbol{S}$ holds. Intuitively, the neighborhood choice within a group of families that share the same response type can be understood as if it were generated by randomized controlled trial RCT where Z determines the neighborhood assignment. If we knew all the families $i \in \mathcal{I}$ that have type $\boldsymbol{S}_i = [t_h, t_m, t_l]$, we would be able to identify the causal effect of low t_l versus high t_h from:

$$E(Y|Z = z_e, \boldsymbol{S} = [t_h, t_m, t_l]') - E(Y|Z = z_c, \boldsymbol{S} = [t_h, t_m, t_l]') \tag{109}$$

$$= E(Y|T = t_l, \boldsymbol{S} = [t_h, t_m, t_l]') - E(Y|T = t_h, \boldsymbol{S} = [t_h, t_m, t_l]'), \text{ due to response type} \tag{110}$$

$$= E(Y(t_l)|T = t_l, \boldsymbol{S} = [t_h, t_m, t_l]') - E(Y(t_h)|T = t_h, \boldsymbol{S} = [t_h, t_m, t_l]'), \quad \text{due to (106)} \tag{111}$$

$$= E(Y(t_l) - Y(t_h)|\boldsymbol{S} = [t_h, t_m, t_l]), \qquad \text{due to } Y(t) \perp\!\!\!\perp T|\boldsymbol{S} \tag{112}$$

Response types control for unobserved characteristics $\boldsymbol{V}$ by generating a useful partition of its support. Holding $\boldsymbol{X}$ fixed, the potential choice $\boldsymbol{T}(z) = f_T(z, \boldsymbol{V})$ depends only on $\boldsymbol{V}$. The set of unobserved characteristics corresponding to response type $\boldsymbol{s} = [t_h, t_m, t_l]$ is given by:

$$\mathcal{V}_{\boldsymbol{s}} = \{\boldsymbol{v} \in \mathit{supp}(\boldsymbol{V}) \text{ such that } f_T(z_c, \boldsymbol{v}) = t_h, f_T(z_8, \boldsymbol{v}) = t_m, f_T(z_e, \boldsymbol{v}) = t_l\}.$$

Events $\boldsymbol{S} = \boldsymbol{s}$ and $\boldsymbol{V} \in \mathcal{V}_{\boldsymbol{s}}$ are equivalent. $Y(t) \perp\!\!\!\perp T|(\boldsymbol{S} = \boldsymbol{s})$ implies that $Y(t) \perp\!\!\!\perp T|(\boldsymbol{V} \in \mathcal{V}_{\boldsymbol{s}})$ holds. Conditioning on $\boldsymbol{S} = \boldsymbol{s}$ is equivalent to conditioning on the set of unobserved variables $\boldsymbol{V} \in \mathcal{V}_{\boldsymbol{s}}$ that renders the choice T statistically independent of the counterfactual outcomes $Y(t)$.[69] As $\boldsymbol{s}$ ranges in $\mathit{supp}(\boldsymbol{S})$, it spans the support of $\boldsymbol{V}$ as $\mathit{supp}(\boldsymbol{V}) = \bigcup_{\boldsymbol{s} \in \mathit{supp}(S)} \mathcal{V}_{\boldsymbol{s}}$.

Response types are not observed, but we can express observed outcomes as a mixture of potential outcomes conditioned on response types as written in equation (8) of the main paper.

E Comparing MTO with Other Choice Models

This section investigate choice models that differ from the choice Model of MTO.

The monotonicity condition of Angrist and Imbens (1995) is widely known among empirical economists. (Vytlacil, 2006) demonstrates that assuming the monotonicity condition of Angrist and Imbens (1995) is equivalent to assume an ordered choice model with random thresholds. Thus, I refer to the monotonicity condition of Angrist and Imbens (1995) as "ordered monotonicity" for sake of clarity of exposition.

Section E.1 shows that the response matrix of MTO does not comply with the ordered monotonicity condition. Section E.2 presents examples of incentives that justify assuming that the ordered monotonicity condition holds. Section E.3 presents the identification analysis for a three-choice model with a three-valued IV where the ordered monotonicity condition holds. Section E.4 examines the MTO model when the Normal Choice Assumption in Section (4.1) is violated. Section E.5 investigates the case of misrepresented choice incentives. Section E.6 investigates the three-choice model with a parallel design described in Kirkeboen, Leuven, and Mogstad (2016).

[69] $\boldsymbol{S}$ plays the role of a control function in Heckman and Robb (1985) and Powell (1994) as well as an unobserved balancing score in Rosenbaum and Rubin (1983).

E.1 Do MTO Incentives Justify an Ordered Choice Model?

A natural inquiry is whether it is possible to model neighborhood choices in MTO as an ordered choice model. Viewing the treatment as ordered is appealing because it relates to the well-known monotonicity condition of Angrist and Imbens (1995) described below:

$$\text{For any } z, z' \ , T_i(z) \leq T_i(z') \forall\, i \text{ or } T_i(z) \geq T_i(z') \forall\, i. \tag{113}$$

in (113), Vytlacil (2006) demonstrates that the monotonicity condition (113) is equivalent to assuming an ordered choice model. For sake of clarity, condition (113) is termed ordered monotonicity henceforward.

Ordered monotonicity (113) benefits from well-establish literature in policy evaluation. In particular, Angrist and Imbens (1995) has shown that the standard 2SLS regression evaluates an interpretable causal parameter under (113). Unfortunately, ordered monotonicity (113) is not compatible with MTO incentives.

Ordered monotonicity (113) is equivalent to state that there exist a sequence of instrumental variables $z_1, ..., z_J$ and such that $T_i(z_1) < ... < T_i(z_J)$ for all agents $i \in \mathcal{I}$. If (113) were true for MTO, we would be able to relabel the instrumental values and neighborhood choices of MTO, say $supp(Z) = \{z_1, z_2, z_3\}$, and $T \in \{1, 2, 3\}$, such that

$$T_i(z_1) \leq T_i(z_2) \leq T_i(z_3) \text{ holds for all } i \in \mathcal{I}. \tag{114}$$

Unfortunately, condition (114) does not hold regardless of how we label instrumental values and neighborhood choices. To see this, let the instrumental values z_c, z_8, z_e be relabeled as z_1, z_2, z_3 and the neighborhood choices z_h, z_m, z_l as $1, 2, 3$. In this notation, the MTO response matrix is given by:

$$\text{Relabeled MTO Response Matrix}: \quad \boldsymbol{R} = \begin{array}{c} \begin{matrix} \boldsymbol{s}_1 & \boldsymbol{s}_2 & \boldsymbol{s}_3 & \boldsymbol{s}_4 & \boldsymbol{s}_5 & \boldsymbol{s}_6 & \boldsymbol{s}_7 \end{matrix} \\ \begin{bmatrix} 1 & 2 & 3 & 1 & 1 & 2 & 1 \\ 1 & 2 & 3 & 2 & 3 & 2 & 2 \\ 1 & 2 & 3 & 3 & 3 & 3 & 1 \end{bmatrix} \end{array} \begin{matrix} T_i(z_1) \\ T_i(z_2) \\ T_i(z_3) \end{matrix} \tag{115}$$

The response types $\boldsymbol{s}_1$ until $\boldsymbol{s}_6$ are weakly increasing, which comply with the monotonicity condition $T_i(z_1) \leq T_i(z_2) \leq T_i(z_3)$. Response type $\boldsymbol{s}_7$ however violates this condition as $T_i(z_2) > T_i(z_3)$. Switching the second and third rows of (115) would make $\boldsymbol{s}_7$ comply with the monotonicity criteria (114), but $\boldsymbol{s}_4$ would violate it. It is easy to see that the monotonicity condition would not be satisfied by relabeling the neighborhood choices either.

E.2 Which Incentives Justify Ordered Monotonicity?

The are several incentive schemes that justify the ordered monotonicity (114). Let the incentive matrix $\boldsymbol{L}$ be the $J \times K$ matrix that characterises the incentives induced by instrumental values in $Z \in \{z_1, ..., z_J\}$, toward choices in $T \in \{1, ..., K\}$. The matrix input $\boldsymbol{L}[z_j, k]$ denotes the incentive for choosing choice $k \in \{1, ..., K\}$ when assigned to instrumental variable $z_j\{z_1, ..., z_J\}$. One incentive scheme that generates the ordered choice models is the presence of increasing incentive increments, that is:

$$\boldsymbol{L}[z_{j+1}, t_k] - \boldsymbol{L}[z_j, t_k] < \boldsymbol{L}[z_{j+1}, t_{k+1}] - \boldsymbol{L}[z_j, t_{k+1}] \text{ for } j \in \{1, ..., J-1\} \text{ and } k \in \{1, ..., K-1\}. \tag{116}$$

Examples of such incentives for three-choice model and a three-valued instrument of MTO are:

$$\boldsymbol{L}_1 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 2 \\ 0 & 2 & 4 \end{bmatrix}_{z_1, z_2, z_3}^{1\ 2\ 3}, \quad \boldsymbol{L}_2 = \begin{bmatrix} 1 & 2 & 3 \\ 1 & 4 & 9 \\ 1 & 8 & 27 \end{bmatrix}_{z_1, z_2, z_3}^{1\ 2\ 3}, or \quad \boldsymbol{L}_3 = \begin{bmatrix} 1 & 1 & 1 \\ 2 & 4 & 8 \\ 4 & 16 & 64 \end{bmatrix}_{z_1, z_2, z_3}^{1\ 2\ 3} \tag{117}$$

The combination of WARP and Normal Choice criteria of Section 4 generates the following choice restriction:

$$\text{If } T_i(z) = t \text{ and } \boldsymbol{L}[z', t'] - \boldsymbol{L}[z, t'] \leq \boldsymbol{L}[z', t] - \boldsymbol{L}[z, t] \text{ then } T_i(z') \neq t'. \tag{118}$$

Choice Rule (118) is intuitive. It states that if an agent i chooses t instead of t' under z, and z' offers greater incentives towards t than t', then agent i will not choose t' under z'. Applying Choice Rule (118) to any of the incentive matrices $\boldsymbol{L}_1$, $\boldsymbol{L}_2$ or $\boldsymbol{L}_3$ in (117) generates the following response matrix:

$$\text{MTO Response Matrix:} \quad \boldsymbol{R} = \begin{array}{c} \begin{matrix} \boldsymbol{s}_1 & \boldsymbol{s}_2 & \boldsymbol{s}_3 & \boldsymbol{s}_4 & \boldsymbol{s}_5 & \boldsymbol{s}_6 & \boldsymbol{s}_7 & \boldsymbol{s}_8 & \boldsymbol{s}_9 & \boldsymbol{s}_{10} \end{matrix} \\ \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 2 & 2 & 2 & 3 \\ 1 & 1 & 1 & 2 & 2 & 3 & 2 & 2 & 3 & 3 \\ 1 & 2 & 3 & 2 & 3 & 3 & 2 & 3 & 3 & 3 \end{bmatrix} \end{array} \begin{matrix} T_i(z_1) \\ T_i(z_2) \\ T_i(z_3) \end{matrix} \tag{119}$$

Response matrix (119) contains all the admissible response types that satisfy the monotonicity condition $T_i(z_1) \leq T_i(z_2) \leq T_i(z_3)$. Indeed, the choices in each of the response types of (119) are weakly increasing. Moreover, there is no response type other than those in (119) that satisfy the ordered monotonicity (114).

E.3 What can Identified by Assuming Ordered Monotonicity?

As mentioned, Response matrix (119) is obtained by assuming the monotonicity assumption (114) of Angrist and Imbens (1995) in the case of a three choice model with a three-valued instrument. The response matrix has 10 response types comprising of 18 counterfactual outcomes condition on response types. For instance, there are six response types that take treatment value the treatment $T = 1$, namely, $\boldsymbol{s}_1, \boldsymbol{s}_2, \boldsymbol{s}_3, \boldsymbol{s}_4, \boldsymbol{s}_5$ and $\boldsymbol{s}_6$. There are also six response types that take treatment value the treatment $T = 2 : \boldsymbol{s}_2, \boldsymbol{s}_4, \boldsymbol{s}_5, \boldsymbol{s}_7, \boldsymbol{s}_8$ and $\boldsymbol{s}_9$. Finally, there are six response types that take treatment value the treatment $T = 3 : \boldsymbol{s}_3, \boldsymbol{s}_5, \boldsymbol{s}_6, \boldsymbol{s}_8, \boldsymbol{s}_9$ and $\boldsymbol{s}_{10}$.

Equation (34) provides the necessary and sufficient criteria to examine the identification of counterfactual outcomes. The criteria can also be use to examine the identification of response type probabilities (by setting the outcome to one). The application of the criteria to the response matrix (119) enables the identification of the following causal parameters:

1. The following response type probabilities are identified:

$$P(\boldsymbol{S}=\boldsymbol{s}_1), P(\boldsymbol{S}=\boldsymbol{s}_{10}), P(\boldsymbol{S}\in\{\boldsymbol{s}_2,\boldsymbol{s}_3\}), P(\boldsymbol{S}\in\{\boldsymbol{s}_6,\boldsymbol{s}_9\}),$$
$$P(\boldsymbol{S}\in\{\boldsymbol{s}_2,\boldsymbol{s}_4,\boldsymbol{s}_7\}), P(\boldsymbol{S}\in\{\boldsymbol{s}_3,\boldsymbol{s}_5,\boldsymbol{s}_8\}), \text{ and } P(\boldsymbol{S}\in\{\boldsymbol{s}_4,\boldsymbol{s}_5,\boldsymbol{s}_6\}).$$

   Note that the probabilities $P(\boldsymbol{S}\in\{\boldsymbol{s}_7,\boldsymbol{s}_8,\boldsymbol{s}_9\})$ and $P(\boldsymbol{S}\in\{\boldsymbol{s}_4,s_5,\boldsymbol{s}_7,\boldsymbol{s}_8\})$ can be obtained as a liner combination of the identified probabilities above.

2. The following counterfactual outcome expectations are identified:

$$E(Y(1)|\boldsymbol{S}=\boldsymbol{s}_1), E(Y(1)|\boldsymbol{S}\in\{\boldsymbol{s}_2,\boldsymbol{s}_3\}), \text{ and } E(Y(1)|\boldsymbol{S}\in\{\boldsymbol{s}_4,\boldsymbol{s}_5,\boldsymbol{s}_6\})$$
$$E(Y(2)|\boldsymbol{S}\in\{\boldsymbol{s}_2,\boldsymbol{s}_4,\boldsymbol{s}_7\}), \; E(Y(2)|\boldsymbol{S}\in\{\boldsymbol{s}_7,\boldsymbol{s}_8,\boldsymbol{s}_9\}), \text{ and } E(Y(2)|\boldsymbol{S}\in\{\boldsymbol{s}_4,\boldsymbol{s}_5,\boldsymbol{s}_7,\boldsymbol{s}_8\})$$
$$E(Y(3)|\boldsymbol{S}=\boldsymbol{s}_{10}), E(Y(3)|\boldsymbol{S}\in\{\boldsymbol{s}_6,\boldsymbol{s}_9\}), \text{ and } E(Y(3)|\boldsymbol{S}\in\{\boldsymbol{s}_3,\boldsymbol{s}_5,\boldsymbol{s}_8\}).$$

Note that there are only two response type probabilities that are point-identified: $P(S=s_1)$, and $P(S=s_{10})$. Moreover, only two out of the 18 counterfactual outcomes listed above are point-identified: $E(Y(1)|\boldsymbol{S}=\boldsymbol{s}_1)$, and $E(Y(3)|\boldsymbol{S}=\boldsymbol{s}_{10})$. The number of identified parameters for MTO are sharply different. The MTO response matrix in (115) has seven response types, all response type probabilities are point-identified, and six out of 12 counterfactual outcomes are point-identified.

E.4 Violating the Normal Choice Assumption

Consider a family i chooses low-poverty neighborhood t_l under control, that is $T_i(z_c)=t_l$. According to WARP, family i will also choose low-poverty neighborhood t_l under the experimental voucher, namely, $T_i(z_e)=t_l$. Suppose this family chooses medium-poverty neighborhood t_m under z_8. This behavior is denoted by the following response type, $\boldsymbol{s}^*=[t_l,t_m,t_l]'$, which violates the normal choice assumption.

The behavior is unlikely because if the family chooses low=poverty neighborhood with no incentives, then their neighborhood choice is likely to remain the same when incentives to choose low-poverty neighborhood are offered. Nevertheless, it is informative to investigate the identification of causal parameters when including the response type $\boldsymbol{s}^*$ to the response matrix of MTO. The response matrix that includes the response type $\boldsymbol{s}^*$ is displayed below:

$$\boldsymbol{R} = \begin{array}{c} \begin{matrix} \boldsymbol{s}_{ah} & \boldsymbol{s}_{am} & \boldsymbol{s}_{al} & \boldsymbol{s}_{fc} & \boldsymbol{s}_{pl} & \boldsymbol{s}_{pm} & \boldsymbol{s}_{ph} & \boldsymbol{s}^* \end{matrix} \\ \begin{bmatrix} t_h & t_m & t_l & t_h & t_h & t_m & t_h & t_l \\ t_h & t_m & t_l & t_m & t_l & t_m & t_m & t_m \\ t_h & t_m & t_l & t_l & t_l & t_l & t_h & t_l \end{bmatrix} \end{array} \begin{matrix} T_i(z_c) \\ T_i(z_8) \\ T_i(z_e) \end{matrix} \tag{120}$$

Equation (34) provides the necessary and sufficient criteria to examine the identification of counterfactual outcomes. As mentioned, the criteria can also be use to examine the identification of response type probabilities by setting the outcome to one. The application of the criteria to the response matrix (120) enables to point-identify the following response type probabilities:

$$P(\boldsymbol{S}=\boldsymbol{s}_{ah}), P(\boldsymbol{S}=\boldsymbol{s}_{am}), P(\boldsymbol{S}=\boldsymbol{s}_{pm}), \text{ and } P(\boldsymbol{S}=\boldsymbol{s}_{ph})$$

The following response type probabilities are partially identified:

$$P(\boldsymbol{S}\in\{\boldsymbol{s}_{al},\boldsymbol{s}_{pl}\}), P(\boldsymbol{S}\in\{\boldsymbol{s}_{al},\boldsymbol{s}^*\}), P(\boldsymbol{S}\in\{\boldsymbol{s}_{fc},\boldsymbol{s}_{pl}\}), \text{ and } P(\boldsymbol{S}\in\{\boldsymbol{s}_{fc},\boldsymbol{s}^*\}).$$

The response type probability $P(\boldsymbol{S} \in \{\boldsymbol{s}_{al}, \boldsymbol{s}^*\})$ is identified by: $P(\boldsymbol{S} \in \{\boldsymbol{s}_{al}, \boldsymbol{s}^*\}) = P(\boldsymbol{S} = \boldsymbol{s}_{al}) + P(\boldsymbol{S} = \boldsymbol{s}^*) = P(T = t_l | Z = z_c)$. The propensity score $P(T = t_l | Z = z_c)$ which is about 3%. Thus the families corresponding to low-poverty always-takes $\boldsymbol{s}_{al}$ and the response type $\boldsymbol{s}^*$ combined account for only 3% of the sample.

It is expected that the share of $\boldsymbol{s}_{al}$-type families to be bigger than the share of $\boldsymbol{s}^*$-type families. Response types $\boldsymbol{s}_{al}$ and $\boldsymbol{s}^*$ share some similarities with the response types $\boldsymbol{s}_{ah} = [t_h, t_h, t_h]'$ and $\boldsymbol{s}_{ph} = [t_h, t_m, t_h]'$. Note that $\boldsymbol{s}_{ph}$ deviates from the high-poverty always-takers by switching the neighborhood choice under z_8 from t_h to t_m. This deviation is economically justifiable since z_8 does not incentivizes t_h, but does for t_m. Figure 4 shows that the share of families corresponding to $\boldsymbol{s}_{ph}$ is less than a third of the share corresponding to $\boldsymbol{s}_{ah}$.

Note that the response type $\boldsymbol{s}^*$ is a deviation from an always-takers in high-poverty by switching the neighborhood choice under z_8 from t_l to t_m. In this case, there are less economic incentives to do so since z_8 incenivizes t_l. A parallel argument suggests that the response type $\boldsymbol{s}^*$ should be less than a third of the share of $\boldsymbol{s}_{ah}$. In this case, the share of $\boldsymbol{s}^*$ is likely to be less 0.75% of the data and is not of primary concern in the empirical evaluation. On the other hand, the counterfactuals for the full-compliers are partially identified. Addressing the problem of partially identification is primary for the evaluation of the intervention. Appendix G.1 provides bounds for the partially identified counterfactuals while Section 5.3 provides a solution to disentangle the partially identified counterfactuals.

E.5 Misrepresented Choice Incentives

Some families assigned to the experimental group receive counseling that motivated them to move away from high-poverty neighborhoods, regardless of whether the new neighborhood was located in a low- or medium-poverty region. This means that the experimental voucher provides two types of incentives. The primary incentive is the rent-subsidy towards low-poverty neighborhoods. However, for same families, the voucher also provides some weak incentives towards medium-poverty neighborhoods. This feature is modeled by the following incentive matrix:

$$\textbf{New Incentive Matrix} \quad \boldsymbol{L} = \begin{array}{c} \begin{matrix} t_h & t_m & t_l \end{matrix} \\ \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0.5 & 1 \end{bmatrix} \end{array} \begin{matrix} z_c \\ z_8 \\ z_e \end{matrix} \tag{121}$$

The matrix states that the experimental provides full incentives towards t_l and partial incentives towards t_m. The values of 0.5 is not essential since the matrix is ordinal. Any value in $(0, 1)$ represents equivalent incentives and generate the same response matrix. Applying the revealed preference analysis of Section 4 to the incentive matrix (121) generates the following response matrix:

$$\boldsymbol{R} = \begin{array}{c} \begin{matrix} \boldsymbol{s}_{ah} & \boldsymbol{s}_{am} & \boldsymbol{s}_{al} & \boldsymbol{s}_{fc} & \boldsymbol{s}_{pl} & \boldsymbol{s}_{pm} & \boldsymbol{s}_{ph} & \boldsymbol{s}^* \end{matrix} \\ \begin{bmatrix} t_h & t_m & t_l & t_h & t_h & t_m & t_h & t_h \\ t_h & t_m & t_l & t_m & t_l & t_m & t_m & t_m \\ t_h & t_m & t_l & t_l & t_l & t_l & t_h & t_m \end{bmatrix} \end{array} \begin{matrix} T_i(z_c) \\ T_i(z_8) \\ T_i(z_e) \end{matrix} \tag{122}$$

The response matrix above has eight response types. The first seven response types are identical to those of the response matrix of MTO in the main paper. The additional response type is

$\boldsymbol{s}^* = [t_h, t_m, t_m]'$.

As mentioned, equation (34) provides the necessary and sufficient criteria to examine the identification of counterfactual outcomes. The criteria can also be use to examine the identification of response type probabilities by setting the outcome to one. The application of the criteria to the response matrix (120) enables to point-identify the following response type probabilities:

$$P(\boldsymbol{S} = \boldsymbol{s}_{ah}), P(\boldsymbol{S} = \boldsymbol{s}_{al}), P(\boldsymbol{S} = \boldsymbol{s}_{pl}), \text{ and } P(\boldsymbol{S} = \boldsymbol{s}_{ph}).$$

The identification of these probabilities are the same as the equations that identify the response type probabilities in the response matrix with seven response types. Thus, these estimates remain the same regardless if the true response matrix contains or do not contain $\boldsymbol{s}^*\}$.

The following response type probabilities are partially identified:

$$P(\boldsymbol{S} \in \{\boldsymbol{s}_{am}, \boldsymbol{s}_{pm}\}), P(\boldsymbol{S} \in \{\boldsymbol{s}_{am}, \boldsymbol{s}^*\}), P(\boldsymbol{S} \in \{\boldsymbol{s}_{fc}, \boldsymbol{s}_{pm}\}), \text{ and } P(\boldsymbol{S} \in \{\boldsymbol{s}_{fc}, \boldsymbol{s}^*\}).$$

The response type probability $P(\boldsymbol{S} \in \{\boldsymbol{s}_{am}, \boldsymbol{s}^*\})$ is identified by: $P(\boldsymbol{S} \in \{\boldsymbol{s}_{am}, \boldsymbol{s}^*\}) = P(\boldsymbol{S} = \boldsymbol{s}_{am}) + P(\boldsymbol{S} = \boldsymbol{s}^*) = P(T = t_m | Z = z_e)$. The estimate of the propensity score $P(T = t_m | Z = z_e)$ is about 4.5%. Thus the response matrix (122) is correct, the families corresponding to medium-poverty always-takes $\boldsymbol{s}_{am}$ and the response type $\boldsymbol{s}^*$ combined account for only 4.5% of the sample.

The response type probability for the low-poverty always takers is 3%, and since medium-poverty neighborhoods are more diverse and distributed over a larger area than the low-poverty ones, it is safe to assume that the response type probability for the medium-poverty always takers $\boldsymbol{s}_{am}$ is greater than the probability for $\boldsymbol{s}_{ah}$. Therefore, it is safe to assume that the upper bound of the probability of the response type $\boldsymbol{s}^*$ is about 1.5% of the total sample. This analysis considers that all the families assigned to the experimental voucher got life-counseling training sections which encouraged them to leave high-poverty neighborhoods. However, only less than a third of the experimental voucher recipients had the advantage of this training (Feins et al., 1997). This implies that a likely value for the upper bound for the probability of the response type $\boldsymbol{s}^* = [t_h, t_m, t_m]'$ is approximately 0.5% of the total sample.

E.6 The Three-choice Model with a Parallel Design

This section considers a three-valued treatment, $T \in \{t_0, t_1, t_2\}$, and a three-valued instrument $Z \in \{z_0, z_1, z_2\}$ where z_1 incentivizes t_1, z_2 incentivizes t_2, and z_0 is the baseline IV-value that offers no incentives. The model is consistent with a three-arm randomized trial design in which z_1, z_2 correspond to the intended treatment t_1, t_2 and z_0 stands for the control group. This is experiment is an example of the so called parallel design in the literature of randomized controlled trials and has been examined by Kirkeboen, Leuven, and Mogstad (2016). Choice incentives are characterized by the following incentive matrix:

$$\boldsymbol{L} = \begin{array}{c} \\ \\ \\ \end{array}\begin{matrix} t_0 & t_1 & t_2 \\ \end{matrix} \quad \boldsymbol{L} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{matrix} z_0 \\ z_1 \\ z_2 \end{matrix} \tag{123}$$

The incentive matrix (123) differs from the MTO incentive matrix (13) as z_1 incentivizes a single choice t_1, while the respective IV value in MTO, z_8, incentivizes two choices: t_m and t_l. MTO incentives justify tree monotonicity conditions (10)–(12), the incentive matrix (123) justifies only two:

$$\mathbf{1}[T_i(z_0) = t_1] \leq \mathbf{1}[T_i(z_1) = t_1] \tag{124}$$
$$\mathbf{1}[T_i(z_0) = t_2] \leq \mathbf{1}[T_i(z_2) = t_2]. \tag{125}$$

Monotonicity condition (124) states that a change in the instrument from z_0 to z_1 induces agents to shift their choice towards t_1 while (125) states that a change from z_0 to z_2 induces agents towards t_2. Panel B of Table (A.5) shows that monotonicity conditions (124)–(125) eliminate 12 out of the 27 possible response types.

The remaining 15 response types are displayed in response matrix below:

$$\boldsymbol{R} = \begin{array}{c} \begin{matrix} \boldsymbol{s}_1 & \boldsymbol{s}_2 & \boldsymbol{s}_3 & \boldsymbol{s}_4 & \boldsymbol{s}_5 & \boldsymbol{s}_6 & \boldsymbol{s}_7 & \boldsymbol{s}_8 & \boldsymbol{s}_9 & \boldsymbol{s}_{10} & \boldsymbol{s}_{11} & \boldsymbol{s}_{12} & \boldsymbol{s}_{13} & \boldsymbol{s}_{14} & \boldsymbol{s}_{15} \end{matrix} \\ \begin{bmatrix} t_0 & t_0 & t_0 & t_0 & t_0 & t_0 & t_0 & t_0 & t_0 & t_2 & t_1 & t_1 & t_2 & t_2 & t_2 \\ t_0 & t_0 & t_0 & t_1 & t_1 & t_1 & t_2 & t_2 & t_2 & t_2 & t_1 & t_1 & t_0 & t_1 & t_2 \\ t_0 & t_1 & t_2 & t_0 & t_1 & t_2 & t_0 & t_1 & t_2 & t_0 & t_1 & t_2 & t_2 & t_2 & t_2 \end{bmatrix} \end{array} \begin{matrix} T_i(z_0) \\ T_i(z_1) \\ T_i(z_2) \end{matrix} \tag{126}$$

The elimination of monotonicity conditions (124)–(125) are not sufficient to point-identify a single counterfactual outcomes or a single response type probabilities. Equation (34) provides the necessary and sufficient criteria to examine the identification of counterfactual outcomes. The application of the criteria to the response matrix above enables to identify the following counterfactual outcomes:

$E(Y(t_1)|\boldsymbol{S} \in \{\boldsymbol{s}_1, \boldsymbol{s}_2, \boldsymbol{s}_3, \boldsymbol{s}_4, \boldsymbol{s}_5, \boldsymbol{s}_6, \boldsymbol{s}_7, \boldsymbol{s}_8, \boldsymbol{s}_9\})$, $E(Y(t_1)|\boldsymbol{S} \in \{\boldsymbol{s}_1, \boldsymbol{s}_2, \boldsymbol{s}_3, \boldsymbol{s}_{13}\})$, $E(Y(t_1)|\boldsymbol{S} \in \{\boldsymbol{s}_1, \boldsymbol{s}_4, \boldsymbol{s}_7, \boldsymbol{s}_{10}\})$,
$E(Y(t_2)|\boldsymbol{S} \in \{\boldsymbol{s}_{11}, \boldsymbol{s}_{12}\})$, $E(Y(t_2)|\boldsymbol{S} \in \{\boldsymbol{s}_4, \boldsymbol{s}_5, \boldsymbol{s}_6, \boldsymbol{s}_{14}\})$, $E(Y(t_2)|\boldsymbol{S} \in \{\boldsymbol{s}_2, \boldsymbol{s}_5, \boldsymbol{s}_{11}\})$,
$E(Y(t_3)|\boldsymbol{S} \in \{\boldsymbol{s}_{10}, \boldsymbol{s}_{13}, \boldsymbol{s}_{14}, \boldsymbol{s}_{15}\})$, $E(Y(t_3)|\boldsymbol{S} \in \{\boldsymbol{s}_7, \boldsymbol{s}_8, \boldsymbol{s}_9, \boldsymbol{s}_{10}, \boldsymbol{s}_{15}\})$, $E(Y(t_3)|\boldsymbol{S} \in \{\boldsymbol{s}_3, \boldsymbol{s}_6, \boldsymbol{s}_9, \boldsymbol{s}_{12}, \boldsymbol{s}_{13}, \boldsymbol{s}_{14}, \boldsymbol{s}_{15}\})$.

Revealed Preference Analysis

Revealed preference analysis is more effective in eliminating response types than monotonicity conditions (124)–(125). Table A.6 applies the WARP choice rule in **P.1** to the incentive matrix (123). There are 22 binding restrictions. Table A.7 summarise these 22 choice restrictions of Table A.6 into the five restrictions. The remaining restrictions do not eliminate any additional response types.

The WARP restriction in **P.1** translates the incentive matrix 30 into the five choice restrictions:

1	$T_i(z_0) = t_0$	$\Rightarrow$	$T_i(z_1) \neq t_2$ and $T_i(z_2) \neq t_1$
2	$T_i(z_0) = t_1$	$\Rightarrow$	$T_i(z_1) = t_1$ and $T_i(z_2) \neq t_0$
3	$T_i(z_0) = t_2$	$\Rightarrow$	$T_i(z_1) \neq t_0$ and $T_i(z_2) = t_2$
4	$T_i(z_1) = t_2$	$\Rightarrow$	$T_i(z_0) = t_2$ and $T_i(z_2) = t_2$
5	$T_i(z_2) = t_1$	$\Rightarrow$	$T_i(z_0) = t_1$ and $T_i(z_1) = t_1$

The first restriction consider an agent that chooses t_0 under z_0. This mean that t_0 is revealed preferred to t_1 and t_2 when no incentives are available. The IV value z_1 offers no incentives towards t_2, hence t_2 is not chosen under z_1. In same token, z_2 does not incentivize t_1 and therefore t_1 is not chosen. Panel C of Table (A.5) shows that these five restrictions eliminate 19 out of the 27 possible response types. The eight response types that survive the elimination process are displayed in the

Table A.5: Elimination of Response types of the Three-choice Model with a Parallel Design

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
$T_i(z_0)$	t_0	t_0	t_0	t_0	t_0	t_0	t_0	t_0	t_0	t_1	t_1	t_1	t_1	t_1	t_1	t_1	t_1	t_1	t_2	t_2	t_2	t_2	t_2	t_2	t_2	t_2	t_2
$T_i(z_1)$	t_0	t_0	t_0	t_1	t_1	t_1	t_2	t_2	t_2	t_0	t_0	t_0	t_1	t_1	t_1	t_2	t_2	t_2	t_0	t_0	t_0	t_1	t_1	t_1	t_2	t_2	t_2
$T_i(z_2)$	t_0	t_1	t_2	t_0	t_1	t_2	t_0	t_1	t_2	t_0	t_1	t_2	t_0	t_1	t_2	t_0	t_1	t_2	t_0	t_1	t_2	t_0	t_1	t_2	t_0	t_1	t_2
Panel B	**Response type Eliminated by Monotonicity Conditions (124)–(125)**																										
Condition 1	✓	✓	✓	✓	✓	✓	✓	✓	✓	✗	✗	✗	✓	✓	✓	✗	✗	✗	✓	✓	✓	✓	✓	✓	✓	✓	✓
Condition 2	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✗	✗	✓	✗	✗	✓	✗	✗	✓
Not Eliminated	1	2	3	4	5	6	7	8	9				13	14	15						21			24			27
Panel C	**Response type Eliminated by Revealed Preference Analysis**																										
Restriction 1	✓	✗	✓	✓	✗	✓	✗	✗	✗	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Restriction 2	✓	✓	✓	✓	✓	✓	✓	✓	✓	✗	✗	✗	✗	✓	✓	✗	✗	✗	✓	✓	✓	✓	✓	✓	✓	✓	✓
Restriction 3	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✗	✗	✗	✗	✗	✓	✗	✗	✓
Restriction 4	✓	✓	✓	✓	✓	✓	✗	✗	✗	✓	✓	✓	✓	✓	✓	✗	✗	✗	✓	✓	✓	✓	✓	✓	✗	✗	✓
Restriction 5	✓	✗	✓	✓	✗	✓	✓	✗	✓	✓	✗	✓	✓	✓	✓	✓	✗	✓	✓	✗	✓	✓	✗	✓	✓	✗	✓
Not Eliminated	1		3	4		6								14	15									24			27

Panel A lists the 27 possible response types that the response variable $S_i = [T_i(z_0)\ ,\ T_i(z_1)\ ,\ T_i(z_2)]$ can take. Rows present the counterfactual choices an agent i could choose if it were assigned to z_0, z_1, and z_2 respectively. Columns present all the values of response type as choices range over $supp(T) = \{t_0, t_1, t_2\}$. Panel B describes an elimination process based on the two monotonicity conditions (124)–(125). These criteria are also stated in Panel B′ below. Panel C describes an elimination process based on the seven choice restrictions generated by the revealed preference analysis. These choice restrictions are also displayed in Panel C′ below.

Panel B′ – Monotonicity Relations

Monotonicity Condition 1	$\mathbf{1}[T_i(z_0) = t_1]$	$\leq$	$\mathbf{1}[T_i(z_1) = t_1]$
Monotonicity Condition 2	$\mathbf{1}[T_i(z_0) = t_2]$	$\leq$	$\mathbf{1}[T_i(z_2) = t_2]$

Panel C′ – Choice Restrictions

Choice Restriction 1	$T_i(z_0) = t_0$	$\Rightarrow$	$T_i(z_1) \neq t_2$	and	$T_i(z_2) \neq t_1$
Choice Restriction 2	$T_i(z_0) = t_1$	$\Rightarrow$	$T_i(z_1) = t_1$	and	$T_i(z_2) \neq t_0$
Choice Restriction 3	$T_i(z_0) = t_2$	$\Rightarrow$	$T_i(z_1) \neq t_0$	and	$T_i(z_2) = t_2$
Choice Restriction 4	$T_i(z_1) = t_2$	$\Rightarrow$	$T_i(z_0) = t_2$	and	$T_i(z_2) = t_2$
Choice Restriction 5	$T_i(z_2) = t_1$	$\Rightarrow$	$T_i(z_0) = t_1$	and	$T_i(z_1) = t_1$

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.

response matrix below:

$$\boldsymbol{R} = \begin{array}{c} \\ \\ \\ \end{array} \begin{array}{c} \begin{array}{cccccccc} \boldsymbol{s}_1 & \boldsymbol{s}_2 & \boldsymbol{s}_3 & \boldsymbol{s}_4 & \boldsymbol{s}_5 & \boldsymbol{s}_6 & \boldsymbol{s}_7 & \boldsymbol{s}_8 \end{array} \\ \begin{bmatrix} t_1 & t_1 & t_0 & t_0 & t_2 & t_0 & t_0 & t_2 \\ t_1 & t_1 & t_1 & t_1 & t_1 & t_0 & t_0 & t_2 \\ t_1 & t_2 & t_0 & t_2 & t_2 & t_0 & t_2 & t_2 \end{bmatrix} \end{array} \begin{array}{l} \\ T_i(z_0) \\ T_i(z_1) \\ T_i(z_2) \end{array}$$

The main paper shows that response matrix above satisfies the monotonicity condition of Angrist and Imbens (1995).

Table A.6: Choice Restrictions Due to WARP

#	Revealed Choice	Incentive Inequalities					Choice Statement
	$T_i(z) = t$	$\boldsymbol{L}[z',t'] - \boldsymbol{L}[z,t'] \le 0 \le \boldsymbol{L}[z',t] - \boldsymbol{L}[z,t]$					$T(z') \neq t'$
1	$T_i(z_0) = t_0,$	$\boldsymbol{L}[z_2,t_1] - \boldsymbol{L}[z_0,t_1] = 0$	$\le$	0	$\le$	$0 = \boldsymbol{L}[z_2,t_0] - \boldsymbol{L}[z_0,t_0]$	$T_i(z_2) \neq t_1$
2	$T_i(z_0) = t_0,$	$\boldsymbol{L}[z_1,t_2] - \boldsymbol{L}[z_0,t_2] = 0$	$\le$	0	$\le$	$0 = \boldsymbol{L}[z_1,t_0] - \boldsymbol{L}[z_0,t_0]$	$T_i(z_1) \neq t_2$
3	$T_i(z_0) = t_1,$	$\boldsymbol{L}[z_1,t_0] - \boldsymbol{L}[z_0,t_0] = 0$	$\le$	0	$\le$	$1 = \boldsymbol{L}[z_1,t_1] - \boldsymbol{L}[z_0,t_1]$	$T_i(z_1) \neq t_0$
4	$T_i(z_0) = t_1,$	$\boldsymbol{L}[z_2,t_0] - \boldsymbol{L}[z_0,t_0] = 0$	$\le$	0	$\le$	$0 = \boldsymbol{L}[z_2,t_1] - \boldsymbol{L}[z_0,t_1]$	$T_i(z_2) \neq t_0$
5	$T_i(z_0) = t_1,$	$\boldsymbol{L}[z_1,t_2] - \boldsymbol{L}[z_0,t_2] = 0$	$\le$	0	$\le$	$1 = \boldsymbol{L}[z_1,t_1] - \boldsymbol{L}[z_0,t_1]$	$T_i(z_1) \neq t_2$
6	$T_i(z_0) = t_2,$	$\boldsymbol{L}[z_1,t_0] - \boldsymbol{L}[z_0,t_0] = 0$	$\le$	0	$\le$	$0 = \boldsymbol{L}[z_1,t_2] - \boldsymbol{L}[z_0,t_2]$	$T_i(z_1) \neq t_0$
7	$T_i(z_0) = t_2,$	$\boldsymbol{L}[z_2,t_0] - \boldsymbol{L}[z_0,t_0] = 0$	$\le$	0	$\le$	$1 = \boldsymbol{L}[z_2,t_2] - \boldsymbol{L}[z_0,t_2]$	$T_i(z_2) \neq t_0$
8	$T_i(z_0) = t_2,$	$\boldsymbol{L}[z_2,t_1] - \boldsymbol{L}[z_0,t_1] = 0$	$\le$	0	$\le$	$1 = \boldsymbol{L}[z_2,t_2] - \boldsymbol{L}[z_0,t_2]$	$T_i(z_2) \neq t_1$
9	$T_i(z_1) = t_0,$	$\boldsymbol{L}[z_0,t_1] - \boldsymbol{L}[z_1,t_1] = -1$	$\le$	0	$\le$	$0 = \boldsymbol{L}[z_0,t_0] - \boldsymbol{L}[z_1,t_0]$	$T_i(z_0) \neq t_1$
10	$T_i(z_1) = t_0,$	$\boldsymbol{L}[z_2,t_1] - \boldsymbol{L}[z_1,t_1] = -1$	$\le$	0	$\le$	$0 = \boldsymbol{L}[z_2,t_0] - \boldsymbol{L}[z_1,t_0]$	$T_i(z_2) \neq t_1$
11	$T_i(z_1) = t_0,$	$\boldsymbol{L}[z_0,t_2] - \boldsymbol{L}[z_1,t_2] = 0$	$\le$	0	$\le$	$0 = \boldsymbol{L}[z_0,t_0] - \boldsymbol{L}[z_1,t_0]$	$T_i(z_0) \neq t_2$
12	$T_i(z_1) = t_2,$	$\boldsymbol{L}[z_0,t_0] - \boldsymbol{L}[z_1,t_0] = 0$	$\le$	0	$\le$	$0 = \boldsymbol{L}[z_0,t_2] - \boldsymbol{L}[z_1,t_2]$	$T_i(z_0) \neq t_0$
13	$T_i(z_1) = t_2,$	$\boldsymbol{L}[z_2,t_0] - \boldsymbol{L}[z_1,t_0] = 0$	$\le$	0	$\le$	$1 = \boldsymbol{L}[z_2,t_2] - \boldsymbol{L}[z_1,t_2]$	$T_i(z_2) \neq t_0$
14	$T_i(z_1) = t_2,$	$\boldsymbol{L}[z_0,t_1] - \boldsymbol{L}[z_1,t_1] = -1$	$\le$	0	$\le$	$0 = \boldsymbol{L}[z_0,t_2] - \boldsymbol{L}[z_1,t_2]$	$T_i(z_0) \neq t_1$
15	$T_i(z_1) = t_2,$	$\boldsymbol{L}[z_2,t_1] - \boldsymbol{L}[z_1,t_1] = -1$	$\le$	0	$\le$	$1 = \boldsymbol{L}[z_2,t_2] - \boldsymbol{L}[z_1,t_2]$	$T_i(z_2) \neq t_1$
16	$T_i(z_2) = t_0,$	$\boldsymbol{L}[z_0,t_1] - \boldsymbol{L}[z_2,t_1] = 0$	$\le$	0	$\le$	$0 = \boldsymbol{L}[z_0,t_0] - \boldsymbol{L}[z_2,t_0]$	$T_i(z_0) \neq t_1$
17	$T_i(z_2) = t_0,$	$\boldsymbol{L}[z_0,t_2] - \boldsymbol{L}[z_2,t_2] = -1$	$\le$	0	$\le$	$0 = \boldsymbol{L}[z_0,t_0] - \boldsymbol{L}[z_2,t_0]$	$T_i(z_0) \neq t_2$
18	$T_i(z_2) = t_0,$	$\boldsymbol{L}[z_1,t_2] - \boldsymbol{L}[z_2,t_2] = -1$	$\le$	0	$\le$	$0 = \boldsymbol{L}[z_1,t_0] - \boldsymbol{L}[z_2,t_0]$	$T_i(z_1) \neq t_2$
19	$T_i(z_2) = t_1,$	$\boldsymbol{L}[z_0,t_0] - \boldsymbol{L}[z_2,t_0] = 0$	$\le$	0	$\le$	$0 = \boldsymbol{L}[z_0,t_1] - \boldsymbol{L}[z_2,t_1]$	$T_i(z_0) \neq t_0$
20	$T_i(z_2) = t_1,$	$\boldsymbol{L}[z_1,t_0] - \boldsymbol{L}[z_2,t_0] = 0$	$\le$	0	$\le$	$1 = \boldsymbol{L}[z_1,t_1] - \boldsymbol{L}[z_2,t_1]$	$T_i(z_1) \neq t_0$
21	$T_i(z_2) = t_1,$	$\boldsymbol{L}[z_0,t_2] - \boldsymbol{L}[z_2,t_2] = -1$	$\le$	0	$\le$	$0 = \boldsymbol{L}[z_0,t_1] - \boldsymbol{L}[z_2,t_1]$	$T_i(z_0) \neq t_2$
22	$T_i(z_2) = t_1,$	$\boldsymbol{L}[z_1,t_2] - \boldsymbol{L}[z_2,t_2] = -1$	$\le$	0	$\le$	$1 = \boldsymbol{L}[z_1,t_1] - \boldsymbol{L}[z_2,t_1]$	$T_i(z_1) \neq t_2$

This table displays the binding choice restrictions generated by the WARP choice rule (Lemma L.1) described below:

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

when applied to the MTO incentive matrix below:

$$\text{Incentive Matrix } \boldsymbol{L} = \begin{array}{c} \begin{matrix} t_0 & t_1 & t_2 \end{matrix} \\ \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \end{array} \begin{matrix} z_0 \\ z_1 \\ z_2 \end{matrix}$$

Table A.7: Summary of Choice Restrictions generated by applying WARP to the Parallel Design Model in Table A.6

#	**Choice Restrictions**				
1,2	$T_i(z_0) = t_0$	$\Rightarrow$	$T_i(z_1) \neq t_2$	and	$T_i(z_2) \neq t_1$
3,4,5	$T_i(z_0) = t_1$	$\Rightarrow$	$T_i(z_1) = t_1$	and	$T_i(z_2) \neq t_0$
6,7,8	$T_i(z_0) = t_2$	$\Rightarrow$	$T_i(z_1) \neq t_0$	and	$T_i(z_2) = t_2$
12,13,14,15	$T_i(z_1) = t_2$	$\Rightarrow$	$T_i(z_0) = t_2$	and	$T_i(z_2) = t_2$
19,20,21,22	$T_i(z_2) = t_1$	$\Rightarrow$	$T_i(z_0) = t_1$	and	$T_i(z_1) = t_1$

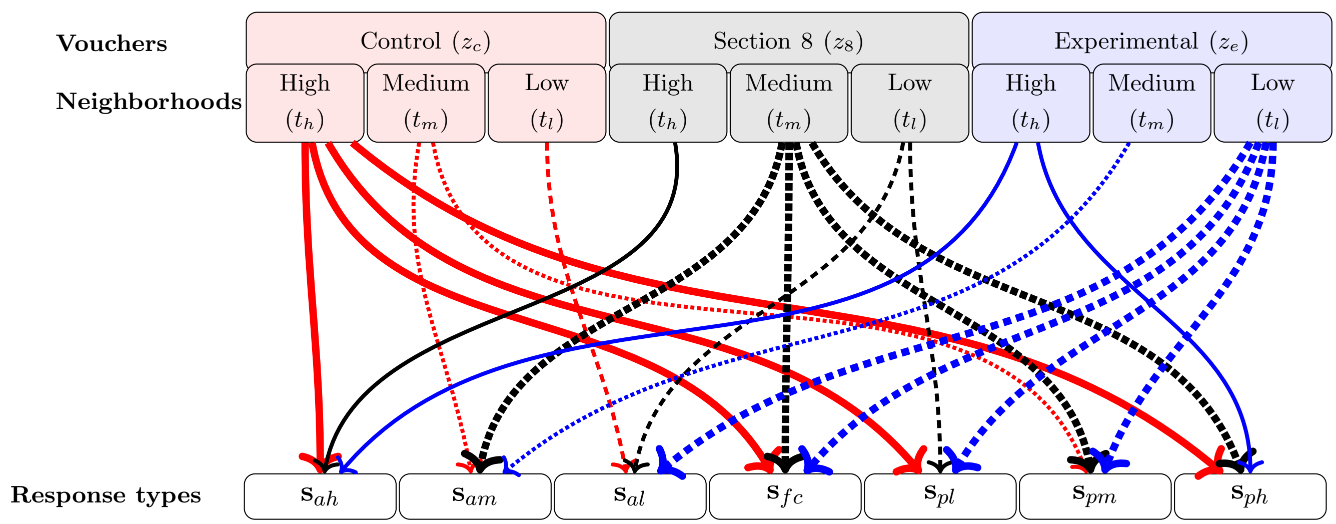
F Exploring the Properties of the Response Matrix

The response matrix summarizes the necessary and sufficient information to investigate the non-parametric identification of response type probabilities and counterfactual outcomes. This section discusses three topics regarding the usage of the response matrix. Section F.1 clarifies how the response matrix enables us to map counterfactuals and observed data. Section F.2 discusses the identification of counterfactual outcomes. The section provides closed-form solutions for each identified counterfactual. Section F.3 explains how to estimate the identified counterfactual outcomes using stands 2SLS regressions.

F.1 Mapping Counterfactual outcomes with Observed Data

The matrix determines a mapping between observed choices and latent response types. For instance, the first row of the response matrix (29) lists the choices for control z_c. Families that choose t_l under z_c can only be low-poverty always-takers $\boldsymbol{s}_{al}$. Families that choose t_m under z_c are a mixture of $\boldsymbol{s}_{al}$ and $\boldsymbol{s}_{al}$, while those who choose t_l under z_c can be of four types: $\boldsymbol{s}_{ah}, \boldsymbol{s}_{am}, \boldsymbol{s}_{fc}, \boldsymbol{s}_{pl}$ or , $\boldsymbol{s}_{ph}$. Figure A.4 displays the mapping generated by the response matrix (29). The identification of causal parameters consists of disentangling this mapping.

Figure A.4: **From Observed Vouchers and Choices to Unobserved Response types**



$T(z_c)$	t_h	t_m	t_l	t_h	t_h	t_m	t_h
$T(z_8)$	t_h	t_m	t_l	t_m	t_l	t_m	t_m
$T(z_e)$	t_h	t_m	t_l	t_l	t_l	t_l	t_h

This figure describes how voucher assignments and neighborhood choices map into the MTO response types. There are three possible voucher assignments: Control (z_c), Section 8 (z_8), or Experimental (z_e). There are three neighborhood choices: high-poverty neighborhood (t_h), medium-poverty neighborhood (t_m) or low-poverty neighborhood (t_l). The combination of voucher assignment and neighborhood choice generate nine possibilities. There are seven response types according to the response matrix $\boldsymbol{R}$ in (29). These response types are denoted by $\boldsymbol{s}_{ah}, \boldsymbol{s}_{am}, \boldsymbol{s}_{al}, \boldsymbol{s}_{fc}, \boldsymbol{s}_{ph}, \boldsymbol{s}_{pm}, \boldsymbol{s}_{pl}$. The mapping between the voucher assignments and neighborhood choices into response types is represented by connecting lines. Solid lines denote the choice of high-poverty neighborhood. Dotted lines denote the choice of medium-poverty neighborhood. Dashed lines denote the choice of low-poverty neighborhood. Bold lines refer to the most frequent neighborhood choice for each voucher assignment.

F.2 Interpreting Identification Results

Equation (8) is central to the identification analysis. It shows that the indicator $\mathbf{1}[T=t|\boldsymbol{S}=\boldsymbol{s},Z=z]$ connects observed data, i.e., the expectation of the outcome multiplied by the choice indicator, with the unobserved parameters we seek to identify, i.e. potential outcomes $E(Y(t)|\boldsymbol{S}=\boldsymbol{s})$ and response type probabilities $P(\boldsymbol{S}=\boldsymbol{s})$. The identification of counterfactual parameters consists of expressing the unobserved variables in the right-hand side of (8) in terms of the observed variables of the left-hand side. This problem is best examined by expressing equation (8) in matrix form:

$$\boldsymbol{Q}_Z(t)\odot\boldsymbol{P}_Z(t)=\boldsymbol{B}_t\cdot\Big(\boldsymbol{Q}_S(t)\odot\boldsymbol{P}_S\Big);\quad t\in\{t_h,t_m,t_l\},\tag{127}$$

$$\begin{aligned}\text{where}\quad \boldsymbol{P}_Z(t)&=[P(T=t|Z=z_c),P(T=t|Z=z_8),P(T=t|Z=z_e)]',\\ \boldsymbol{Q}_Z(t)&=[E(Y|T=t,Z=z_c),E(Y|T=t,Z=z_8),E(Y|T=t,Z=z_e)]',\\ \boldsymbol{P}_S&=[P(\boldsymbol{S}=\boldsymbol{s}_{ah}),P(\boldsymbol{S}=\boldsymbol{s}_{am}),P(\boldsymbol{S}=\boldsymbol{s}_{al}),P(\boldsymbol{S}=\boldsymbol{s}_{fc}),P(\boldsymbol{S}=\boldsymbol{s}_{pl}),P(\boldsymbol{S}=\boldsymbol{s}_{pm}),P(\boldsymbol{S}=\boldsymbol{s}_{ph})]',\\ \boldsymbol{Q}_S(t)&=[E(Y(t)|\boldsymbol{S}=\boldsymbol{s}_{ah}),...,E(Y(t)|\boldsymbol{S}=\boldsymbol{s}_{ph})]',\\ \boldsymbol{B}_t&=\mathbf{1}[\boldsymbol{R}=t];t\in\{t_l,t_m,t_h\}.\end{aligned}$$

$\boldsymbol{P}_Z(t)$ denotes the observed vector of propensity scores. $\boldsymbol{Q}_Z(t)$ denotes the observed vector of conditional outcomes. $\boldsymbol{P}_S$ is the 7×1 vector of response type probabilities. $\boldsymbol{Q}_S(t)$ is the unobserved vector of counterfactual outcome means. $\boldsymbol{B}_t=\mathbf{1}[\boldsymbol{R}=t]$ denotes the 3×7 binary matrix that indicates which elements in $\boldsymbol{R}$ are equal to $t\in\{t_h,t_m,t_l\}$ and $\odot$ denotes the Hadamard product (element-wise multiplication).

The the binary matrices $\boldsymbol{B}_t=\mathbf{1}[\boldsymbol{R}=t]$ for t_l, t_m, and t_h are displayed in equations (129),(130) and (131) respectively. It is useful to decompose each binary matrix $\boldsymbol{B}_t$ into $\boldsymbol{B}_t=\boldsymbol{C}_t\cdot\boldsymbol{A}_t$, where $\boldsymbol{C}_t$ is the array the non-zero columns of $\boldsymbol{B}_t$ and $\boldsymbol{A}_t$ is a mapping between the vectors in $\boldsymbol{C}_t$ and $\boldsymbol{B}_t$. Specifically, the response matrix $\boldsymbol{R}$ is decomposed as:

$$\boldsymbol{R}\equiv\sum_{t\in supp(T)}t\cdot\boldsymbol{B}_t=\sum_{t\in supp(T)}t\cdot\boldsymbol{C}_t\boldsymbol{A}_t,\tag{128}$$

where $\boldsymbol{B}_t,\boldsymbol{C}_t,\boldsymbol{A}_t$ for $t\in\{t_h,t_m,t_h\}$ are given by:

$$\boldsymbol{B}_{t_h}=\begin{array}{c}\begin{matrix}\boldsymbol{s}_{ah}&\boldsymbol{s}_{am}&\boldsymbol{s}_{al}&\boldsymbol{s}_{fc}&\boldsymbol{s}_{pl}&\boldsymbol{s}_{pm}&\boldsymbol{s}_{ph}\end{matrix}\\\begin{bmatrix}1&0&0&1&1&0&1\\1&0&0&0&0&0&0\\1&0&0&0&0&0&1\end{bmatrix}\end{array}=\underbrace{\begin{array}{c}\begin{matrix}\boldsymbol{s}_{fc},\boldsymbol{s}_{pl}&\boldsymbol{s}_{ph}&\boldsymbol{s}_{ah}\end{matrix}\\\begin{bmatrix}1&1&1\\0&0&1\\0&1&1\end{bmatrix}\end{array}}_{\boldsymbol{C}_{t_h}}\cdot\underbrace{\begin{array}{c}\begin{matrix}\boldsymbol{s}_{ah}&\boldsymbol{s}_{am}&\boldsymbol{s}_{al}&\boldsymbol{s}_{fc}&\boldsymbol{s}_{pl}&\boldsymbol{s}_{pm}&\boldsymbol{s}_{ph}\end{matrix}\\\begin{bmatrix}0&0&0&1&1&0&0\\0&0&0&0&0&0&1\\1&0&0&0&0&0&0\end{bmatrix}\end{array}}_{\boldsymbol{A}_{t_h}}\tag{129}$$

$$\boldsymbol{B}_{t_m}=\begin{array}{c}\begin{matrix}\boldsymbol{s}_{ah}&\boldsymbol{s}_{am}&\boldsymbol{s}_{al}&\boldsymbol{s}_{fc}&\boldsymbol{s}_{pl}&\boldsymbol{s}_{pm}&\boldsymbol{s}_{ph}\end{matrix}\\\begin{bmatrix}0&1&0&0&0&1&0\\0&1&0&1&0&1&1\\0&1&0&0&0&0&0\end{bmatrix}\end{array}=\underbrace{\begin{array}{c}\begin{matrix}\boldsymbol{s}_{fc},\boldsymbol{s}_{ph}&\boldsymbol{s}_{pm}&\boldsymbol{s}_{am}\end{matrix}\\\begin{bmatrix}0&1&1\\1&1&1\\0&0&1\end{bmatrix}\end{array}}_{\boldsymbol{C}_{t_m}}\cdot\underbrace{\begin{array}{c}\begin{matrix}\boldsymbol{s}_{ah}&\boldsymbol{s}_{am}&\boldsymbol{s}_{al}&\boldsymbol{s}_{fc}&\boldsymbol{s}_{pl}&\boldsymbol{s}_{pm}&\boldsymbol{s}_{ph}\end{matrix}\\\begin{bmatrix}0&0&0&1&0&0&1\\0&0&0&0&0&1&0\\0&1&0&0&0&0&0\end{bmatrix}\end{array}}_{\boldsymbol{A}_{t_m}}\tag{130}$$

$$
\boldsymbol{B}_{t_l} = \overset{\begin{matrix}\boldsymbol{s}_{ah} & \boldsymbol{s}_{am} & \boldsymbol{s}_{al} & \boldsymbol{s}_{fc} & \boldsymbol{s}_{pl} & \boldsymbol{s}_{pm} & \boldsymbol{s}_{ph}\end{matrix}}{\begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 1 & 0 \end{bmatrix}} = \underbrace{\overset{\begin{matrix}\boldsymbol{s}_{fc},\boldsymbol{s}_{pm} & \boldsymbol{s}_{pl} & \boldsymbol{s}_{al}\end{matrix}}{\begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}}}_{\boldsymbol{C}_{t_l}} \cdot \underbrace{\overset{\begin{matrix}\boldsymbol{s}_{ah} & \boldsymbol{s}_{am} & \boldsymbol{s}_{al} & \boldsymbol{s}_{fc} & \boldsymbol{s}_{pl} & \boldsymbol{s}_{pm} & \boldsymbol{s}_{ph}\end{matrix}}{\begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \end{bmatrix}}}_{\boldsymbol{A}_{t_l}} \tag{131}
$$

Matrices $\boldsymbol{C}_t$ and $\boldsymbol{A}_t$ are used to generate a closed-form solution for the nonparametric identification of counterfactual outcomes. The main paper shows that for each neighborhood choice, we can reorder the columns and rows of the response matrix $\boldsymbol{R}$ to generate a lower-triangular matrix (see equations (35) and (42)). This triangular property means that for each t we have that:

$$
\text{for any} z, z' \in supp(Z), \text{ we have that } \boldsymbol{B}_t[z, \boldsymbol{s}] \leq \boldsymbol{B}_t[z', \boldsymbol{s}] \forall\, \boldsymbol{s} \text{ or } \boldsymbol{B}_t[z, \boldsymbol{s}] \geq \boldsymbol{B}_t[z', \boldsymbol{s}] \forall\, \boldsymbol{s}. \tag{132}
$$

Equation (132) is equivalent to state that there exists a sequence of of IV-values $z_1^{(t)}, ... z_N^{(t)}$ of the values in $supp(Z) \equiv \{z_1, ..., z_N\}$ such that:

$$
\boldsymbol{B}_t[z_k^{(t)}, \boldsymbol{s}] \leq \boldsymbol{B}_t[z_{k+1}^{(t)}, \boldsymbol{s}] \;\forall\, \boldsymbol{s} \in supp(\boldsymbol{S}); k = 1, .., N-1. \tag{133}
$$

This triangular property implies that matrices $\boldsymbol{C}_t$ in the decompositions (129)–(131) are of full row-rank. We can then use the generalized solution of linear equations in Magnus and Neudecker (1999) to identify counterfactual outcomes the following equation:

$$
\Big(\boldsymbol{A}_t\big(\boldsymbol{Q}_S(t) \odot \boldsymbol{P}_S\big) \div \Big(\boldsymbol{A}_t \boldsymbol{P}_S\Big) = \Big((\boldsymbol{C}_t'\boldsymbol{C}_t)^{-1}\boldsymbol{C}_t'\big(\boldsymbol{Q}_Z(t) \odot \boldsymbol{P}_Z(t)\Big) \div \Big(((\boldsymbol{C}_t'\boldsymbol{C}_t)^{-1}\boldsymbol{C}_t'\boldsymbol{P}_Z(t)\Big) \tag{134}
$$

where $\div$ denotes element-wise division,[70] and $\boldsymbol{A}_t$ stems from the decomposition $\boldsymbol{B}_t = \boldsymbol{C}_t\boldsymbol{A}_t$ as in (129)–(131) Moreover, in the case where $\boldsymbol{C}_t$ is invertible, equation (134) can be further simplified as:

$$
\underbrace{\Big(\boldsymbol{A}_t\big(\boldsymbol{Q}_S(t) \odot \boldsymbol{P}_S\Big) \div \Big(\boldsymbol{A}_t \boldsymbol{P}_S\Big)}_{\text{Identified Counterfactual Outcomes}} = \underbrace{\Big(\boldsymbol{C}_t^{-1}\big(\boldsymbol{Q}_Z(t) \odot \boldsymbol{P}_Z(t)\Big) \div \Big((\boldsymbol{C}_t^{-1}\boldsymbol{P}_Z(t)\Big)}_{\text{Identification Formulas}} \tag{135}
$$

The right-hand side of (135) summarizes all identified counterfactual outcomes. The left-hand side of (135) generates identification formulas. Equations (136)–(138) exemplify the left-hand side of (135) for t_h.

$$
\boldsymbol{A}_{t_h}\big(\boldsymbol{Q}_S(t_h) \odot \boldsymbol{P}_S\big) = \begin{bmatrix} E(Y(t_h)|\boldsymbol{S} = \boldsymbol{s}_{fc})\, P(\boldsymbol{S} = \boldsymbol{s}_{fc}) + E(Y(t_h)|\boldsymbol{S} = \boldsymbol{s}_{pl})\, P(\boldsymbol{S} = \boldsymbol{s}_{pl}) \\ E(Y(t_h)|\boldsymbol{S} = \boldsymbol{s}_{ph})\, P(\boldsymbol{S} = \boldsymbol{s}_{ph}) \\ E(Y(t_h)|\boldsymbol{S} = \boldsymbol{s}_{ah})\, P(\boldsymbol{S} = \boldsymbol{s}_{ah}) \end{bmatrix} \tag{136}
$$

$$
\boldsymbol{A}_{t_h}\boldsymbol{P}_S = \begin{bmatrix} P(\boldsymbol{S} = \boldsymbol{s}_{fc}) + P(\boldsymbol{S} = \boldsymbol{s}_{pl}) \\ P(\boldsymbol{S} = \boldsymbol{s}_{ph}) \\ P(\boldsymbol{S} = \boldsymbol{s}_{ah}) \end{bmatrix} \tag{137}
$$

$$
\therefore \Big(\boldsymbol{A}_{t_h}\big(\boldsymbol{Q}_S(t_h) \odot \boldsymbol{P}_S\big)\Big) \div \Big(\boldsymbol{A}_{t_h}\boldsymbol{P}_S\Big) = \begin{bmatrix} E(Y(t_h)|\boldsymbol{S} \in \boldsymbol{s}_{fc}, \boldsymbol{s}_{pl}) \\ E(Y(t_h)|\boldsymbol{S} = \boldsymbol{s}_{ph}) \\ E(Y(t_h)|\boldsymbol{S} = \boldsymbol{s}_{ah}) \end{bmatrix} \tag{138}
$$

[70] Let $\boldsymbol{A}, \boldsymbol{B}$ be two vectors of same length, then $\boldsymbol{A} \div \boldsymbol{B} \equiv \text{diag}(\boldsymbol{B})^{-1}\boldsymbol{A}$, where $\text{diag}(\cdot)$ is the operator that transform a vector into a diagonal matrix.

The right-hand side of (135) for t_h employs the matrix $\boldsymbol{C}_{t_h}^{-1}$ displayed in equation (139):

$$\boldsymbol{C}_{t_h} = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix} \Rightarrow \boldsymbol{C}_{t_h}^{-1} = \begin{bmatrix} 1 & 0 & -1 \\ 0 & -1 & 1 \\ 0 & 1 & 0 \end{bmatrix}, \tag{139}$$

Equations (140) and (141) exemplify the numerator and the denominators of the right-hand side of (135) for t_h :

$$\begin{aligned} \boldsymbol{C}_{t_h}^{-1}\boldsymbol{Q}_Z(t_h) \odot \boldsymbol{P}_Z(t_h) &= \begin{bmatrix} E(Y|T=t_h, Z=z_c)\, P(T=t_h|Z=z_c) - E(Y|T=t_h, Z=z_e)\, P(T=t_h|Z=z_e) \\ E(Y|T=t_h, Z=z_e)\, P(T=t_h|Z=z_e) - E(Y|T=t_h, Z=z_8)\, P(T=t_h|Z=z_8) \\ E(Y|T=t_h, Z=z_8)\, P(T=t_h|Z=z_8) \end{bmatrix}, \\ &= \begin{bmatrix} E(Y \cdot D_{t_h}|Z=z_c) - E(Y \cdot \mathbf{1}[T=t_h]|Z=z_e) \\ E(Y \cdot D_{t_h}|Z=z_e) - E(Y \cdot \mathbf{1}[T=t_h]|Z=z_8) \\ E(Y \cdot D_{t_h}|Z=z_8) \end{bmatrix}, \end{aligned} \tag{140}$$

$$\text{and } \boldsymbol{C}_{t_h}^{-1}\boldsymbol{P}_Z(t_h) = \begin{bmatrix} P(T=t_h|Z=z_c) - P(T=t_h|Z=z_e) \\ P(T=t_h|Z=z_e) - P(T=t_h|Z=z_8) \\ P(T=t_h|Z=z_8) \end{bmatrix}, \tag{141}$$

The final equation for t_h is presented in (142). The left-hand side of (142) lists all the identified counterfactual outcome means of $Y(t_h)$. The right-hand side provides the identification formulas that can be evaluated from observed data.

$$\therefore \underbrace{\begin{bmatrix} E(Y(t_h)|\boldsymbol{S} \in \boldsymbol{s}_{fc}, \boldsymbol{s}_{pl}) \\ E(Y(t_h)|\boldsymbol{S} = \boldsymbol{s}_{ph}) \\ E(Y(t_h)|\boldsymbol{S} = \boldsymbol{s}_{ah}) \end{bmatrix}}_{\boldsymbol{A}_{t_h}(\boldsymbol{Q}_S(t_h) \odot \boldsymbol{P}_S) \div \boldsymbol{A}_{t_h}\boldsymbol{P}_S} = \underbrace{\begin{bmatrix} \frac{E(Y \cdot D_{t_h}|Z=z_c) - E(Y \cdot D_{t_h}|Z=z_e)}{P(T=t_h|Z=z_c) - P(T=t_h|Z=z_e)} \\ \frac{E(Y \cdot D_{t_h}|Z=z_e) - E(Y \cdot D_{t_h}|Z=z_8)}{P(T=t_h|Z=z_e) - P(T=t_h|Z=z_8)} \\ \frac{E(Y \cdot D_{t_h}|Z=z_8)}{P(T=t_h|Z=z_8)} \end{bmatrix}}_{(\boldsymbol{C}_{t_h}^{-1}\boldsymbol{Q}_Z(t_h) \odot \boldsymbol{P}_Z(t_h)) \div \boldsymbol{C}_{t_h}^{-1}\boldsymbol{P}_Z(t_h)}. \tag{142}$$

Equations (143)–(144) arise from applying formula (135) to t_m and t_l :

$$\underbrace{\begin{bmatrix} E(Y(t_m)|\boldsymbol{S} \in \{\boldsymbol{s}_{fc}, \boldsymbol{s}_{ph}\}) \\ E(Y(t_m)|\boldsymbol{S} = \boldsymbol{s}_{pm}) \\ E(Y(t_m)|\boldsymbol{S} = \boldsymbol{s}_{am}) \end{bmatrix}}_{\boldsymbol{A}_{t_m}(\boldsymbol{Q}_S(t_m) \odot \boldsymbol{P}_S) \div \boldsymbol{A}_{t_m}\boldsymbol{P}_S} = \underbrace{\begin{bmatrix} \frac{E(Y \cdot D_{t_m}|Z=z_8) - E(Y \cdot D_{t_m}|Z=z_c)}{P(T=t_m|Z=z_8) - P(T=t_m|Z=z_c)} \\ \frac{E(Y \cdot D_{t_m}|Z=z_c) - E(Y \cdot D_{t_m}|Z=z_e)}{P(T=t_m|Z=z_c) - P(T=t_m|Z=z_e)} \\ \frac{E(Y \cdot D_{t_m}|Z=z_e)}{P(T=t_m|Z=z_e)} \end{bmatrix}}_{\boldsymbol{C}_{t_m}^{-1}(\boldsymbol{Q}_Z(t_m) \odot \boldsymbol{P}_Z(t_m)) \div \boldsymbol{C}_{t_m}^{-1}\boldsymbol{P}_Z(t_m)}, \tag{143}$$

$$\underbrace{\begin{bmatrix} E(Y(t_l)|\boldsymbol{S} \in \{\boldsymbol{s}_{fc}, \boldsymbol{s}_{pm}\}) \\ E(Y(t_l)|\boldsymbol{S} = \boldsymbol{s}_{pl}) \\ E(Y(t_l)|\boldsymbol{S} = \boldsymbol{s}_{al}) \end{bmatrix}}_{\boldsymbol{A}_{t_l}(\boldsymbol{Q}_S(t_l) \odot \boldsymbol{P}_S) \div \boldsymbol{A}_{t_l}\boldsymbol{P}_S} = \underbrace{\begin{bmatrix} \frac{E(Y \cdot D_{t_l}|Z=z_e) - E(Y \cdot D_{t_l}|Z=z_8)}{P(T=t_l|Z=z_e) - P(T=t_l|Z=z_8)} \\ \frac{E(Y \cdot D_{t_l}|Z=z_8) - E(Y \cdot D_{t_l}|Z=z_c)}{P(T=t_l|Z=z_8) - P(T=t_l|Z=z_c)} \\ \frac{E(Y \cdot D_{t_l}|Z=z_c)}{P(T=t_l|Z=z_c)} \end{bmatrix}}_{\boldsymbol{C}_{t_l}^{-1}(\boldsymbol{Q}_Z(t_l) \odot \boldsymbol{P}_Z(t_l)) \div \boldsymbol{C}_{t_l}^{-1}\boldsymbol{P}_Z(t_l)}. \tag{144}$$

Response type probabilities can be identified by equations $\boldsymbol{A}_t\boldsymbol{P}_S = \boldsymbol{C}_{t_h}^{-1}\boldsymbol{P}_Z(t_h)$ for $t = t_h, t_l, t_m$:

$$\therefore \underbrace{\begin{bmatrix} P(\boldsymbol{S} \in \boldsymbol{s}_{fc}, \boldsymbol{s}_{pl}) \\ P(\boldsymbol{S} = \boldsymbol{s}_{ph}) \\ P(\boldsymbol{S} = \boldsymbol{s}_{ah}) \end{bmatrix}}_{\boldsymbol{A}_{t_h}\boldsymbol{P}_S} = \underbrace{\begin{bmatrix} P(T=t_h|Z=z_c) - P(T=t_h|Z=z_e) \\ P(T=t_h|Z=z_e) - P(T=t_h|Z=z_8) \\ P(T=t_h|Z=z_8) \end{bmatrix}}_{\boldsymbol{C}_{t_h}^{-1}\boldsymbol{P}_Z(t_h)}. \tag{145}$$

$$\underbrace{\begin{bmatrix} P(\boldsymbol{S} \in \{\boldsymbol{s}_{fc}, \boldsymbol{s}_{ph}\}) \\ P(\boldsymbol{S} = \boldsymbol{s}_{pm}) \\ P(\boldsymbol{S} = \boldsymbol{s}_{am}) \end{bmatrix}}_{\boldsymbol{A}_{t_m}\boldsymbol{P}_S} = \underbrace{\begin{bmatrix} P(T = t_m|Z = z_8) - P(T = t_m|Z = z_c) \\ P(T = t_m|Z = z_c) - P(T = t_m|Z = z_e) \\ P(T = t_m|Z = z_e) \end{bmatrix}}_{\boldsymbol{C}_{t_m}^{-1}\boldsymbol{P}_{Z}(t_m)}, \tag{146}$$

$$\underbrace{\begin{bmatrix} P(\boldsymbol{S} \in \{\boldsymbol{s}_{fc}, \boldsymbol{s}_{pm}\}) \\ P(\boldsymbol{S} = \boldsymbol{s}_{pl}) \\ P(\boldsymbol{S} = \boldsymbol{s}_{al}) \end{bmatrix}}_{\boldsymbol{A}_{t_l}\boldsymbol{P}_S} = \underbrace{\begin{bmatrix} P(T = t_l|Z = z_e) - P(T = t_l|Z = z_8) \\ P(T = t_l|Z = z_8) - P(T = t_l|Z = z_c) \\ P(T = t_l|Z = z_c) \end{bmatrix}}_{\boldsymbol{C}_{t_l}^{-1}\boldsymbol{P}_{Z}(t_l)}. \tag{147}$$

F.3 Relation to Previous Literature on Binary Treatments

We can connect the triangular property of the response matrices in (35), (42), and (43), to the IV literature of binary choice models. Similar to Imbens and Angrist (1994), counterfactual outcomes can be estimated by Two-Stage Least Squares (2SLS). The identification of $E(Y(t_l)|\boldsymbol{S} = \boldsymbol{s}_{pl})$ in (39) depends on z_8 and z_c. According to equation (144), the counterfactual is identified by the following equations:

$$E(Y(t_l)|\boldsymbol{S} = \boldsymbol{s}_{pl}) = \frac{E(Y \cdot D_{t_l}|Z = z_8) - E(Y \cdot D_{t_l}|Z = z_c)}{P(T = t_l|Z = z_8) - P(T = t_l|Z = z_c)}.$$

The equation is closely related with the LATE equation of Imbens and Angrist (1994). It can be estimated by the 2SLS (148)–(149) that regresses the choice indicator D_{t_l} on two IV indicators, $\mathbf{1}[Z = z_8]$ and $\mathbf{1}[Z = z_c]$ without a constant term (first stage) and then regresses the interaction YD_{t_l} on a constant and the fitted values $\hat{D}_{t_l}$ (second stage):

$$\text{First Stage:} \quad D_{t_l} = \gamma_1\mathbf{1}[Z = z_8] + \gamma_2\mathbf{1}[Z = z_c] + \epsilon_D \tag{148}$$

$$\text{Second Stage:} \quad YD_{t_l} = \beta_0 + \beta_{IV}\hat{D}_{t_l} + \epsilon_Y, \tag{149}$$

γ_1, γ_2 are linear coefficients of the first stage, β_0 is the intercept of the second stage, and β_{IV} is the linear coefficient that estimates $E(Y(t_l)|\boldsymbol{S} = \boldsymbol{s}_{pl})$. We can estimate different counterfactual outcomes by varying the IV-indicators and neighborhood choices as listed in Table A.8.

Table A.8: Two-Stage Least Square Estimation for Identified Parameters

Data Transformations				Identified Parameters
Endogenous Variables Choice Indicator	**Dependent Variable** Outcome Interaction	**Instrumental Variable** IV Indicators		
$D_{t_h} \equiv \mathbf{1}[T = t_h]$	$D_{t_h} \cdot Y$	$\mathbf{1}[Z = z_c]$	$\mathbf{1}[Z = z_e]$	$E(Y(t_h)\|\boldsymbol{S} \in \{\boldsymbol{s}_{fc}, \boldsymbol{s}_{pl}\})$
		$\mathbf{1}[Z = z_8]$	$\mathbf{1}[Z = z_e]$	$E(Y(t_h)\|\boldsymbol{S} = \boldsymbol{s}_{ph})$
$D_{t_m} \equiv \mathbf{1}[T = t_m]$	$D_{t_m} \cdot Y$	$\mathbf{1}[Z = z_c]$	$\mathbf{1}[Z = z_8]$	$E(Y(t_m)\|\boldsymbol{S} \in \{\boldsymbol{s}_{fc}, \boldsymbol{s}_{ph}\})$
		$\mathbf{1}[Z = z_c]$	$\mathbf{1}[Z = z_e]$	$E(Y(t_m)\|\boldsymbol{S} = \boldsymbol{s}_{pm})$
$D_{t_l} \equiv \mathbf{1}[T = t_l]$	$D_{t_l} \cdot Y$	$\mathbf{1}[Z = z_c]$	$\mathbf{1}[Z = z_8]$	$E(Y(t_l)\|\boldsymbol{S} = \boldsymbol{s}_{pl})$
		$\mathbf{1}[Z = z_8]$	$\mathbf{1}[Z = z_e]$	$E(Y(t_l)\|\boldsymbol{S} \in \{\boldsymbol{s}_{fc}, \boldsymbol{s}_{pm}\})$

This table lists the counterfactual outcome means estimated by 2SLS procedures. The first stage estimates use two IV indicators (columns 3 and 4) that are multiplied by γ_1, γ_2 in (148). The choice indicator (column 1) is the endogenous variable estimated in the first stage (148). The second stage uses the interaction of the outcome and the choice indicator (column 2) as dependent variable and uses the estimate of the first stage, which is multiplied by the linear coefficient β_{IV} in . The last column lists the identified counterfactual outcome mean.

We can control for pre-program variables $\boldsymbol{X}$ in a linear fashion by including these variables as covariates in the 2SLS regressions. Angrist and Imbens (1995) show that the 2SLS estimate is a weighted average of the counterfactual outcomes conditioned on the covariates. Weights consist of the variance of the choice indicators conditioned on the covariates.

Abadie (2003) proposes a κ-weighting scheme that nonparametrically controls for baseline variables in the LATE model.[71] The triangular property in (35) and (42) enables us to extend Abadie's κ to the case of multiple choices.

Counterfactual outcome $E\left(Y(t_l)|\boldsymbol{S}=\boldsymbol{s}_{pl}\right)$ is identified as a ratio of two *matching estimators* that depend on z_8 and z_c. This counterfactual outcome can also be expressed in (150) as the expectation of the observed outcome Y multiplied by a weighting function $\kappa(t_l, \boldsymbol{s}_{pl})$ in (151) which depends on z_8 and z_c.

$$E\left(Y(t_l)|\boldsymbol{S}=\boldsymbol{s}_{pl}\right) = E\left(Y \cdot \frac{\kappa(t_l, \boldsymbol{s}_{pl})}{E(\kappa(t_l, \boldsymbol{s}_{pl}))}\right), \tag{150}$$

$$\text{such that } \kappa(t_l, \boldsymbol{s}_{pl}) = D_{t_l}\left(\frac{\mathbf{1}[Z=z_8]}{P(Z=z_8|\boldsymbol{X})} - \frac{\mathbf{1}[Z=z_c]}{P(Z=z_c|\boldsymbol{X})}\right). \tag{151}$$

Equation (150) also holds if Y were to be replaced by any measurable function $g(Y, D_{t_l}, \boldsymbol{X})$. The κ-weighting in (151) can be evaluated from data. It consists of the choice indicator D_{t_l} multiplied by the difference between IV indicators of z_8 and z_c divided by their respective probabilities conditional on baseline variables $\boldsymbol{X}$. $E\left(Y(t_l)|\boldsymbol{S}=\boldsymbol{s}_{pl}\right)$ can be estimated by the sample counterpart of (150), that is, $\sum_i Y_i \cdot \omega_i$, where $\omega_i = \kappa_i(t_l, \boldsymbol{s}_{pl})/(\sum_i \kappa_i(t_l, \boldsymbol{s}_{pl}))$ are weights that sum to one and $\kappa_i(t_l, \boldsymbol{s}_{pl})$ is the κ-weight of family i.

The practical use of the κ-weights is to evaluate causal parameters via conventional estimation procedures that reweighed data according to the estimated values of κ. An example of an estimation procedure for $E\left(Y(t_l)|\boldsymbol{S}=\boldsymbol{s}_{pl}\right)$ is: (1) estimate $P(Z=z_8|\boldsymbol{X}), P(Z=z_c|\boldsymbol{X})$; (2) construct weights $\hat{\kappa}(t_l, \boldsymbol{s}_{pl})$ as in (151); (3) estimate β_1 in regression $Y \cdot D_{t_l} = \beta_0 + \beta_1 \hat{D}_{t_l} + \beta_2 \boldsymbol{X} + \epsilon_Y$ via weighted least squares (WLS) that employ weights $\hat{\kappa}(t_l, \boldsymbol{s}_{pl})$. The WLS solves the sample analog of $(\beta_0, \beta_1, \beta_2) =$

[71] Abadie (2003) shows that the counterfactual outcomes of the LATE model can be evaluated by a weighted average of the outcome across the whole population. He names the weighting functions κ.

$\arg\min_{b_0,b_1,b_2} E(\kappa \cdot g(Y, D, \boldsymbol{X}))$, where $g(Y, D, \boldsymbol{X}) = (YD_{t_l} - (b_0 + b_1\hat{D}_{t_l} + b_2\boldsymbol{X}))^2$.

Weights κ for counterfactual outcomes in Table A.8 can be obtained by replacing D_{t_l}, z_8, z_c in (151) by their corresponding neighborhood choice and IV indicators.

G Identification and Estimation of Counterfactual Outcomes

This section provides further information on the solution to the problem of partial identification in MTO. Section G.1 estimates bounds for the partially identified effects. Section G.2 presents the identification strategy for all partially identified counterfactual outcomes in the same fashion that Section 5.3 describes the identification of partially identified outcomes for low poverty neighborhood. Section G.3 discusses the propensity score estimator in greater detail.

G.1 Estimating Bounds for Partially Identified Counterfactuals

The primary goal of the MTO evaluation is to assess the neighborhood effects for the full-compliers $\boldsymbol{s}_{fc}$. This response type consists of families that are most responsive to vouchers incentives. Full compliers also comprise the largest set of families among the compliers. We are particularly interested in evaluating the neighborhood effect that compares low- and high-poverty neighborhoods, that is, $E(Y(t_l) - Y(t_h)|\boldsymbol{S} = \boldsymbol{s}_{fc})$.

Unfortunately, according to Theorem **T.1**, the counterfactual outcomes of the full-compliers are not point-identified. Instead, each of the counterfactuals for the full-compliers are partially identified: $E(Y(t_h)|\boldsymbol{S} \in \{\boldsymbol{s}_{fc}, \boldsymbol{s}_{pl}\})$, $E(Y(t_m)|\boldsymbol{S} \in \{\boldsymbol{s}_{fc}, \boldsymbol{s}_{ph}\})$, and $E(Y(t_l)|\boldsymbol{S} \in \{\boldsymbol{s}_{fc}, \boldsymbol{s}_{pm}\})$.

One way to evaluate neighborhood effects is to look for additional assumptions that would allow us to identify these counterfactual outcomes. This is done in Section 5 of the main paper. Another approach, which is pursued here, is to evaluate bounds for the counterfactual outcomes of full-compliers. Let's take the counterfactual for low-poverty neighborhood $E(Y(t_l)|\boldsymbol{S} \in \{\boldsymbol{s}_{fc}, \boldsymbol{s}_{pm}\})$ as our main example for bound analysis. The identification equation for this counterfactual is:

$$E(Y(t_l)|\boldsymbol{S} \in \{\boldsymbol{s}_{fc}, \boldsymbol{s}_{pm}\}) = \frac{E(YD_{t_l}|Z = z_e) - E(YD_{t_l}|Z = z_8)}{P_{t_l}(z_e) - P_{t_l}(z_8)} \tag{152}$$

$$\tag{153}$$

It is useful to rewrite the numerator in the right-hand side of (152) in the following fashion:

$$\begin{aligned} E(YD_{t_l}|Z = z_e) - E(YD_{t_l}|Z = z_8) &= \frac{E(YD_{t_l} \cdot \mathbf{1}[Z = z_e])}{P(Z = z_e)} - \frac{E(YD_{t_l} \cdot \mathbf{1}[Z = z_8])}{P(Z = z_8)} \\ &= \frac{E(Y \cdot \mathbf{1}[Z = z_e]|T = t_l)P(T = t_l)}{P(Z = z_e)} - \frac{E(Y \cdot \mathbf{1}[Z = z_8]|T = t_l)P(T = t_l)}{P(Z = z_8)} \\ &= E\left(Y \cdot \left(\frac{\mathbf{1}[Z = z_e]}{P(Z = z_e)} - \frac{\mathbf{1}[Z = z_8]}{P(Z = z_8)}\right)\middle| T = t_l\right) P(T = t_l) \end{aligned}$$

We can then rewrite the identification formula (152) as following:

$$E(Y(t_l)|\boldsymbol{S} \in \{\boldsymbol{s}_{fc}, \boldsymbol{s}_{pm}\}) = E\left(Y \cdot \kappa_l | T = t_l\right) \tag{154}$$

$$\text{where } \kappa_l = \left(\frac{\mathbf{1}[Z = z_e]}{P(Z = z_e)} - \frac{\mathbf{1}[Z = z_8]}{P(Z = z_8)}\right) \cdot \frac{P(T = t_l)}{P_{t_l}(z_e) - P_{t_l}(z_8)}. \tag{155}$$

The expression above is closely related to the κ-weighting scheme of Abadie (2003). The cumulative distribution of the counterfactual $Y(t_l)$ conditioned on $\boldsymbol{S} \in \{\boldsymbol{s}_{fc}, \boldsymbol{s}_{pm}\}$ is identified by $P(Y(t_l) \leq$

$y|\boldsymbol{S} \in \{\boldsymbol{s}_{fc}, \boldsymbol{s}_{pm}\}) = E(\mathbf{1}[Y \leq y] \cdot \kappa_l)$. The representation in (154)–(155) is useful to show that the expectation and the distribution of $Y(t_l)$ conditioned on $\boldsymbol{S} \in \{\boldsymbol{s}_{fc}, \boldsymbol{s}_{pm}\}$ are identified and they only depend on the distribution of Y and Z conditioned on $T = t_l$. In the same token, the identification of the partially identified counterfactuals for the neighborhood choices t_h and t_m depend only on the distribution of Y, Z conditioned on t_h and t_m respectively. This is relevant because we can examine the bounds corresponding to each neighborhood choice separately.

It is useful to express the mean of the partially identified counterfactual outcome as the following mixture:

$$E(Y(t_l)|\boldsymbol{S} \in \{\boldsymbol{s}_{fc}, \boldsymbol{s}_{pm}\}) = E(Y(t_l)|\boldsymbol{S} = \boldsymbol{s}_{fc}) \cdot \omega_l + E(Y(t_l)|\boldsymbol{S} = \boldsymbol{s}_{pm}) \cdot (1 - \omega_l) \tag{156}$$

$$\text{where } \omega_l = \frac{P(\boldsymbol{S} = \boldsymbol{s}_{fc})}{P(\boldsymbol{S} = \boldsymbol{s}_{fc}) + P(\boldsymbol{S} = \boldsymbol{s}_{pm})}. \tag{157}$$

The weight ω_l is known since the response-type probabilities are identified. Thus, the problem of assessing bounds for $E(Y(t_l)|\boldsymbol{S} = \boldsymbol{s}_{fc})$ can be understood as the standard problem of evaluating bounds for the mean of an unobserved potential outcome that is a component of a known mixing distribution $P(Y(t_l) \leq y|\boldsymbol{S} \in \{\boldsymbol{s}_{fc}, \boldsymbol{s}_{pm}\})$ with a known mixing probability ω_l. This problem is investigated by Horowitz and Manski (1995), who presents the sharp bounds for the mean potential outcome in Proposition 4 of Section 3.2. The bounds for a real-valued random variable Y are given by:

$$E(Y(t_l)|\boldsymbol{S} = \boldsymbol{s}_{fc}) \in \left[\frac{1}{\omega_l} \int_{-\infty}^{q_l(\omega_l)} y \, dQ_l(y) \, , \, \frac{1}{\omega_l} \int_{q_l(1-\omega_l)}^{\infty} y \, dQ_l(y) \right], \tag{158}$$

where $Q_l(y) = P(Y(t_l) \leq y|\boldsymbol{S} \in \{\boldsymbol{s}_{fc}, \boldsymbol{s}_{pm}\})$ is the cumulative distribution function of $Y(t_l)$ conditioned on $\boldsymbol{S} \in \{\boldsymbol{s}_{fc}, \boldsymbol{s}_{pm}\}$, and $q_l(\omega_l) = \inf_y\{Q_l(y) \geq \omega_l\}$ is the ω_l-quantile of $Q_l(y)$. The bounds for $E(Y(t_l)|\boldsymbol{S} = \boldsymbol{s}_{fc})$ in (158) are given by trimmed means $[E(Y(t_l)|\boldsymbol{S} \in \{\boldsymbol{s}_{fc}, \boldsymbol{s}_{pm}\}, Y(t_l) \leq q_l(\omega_l)), E(Y(t_l)|\boldsymbol{S} \in \{\boldsymbol{s}_{fc}, \boldsymbol{s}_{pm}\}, Y(t_l) > q_l(1 - \omega_l))]$.

The bounds for $E(Y(t_l)|\boldsymbol{S} = \boldsymbol{s}_{fc})$ are estimated as the empirical counterpart of equation (158). This estimation requires the evaluation of the empirical cumulative distribution function of $Y(t_l)$ conditioned on $\boldsymbol{S} \in \{\boldsymbol{s}_{fc}, \boldsymbol{s}_{pm}\}$, that is, $Q_l(y) = E(\mathbf{1}[Y(t_l) \leq y]|\boldsymbol{S} \in \{\boldsymbol{s}_{fc}, \boldsymbol{s}_{pm}\})$ for all $y \in supp(Y)$. Note that Sections G.4–G.5 describe how to estimate the counterfactual mean $E(Y(t_l)|\boldsymbol{S} = \boldsymbol{s}_{fc})$. The cumulative distribution $Q_l(y)$ is estimated using the same method when replacing the outcome Y by the indicator $\mathbf{1}[Y \leq y]$.

Table A.9 provides estimated bounds for all the counterfactual outcomes and neighborhood effects for the full-compliers $\boldsymbol{s}_{fc}$. The bounds for the counterfactual outcomes are informative, but the bounds for the neighborhood effects are not. This is not entirely unexpected, as these results are consistent with a sizeable IV literature which often finds that bounds for treatment effects are commonly wide and rarely informative (Brinch et al., 2017; Heckman and Vytlacil, 2007).

G.2 Identification Strategy for Partially Identified Counterfactuals

Section 5.3 of the main paper describes the problem of partial identification for the low-poverty neighborhood t_l, where we seek to decompose the identified counterfactual mean $E(Y(t_l)|\boldsymbol{S} \in \{\boldsymbol{s}_{fc}, \boldsymbol{s}_{pm}\})$ into $E(Y(t_l)|\boldsymbol{S} = \boldsymbol{s}_{fc})$ and $E(Y(t_l)|\boldsymbol{S} = \boldsymbol{s}_{pm})$.

The partially identified counterfactual $E(Y(t_l)|\boldsymbol{S} \in \{\boldsymbol{s}_{fc}, \boldsymbol{s}_{pm}\})$ is identified by the LATE parameter that compares the IV values z_8, z_e. According to Theorem **T.4**, $E(Y(t_l)|\boldsymbol{S} \in \{\boldsymbol{s}_{fc}, \boldsymbol{s}_{pm}\})$ can be expressed as the integral of the response function $E(Y(t_l)|U_{t_l} = u)$ over the propensity score interval $[P_{t_l}(z_8), P_{t_l}(z_e)]$. Note that the length of the interval is equal to the response type

Table A.9: Estimated Bounds for Counterfactual Outcomes and Treatment Effects of Full-compliers

	Counterfactual Outcomes full-compliers $\boldsymbol{s}_{fc}$			**Neighborhood Effects** full-compliers $\boldsymbol{s}_{fc}$		
Choices	$E(Y(t_h)\|\boldsymbol{s}_{fc})$	$E(Y(t_m)\|\boldsymbol{s}_{fc})$	$E(Y(t_l)\|\boldsymbol{s}_{fc})$	$E(Y(t_l)-Y(t_h)\|\boldsymbol{s}_{fc})$	$E(Y(t_l)-Y(t_m)\|\boldsymbol{s}_{fc})$	$E(Y(t_m)-Y(t_h)\|\boldsymbol{s}_{fc})$
Income of Family Head	9.868	11.203	11.924	2.056	0.721	1.334
Bounds	[8.269, 11.464]	[6.291, 14.905]	[8.470, 14.898]	[−2.993, 6.629]	[−6.435, 8.607]	[−5.173, 6.636]
Income of Head and Spouse	11.533	11.062	12.411	0.878	1.349	−0.471
Bounds	[9.630, 13.822]	[6.639, 15.841]	[9.086, 15.104]	[−4.736, 5.474]	[−6.755, 8.464]	[−7.183, 6.211]
Total household income	12.844	12.296	14.747	1.902	2.451	−0.549
Bounds	[10.884, 15.400]	[8.047, 17.701]	[11.335, 17.208]	[−4.065, 6.324]	[−6.366, 9.161]	[−7.353, 6.817]
Above Poverty Line	0.239	0.303	0.347	0.108	0.044	0.064
Bounds	[0.145, 0.281]	[0.000, 0.373]	[0.171, 0.420]	[−0.109, 0.276]	[−0.202, 0.420]	[−0.281, 0.228]
Employed without welfare	0.414	0.392	0.527	0.113	0.135	−0.022
Bounds	[0.357, 0.482]	[0.175, 0.636]	[0.370, 0.597]	[−0.113, 0.240]	[−0.266, 0.422]	[−0.307, 0.279]
Currently on welfare	0.351	0.256	0.229	−0.121	−0.026	−0.095
Bounds	[0.274, 0.409]	[0.000, 0.425]	[0.117, 0.349]	[−0.292, 0.075]	[−0.308, 0.349]	[−0.409, 0.151]
Job tenure	0.343	0.324	0.431	0.088	0.107	−0.019
Bounds	[0.249, 0.382]	[0.086, 0.589]	[0.269, 0.506]	[−0.113, 0.256]	[−0.320, 0.420]	[−0.296, 0.340]
Economic self-sufficiency	0.155	0.235	0.220	0.065	−0.015	0.080
Bounds	[0.060, 0.194]	[0.000, 0.309]	[0.022, 0.258]	[−0.172, 0.198]	[−0.286, 0.258]	[−0.194, 0.249]
Neighborhood Poverty	39.948	27.078	6.692	−33.256	−20.387	−12.869
Bounds	[35.716, 43.894]	[23.479, 36.152]	[6.373, 8.701]	[−37.521, −27.015]	[−29.779, −14.779]	[−20.415, 0.436]

This table presents the bounds for mean outcomes (columns 3–5) and neighborhood effects (columns 6–8) of economic outcomes for the full-compliers $\boldsymbol{s}_{fc}$. The fist row of each outcome displays the point-estimates are also shown in Tables 6–7 of the main paper. The second row displays the estimated bounds corresponding to each parameter. All estimates use the adult sampling weights of MTO interim evaluation and are conditioned on the site of intervention and the baseline variables discussed in Section 2. Appendix G.5 describes the estimation procedure in detail.

probability $(\boldsymbol{S} \in \{\boldsymbol{s}_{fc}, \boldsymbol{s}_{pm}\}) = P_{t_l}(z_e) - P_{t_l}(z_8)$.

The strategy to decompose $E(Y(t_l)|\boldsymbol{S} \in \{\boldsymbol{s}_{fc}, \boldsymbol{s}_{pm}\})$ consists of two steps. The first one is to split the interval $[P_{t_l}(z_8), P_{t_l}(z_e)]$ be split into $[P_{t_l}(z_8), P_{t_l}(z_8) + P(\boldsymbol{S} = \boldsymbol{s}_{fc})]$, corresponding to $\boldsymbol{s}_{fc}$, and $[P_{t_l}(z_8) + P(\boldsymbol{S} = \boldsymbol{s}_{fc}), P_{t_l}(z_e)]$, corresponding $\boldsymbol{s}_{pm}$. The equations associated with this decomposition are listed below:

$$E(Y(t_l)|\boldsymbol{S} \in \{\boldsymbol{s}_{fc}, \boldsymbol{s}_{pm}\}) = \frac{\int_{P_{t_l}(z_8)}^{P_{t_l}(z_e)} E(Y(t_l)|U_{t_l} = u)du}{P(\boldsymbol{S} = \boldsymbol{s}_{fc})} = \frac{E(YD_{t_l}|P_{t_l} = P_{t_l}(z_e)) - E(YD_t|P_{t_l} = P_{t_l}(z_8))}{P_{t_l}(z_e) - P_{t_l}(z_8)},$$

$$E(Y(t_l)|\boldsymbol{S} = \boldsymbol{s}_{fc}) = \frac{\int_{P_{t_l}(z_8)}^{p^*} E(Y(t_l)|U_{t_l} = u)du}{P(\boldsymbol{S} = \boldsymbol{s}_{fc})} = \frac{E(YD_{t_l}|P_{t_l} = p^*) - E(YD_t|P_{t_l} = P_{t_l}(z_8))}{p^* - P_{t_l}(z_8)},$$

$$E(Y(t_l)|\boldsymbol{S} = \boldsymbol{s}_{pm}) = \frac{\int_{p^*}^{P_{t_l}(z_e)} E(Y(t_l)|U_{t_l} = u)du}{P(\boldsymbol{S} = \boldsymbol{s}_{pm})} = \frac{E(YD_{t_l}|P_{t_l} = P_{t_l}(z_e)) - E(YD_t|P_{t_l} = p^*)}{P_{t_l}(z_e) - p^*},$$

where $P_{t_l} \equiv P(T = t_l|Z)$ is the propensity score and $p^* = P_{t_l}(z_8) + P(\boldsymbol{S} = \boldsymbol{s}_{fc})$.

The second step consists of using a parametric function to evaluate the LATE-type equations above using a propensity score estimator.

A symmetric argument applies to the decomposition of the partially identified counterfactuals for choices t_h and t_m. In the case of t_m, we seek to decompose $E(Y(t_m)|\boldsymbol{S} \in \{\boldsymbol{s}_{fc}, \boldsymbol{s}_{ph}\})$ into $E(Y(t_m)|\boldsymbol{S} = \boldsymbol{s}_{fc})$ and $E(Y(t_m)|\boldsymbol{S} = \boldsymbol{s}_{ph})$. The partially identified counterfactual $E(Y(t_m)|\boldsymbol{S} \in \{\boldsymbol{s}_{fc}, \boldsymbol{s}_{ph}\})$ is identified by the LATE parameter that compares the IV values z_c, z_8. The identification strategy splits propensity score interval $[P_{t_m}(z_c), P_{t_l}(z_8)]$ associated with the response types $\{\boldsymbol{s}_{fc}, \boldsymbol{s}_{ph}\}$ into $[P_{t_m}(z_c), P_{t_l}(z_c) + P(\boldsymbol{S} = \boldsymbol{s}_{fc})]$, corresponding to $\boldsymbol{s}_{fc}$, and $[P_{t_m}(z_c) + P(\boldsymbol{S} = \boldsymbol{s}_{fc}), P_{t_m}(z_8)]$, corresponding $\boldsymbol{s}_{ph}$. The equations associated with this decomposition are listed below:

$$E(Y(t_m)|\boldsymbol{S} \in \{\boldsymbol{s}_{fc}, \boldsymbol{s}_{ph}\}) = \frac{\int_{P_{t_m}(z_c)}^{P_{t_m}(z_8)} E(Y(t_m)|U_{t_m} = u)du}{P(\boldsymbol{S} = \boldsymbol{s}_{fc})} = \frac{E(YD_{t_m}|P_{t_m} = P_{t_m}(z_8)) - E(YD_t|P_{t_m} = P_{t_m}(z_c))}{P_{t_m}(z_8) - P_{t_m}(z_c)},$$

$$E(Y(t_m)|\boldsymbol{S} = \boldsymbol{s}_{fc}) = \frac{\int_{P_{t_m}(z_c)}^{p^*} E(Y(t_m)|U_{t_m} = u)du}{P(\boldsymbol{S} = \boldsymbol{s}_{fc})} = \frac{E(YD_{t_m}|P_{t_m} = p^*) - E(YD_t|P_{t_m} = P_{t_m}(z_c))}{p^* - P_{t_m}(z_c)},$$

$$E(Y(t_m)|\boldsymbol{S} = \boldsymbol{s}_{ph}) = \frac{\int_{p^*}^{P_{t_m}(z_8)} E(Y(t_m)|U_{t_m} = u)du}{P(\boldsymbol{S} = \boldsymbol{s}_{ph})} = \frac{E(YD_{t_m}|P_{t_m} = P_{t_m}(z_8)) - E(YD_t|P_{t_m} = p^*)}{P_{t_m}(z_8) - p^*},$$

where $P_{t_m} \equiv P(T = t_m|Z)$ is the propensity score and $p^* = P_{t_m}(z_c) + P(\boldsymbol{S} = \boldsymbol{s}_{fc})$.

In the case of t_h, we seek to decompose $E(Y(t_h)|\boldsymbol{S} \in \{\boldsymbol{s}_{fc}, \boldsymbol{s}_{pl}\})$ into $E(Y(t_h)|\boldsymbol{S} = \boldsymbol{s}_{fc})$ and $E(Y(t_h)|\boldsymbol{S} = \boldsymbol{s}_{pl})$. The partially identified counterfactual $E(Y(t_h)|\boldsymbol{S} \in \{\boldsymbol{s}_{fc}, \boldsymbol{s}_{pl}\})$ is identified by the LATE parameter that compares the IV values z_e, z_c. The identification strategy splits propensity score interval $[P_{t_h}(z_e), P_{t_l}(z_c)]$ associated with the response types $\{\boldsymbol{s}_{fc}, \boldsymbol{s}_{pl}\}$ into $[P_{t_h}(z_e), P_{t_l}(z_e) + P(\boldsymbol{S} = \boldsymbol{s}_{fc})]$, corresponding to $\boldsymbol{s}_{fc}$, and $[P_{t_h}(z_e) + P(\boldsymbol{S} = \boldsymbol{s}_{fc}), P_{t_h}(z_c)]$, corresponding $\boldsymbol{s}_{pl}$. The equations associated with this decomposition are listed below:

$$E(Y(t_h)|\boldsymbol{S} \in \{\boldsymbol{s}_{fc}, \boldsymbol{s}_{pl}\}) = \frac{\int_{P_{t_h}(z_e)}^{P_{t_h}(z_c)} E(Y(t_h)|U_{t_h} = u)du}{P(\boldsymbol{S} = \boldsymbol{s}_{fc})} = \frac{E(YD_{t_h}|P_{t_h} = P_{t_h}(z_c)) - E(YD_t|P_{t_h} = P_{t_h}(z_e))}{P_{t_h}(z_c) - P_{t_h}(z_e)},$$

$$E(Y(t_h)|\boldsymbol{S} = \boldsymbol{s}_{fc}) = \frac{\int_{P_{t_h}(z_e)}^{p^*} E(Y(t_h)|U_{t_h} = u)du}{P(\boldsymbol{S} = \boldsymbol{s}_{fc})} = \frac{E(YD_{t_h}|P_{t_h} = p^*) - E(YD_t|P_{t_h} = P_{t_h}(z_e))}{p^* - P_{t_h}(z_e)},$$

$$E(Y(t_h)|\boldsymbol{S} = \boldsymbol{s}_{pl}) = \frac{\int_{p^*}^{P_{t_h}(z_c)} E(Y(t_h)|U_{t_h} = u)du}{P(\boldsymbol{S} = \boldsymbol{s}_{pl})} = \frac{E(YD_{t_h}|P_{t_h} = P_{t_h}(z_c)) - E(YD_t|P_{t_h} = p^*)}{P_{t_h}(z_c) - p^*},$$

where $P_{t_h} \equiv P(T = t_h|Z)$ is the propensity score and $p^* = P_{t_h}(z_e) + P(\boldsymbol{S} = \boldsymbol{s}_{fc})$.

Note that all three identification strategies split the propensity score intervals such that the first interval corresponds to the full-compliers $\boldsymbol{s}_{fc}$ while the second interval is ascribed to a partial-complier. Specifically, the first interval $[P_{t_l}(z_8), P_{t_l}(z_8) + P(\boldsymbol{S} = \boldsymbol{s}_{fc})]$ of $[P_{t_l}(z_8), P_{t_l}(z_e)]$ for t_l corresponds to full-complier $\boldsymbol{s}_{fc}$, while the second interval $[P_{t_l}(z_8) + P(\boldsymbol{S} = \boldsymbol{s}_{fc}), P_{t_l}(z_e)]$ corresponding the partial-complier $\boldsymbol{s}_{pm}$. This ordering is maintained for choices t_m and t_h as well. In the case of t_m, the first interval corresponds to full-complier $\boldsymbol{s}_{fc}$ while the second to $\boldsymbol{s}_{ph}$. In the case of t_h, the first interval corresponds to full-complier $\boldsymbol{s}_{fc}$ while the second to $\boldsymbol{s}_{pl}$.

It is natural to question if it is possible to assign the second interval to the full-compliers instead of the first interval. It turns out that this is not a viable strategy. Theorem **T.5** explains that assigning the second interval to the full-compliers (instead of the first interval) is only possible if the response-type probabilities of at least two partial-compliers is zero, which is against the empirical evidence in MTO.

Theorem T.5. Consider the IV model characterized by assumptions (1)–(3) in which the response matrix (29) holds, and each choice indicator is given by equation (46), that is, $D_t = \mathbf{1}[P_t(Z) \geq U_t]; U_t \sim Unif[0,1]$ for $t \in \{t_h, t_m, t_l\}$. Suppose that the intervals corresponding to the partial-compliers $(\boldsymbol{s}_{pm}, \boldsymbol{s}_{ph}, \boldsymbol{s}_{pl})$ in the support of $(U_{t_l}, U_{t_m}, U_{t_h})$ preceded the intervals corresponding to the full-complier $\boldsymbol{s}_{fc}$. Then it must be the case that the response type probability of at least two of these partial-compliers is zero. On the other hand, there are no probability constraints if the intervals corresponding to the full-compliers $\boldsymbol{s}_{fc}$ precede the intervals corresponding to the partial-compliers $(\boldsymbol{s}_{pm}, \boldsymbol{s}_{ph}, \boldsymbol{s}_{pl})$ in $(U_{t_l}, U_{t_m}, U_{t_h})$.

Proof. See Appendix G.2.1 for proof. □

G.2.1 Proof of Theorem T.5

Assumptions (1)–(3) enable us to relate propensity scores and response type probabilities by the following equation:

$$P_t(z) \equiv P(T = t|Z = z) = \sum_{\boldsymbol{s} \in supp(\boldsymbol{S})} \mathbf{1}[T = t|\boldsymbol{S} = \boldsymbol{s}, Z = z]\, P(\boldsymbol{S} = \boldsymbol{s}). \tag{159}$$

The equation above shows that each propensity score equals a sum of response type probabilities. The triangular property of the MTO response matrix in (35), (42), and (43) enable us to map each propensity score to nested sets of response types. In the case of t_l, we can use equation (159) and

$\boldsymbol{R}_l$ in (35) to express the propensity scores as:

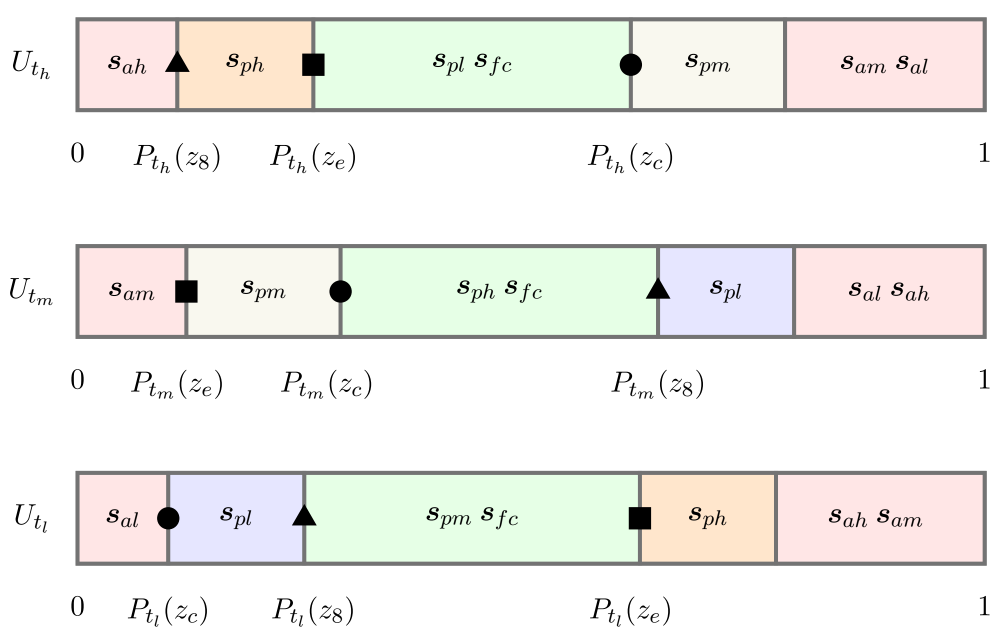
$$P_{t_l}(z_c) = \sum_{\boldsymbol{s} \in \{\boldsymbol{s}_{al}\}} P(\boldsymbol{S} = \boldsymbol{s}) \Rightarrow P_{t_l}(z_c) = P(\boldsymbol{S} \in \{\boldsymbol{s}_{al}\})$$
$$P_{t_l}(z_8) = \sum_{\boldsymbol{s} \in \{\boldsymbol{s}_{al}, \boldsymbol{s}_{pl}\}} P(\boldsymbol{S} = \boldsymbol{s}) \Rightarrow P_{t_l}(z_8) = P(\boldsymbol{S} \in \{\boldsymbol{s}_{al}, \boldsymbol{s}_{pl}\})$$
$$P_{t_l}(z_e) = \sum_{\boldsymbol{s} \in \{\boldsymbol{s}_{al}, \boldsymbol{s}_{pl}, \boldsymbol{s}_{fc}, \boldsymbol{s}_{pm}\}} P(\boldsymbol{S} = \boldsymbol{s}) \Rightarrow P_{t_l}(z_e) = P(\boldsymbol{S} \in \{\boldsymbol{s}_{al}, \boldsymbol{s}_{pl}, \boldsymbol{s}_{fc}, \boldsymbol{s}_{pm}\})$$

Note that the triangular property of the MTO response matrix maps the propensity scores into a family of nested sets of response types. This sequence of nested sets determines the ordering of the response types along the support of the variable U_t of the choice indicator $D_t = \mathbf{1}[P_t(Z) \geq U_t]$. Figure 2 displays the ordering of response types for t_l. In summary, the sequence of the response types associated with each variable $U_t; t \in \{t_h, t_m, , t_l\}$ is given below:

- The sequence of response types associated with U_{t_h} is: $(\boldsymbol{s}_{ah}, \boldsymbol{s}_{ph}, \{\boldsymbol{s}_{fc}, \boldsymbol{s}_{pl}\}, \{\boldsymbol{s}_{pm}, \boldsymbol{s}_{am}, \boldsymbol{s}_{al}\})$
- The sequence of response types associated with U_{t_m} is: $(\boldsymbol{s}_{am}, \boldsymbol{s}_{pm}, \{\boldsymbol{s}_{fc}, \boldsymbol{s}_{ph}\}, \{\boldsymbol{s}_{pl}, \boldsymbol{s}_{al}, \boldsymbol{s}_{ah}\})$
- The sequence of response types associated with U_{t_l} is: $(\boldsymbol{s}_{al}, \boldsymbol{s}_{pl}, \{\boldsymbol{s}_{fc}, \boldsymbol{s}_{pm}\}, \{\boldsymbol{s}_{ph}, \boldsymbol{s}_{ah}, \boldsymbol{s}_{am}\})$

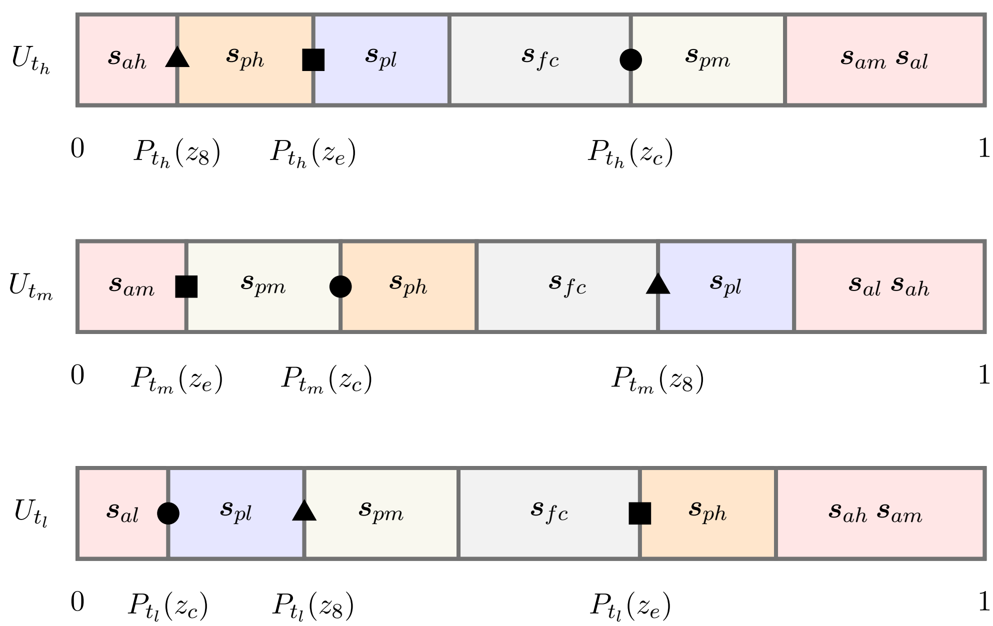
Figure A.5 presents a diagram that displays the ordering of the response types associated with each variable U_t.

Figure A.5: Order of Response types for Each Neighborhood Choice due to MTO Response Matrix



We seek to split three intervals: (1) $\{\boldsymbol{s}_{fc}, \boldsymbol{s}_{pl}\}$ associated with U_{t_h}; (2) $\{\boldsymbol{s}_{fc}, \boldsymbol{s}_{ph}\}$ associated with U_{t_m}; (3) $\{\boldsymbol{s}_{fc}, \boldsymbol{s}_{pm}\}$ associated with U_{t_l}. The theorem considers the ordering of the response types in Figure A.5. It investigates the case where the partial compliers $\boldsymbol{s}_{pl}, \boldsymbol{s}_{ph}, \boldsymbol{s}_{pm}$ precede the full-compliers $\boldsymbol{s}_{fc}$. This ordering scheme is displayed in the diagram of Figure A.6.

Figure A.6: Order of Response types Assuming that Partial-compliers precede Full-compliers



It is useful to clarify the choice scheme displayed by the diagram of Figure A.6.

Consider the top bar which refers to variable U_{t_h}. The sequence of response types displayed in the bar is $\boldsymbol{s}_{ah}, \boldsymbol{s}_{ph}, \boldsymbol{s}_{pl}, \boldsymbol{s}_{fc}, \boldsymbol{s}_{pm}, \{\boldsymbol{s}_{am}, \boldsymbol{s}_{al}\}$. According to the choice equation $D_{t_h} = \mathbf{1}[p_{t_h} \geq U_{t_h}]$, this sequence of response types implies a specific choice scheme. If the propensity score of choice t_h is set to $p_{t_h} = P(\boldsymbol{S} = \boldsymbol{s}_{al})$, then, $\boldsymbol{s}_{ah}$-families choose t_h while the other family types will not. If the propensity score of choice t_h is set to $p_{t_h} = P(\boldsymbol{S} \in \{\boldsymbol{s}_{al}, \boldsymbol{s}_{ph}\})$, then, families of type $\boldsymbol{s}_{ah}$ and $\boldsymbol{s}_{ph}$ choose t_h while the other family types will not. This pattern continues according to the sequence of the response types displayed for t_h. The symmetric choice procedure also holds for the remaining choices.

We seek to investigate the choice scheme generated by assuming that partial-compliers precede the full-compliers. In the case of t_h, it means that $\boldsymbol{s}_{pl}$ precedes $\boldsymbol{s}_{fc}$. Thus considering setting the propensity score of choice t_h to $p_{t_h} = P(\boldsymbol{S} \in \{\boldsymbol{s}_{ah}, \boldsymbol{s}_{ph}, \boldsymbol{s}_{pl}\})$. This means that families of type $\boldsymbol{s}_{ah}, \boldsymbol{s}_{ph}$ and $\boldsymbol{s}_{pl}$ choose t_h, while the families associated with the remaining response types, that is, $\boldsymbol{s}_{fc}, \boldsymbol{s}_{pm}, \boldsymbol{s}_{am}$ and $\boldsymbol{s}_{al}$ do not choose t_h. In particular, the full-compliers $\boldsymbol{s}_{fc}$ must choose t_m or t_l. These two possibilities are considered below:

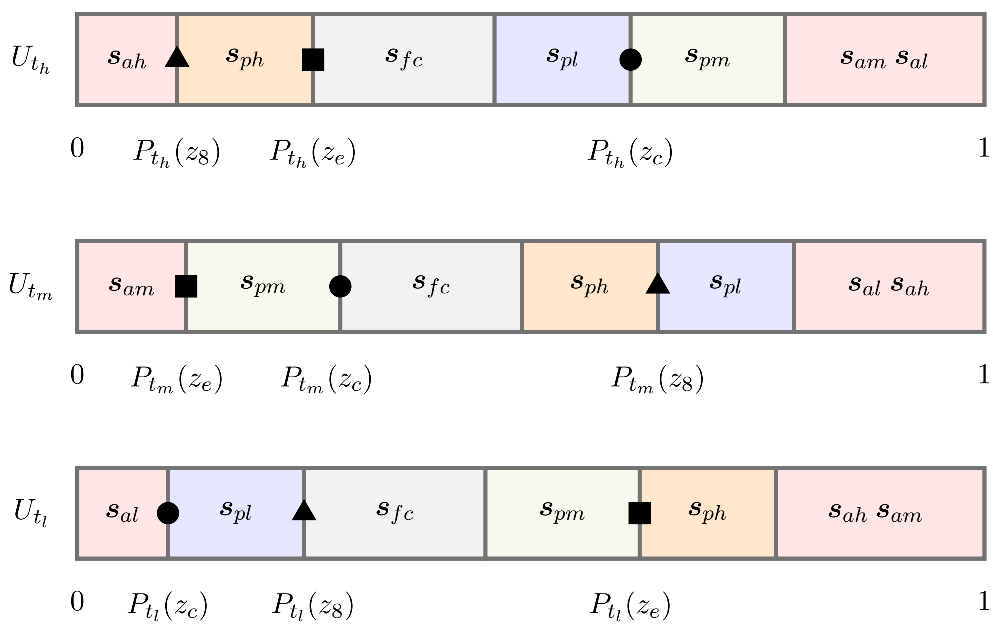
- Suppose that full-compliers $\boldsymbol{s}_{fc}$ choose t_m. Thus, according to the sequence of response types of U_m (second bar), it must be the case that $p_m \geq P(\boldsymbol{S} \in \{\boldsymbol{s}_{am}, \boldsymbol{s}_{pm}, \boldsymbol{s}_{ph}, \boldsymbol{s}_{fc}\})$. Note that, according to equation $D_{t_h} = \mathbf{1}[p_{t_h} \geq U_{t_h}]$, it implies that families of type $\boldsymbol{s}_{am}, \boldsymbol{s}_{pm}$ and $\boldsymbol{s}_{ph}$ also choose t_m. There lies a contradiction since we have established that families of type $\boldsymbol{s}_{ph}$ were already choosing t_h.

- Suppose that full-compliers $\boldsymbol{s}_{fc}$ choose t_l. Thus, according to the sequence of response types of U_l (third bar), it must be the case that $p_l \geq P(\boldsymbol{S} \in \{\boldsymbol{s}_{al}, \boldsymbol{s}_{pl}, \boldsymbol{s}_{pm}, \boldsymbol{s}_{fc}\})$. Thus, according to equation $D_{t_l} = \mathbf{1}[p_{t_l} \geq U_{t_l}]$, families of type $\boldsymbol{s}_{al}, \boldsymbol{s}_{pl}$ and $\boldsymbol{s}_{pm}$ also choose t_l. There lies another contradiction since we have established that families of type $\boldsymbol{s}_{pl}$ were already choosing t_h.

In summary, for the full-compliers to choose t_m or t_l, it must be the case that either $\boldsymbol{s}_{pl}$ or $\boldsymbol{s}_{ph}$ do not exist. That is to say that either $P(\boldsymbol{S} = \boldsymbol{s}_{pl}) - 0$ or $P(\boldsymbol{S} = \boldsymbol{s}_{ph}) - 0$. The same rationale applies to the analysis of choices t_m and t_l. Namely, if we set $p_{t_m} = P(\boldsymbol{S} \in \{\boldsymbol{s}_{am}, \boldsymbol{s}_{pm}, \boldsymbol{s}_{ph}\})$, then the full-compliers $\boldsymbol{s}_{fc}$ must choose either t_h or t_l, and this choice behavior is possible only if $P(\boldsymbol{S} = \boldsymbol{s}_{pm}) - 0$ or $P(\boldsymbol{S} = \boldsymbol{s}_{ph}) - 0$. Finally, if we set $p_{t_l} = P(\boldsymbol{S} \in \{\boldsymbol{s}_{al}, \boldsymbol{s}_{pl}, \boldsymbol{s}_{pm}\})$, then the full-compliers $\boldsymbol{s}_{cf}$ must choose either t_h or t_m, and this choice behavior is possible only if $P(\boldsymbol{S} = \boldsymbol{s}_{pl}) - 0$ or $P(\boldsymbol{S} = \boldsymbol{s}_{pm}) - 0$.

The central message conveyed by this theorem is that assuming that partial-compliers precede full-compliers generates a contradictory choice scheme. Fortunately, these contradictions do not occur when full-compliers precede the partial-compliers. Figure A.7 presents the sequence of response types in which the full-compliers precede the partial-compliers.

Figure A.7: Order of Response types Assuming that Full-compliers precede Partial-compliers



Consider the sequence of response types of U_{t_h} displayed in the first bar of Figure A.7. Setting the propensity score of choice t_h to $p_{t_h} = P(\boldsymbol{S} \in \{\boldsymbol{s}_{ah}, \boldsymbol{s}_{ph}, \boldsymbol{s}_{fc}\})$. This means that families of type $\boldsymbol{s}_{al}, \boldsymbol{s}_{ph}$ and $\boldsymbol{s}_{fc}$ choose t_h. The families associated with the remaining response types, that is, $\boldsymbol{s}_{pl}, \boldsymbol{s}_{pm}, \boldsymbol{s}_{am}$ and $\boldsymbol{s}_{al}$ do not choose t_h. This choice behavior is consistent with setting the propensity score of choice t_m to $p_{t_m} = P(\boldsymbol{S} \in \{\boldsymbol{s}_{am}, \boldsymbol{s}_{pm}\})$, and the propensity score of choice t_l to $p_{t_l} = P(\boldsymbol{S} \in \{\boldsymbol{s}_{al}, \boldsymbol{s}_{pl}\})$.

In the case of choice t_m, the sequence of response types in Figure A.7 also provides consistent choice behaviors when we set the propensity scores to $p_{t_m} = P(\boldsymbol{S} \in \{\boldsymbol{s}_{am}, \boldsymbol{s}_{pm}, \boldsymbol{s}_{fc}\})$, $p_{t_h} = P(\boldsymbol{S} \in \{\boldsymbol{s}_{ah}, \boldsymbol{s}_{ph}\})$, and $p_{t_l} = P(\boldsymbol{S} \in \{\boldsymbol{s}_{al}, \boldsymbol{s}_{pl}\})$.

In the case of choice t_l, the sequence of response types in Figure A.7 also provides consistent choice behaviors when we set the propensity scores to $p_{t_l} = P(\boldsymbol{S} \in \{\boldsymbol{s}_{al}, \boldsymbol{s}_{pl}, \boldsymbol{s}_{fc}\})$, $p_{t_m} = P(\boldsymbol{S} \in \{\boldsymbol{s}_{am}, \boldsymbol{s}_{pm}\})$, and $p_{t_h} = P(\boldsymbol{S} \in \{\boldsymbol{s}_{ah}, \boldsymbol{s}_{ph}\})$.

G.3 The Propensity Score Estimator with Covariates

Consider any counterfactual outcome mean $E(Y(t)|\boldsymbol{S}=\boldsymbol{s})$ for a choice $t \in \{t_h, t_m, t_l\}$ identified by the LATE-type parameter that compares the values $z, z' \in \{z_c, z_8, z_e\}$ such that such that $P_t(z') > P_t(z)$. According to Theorem T.4, we can express this counterfactual as following:

$$E(Y(t)|\boldsymbol{S}=\boldsymbol{s}) = \frac{\int_{P_t(z)}^{P_t(z')} E(Y(t)|U_t=u)du}{P_t(z)-P_t(z')}, \tag{160}$$

In summary, equation (160) simply describes a connection between instrumental values z, z', neighborhood choice t and their associated with response type $\boldsymbol{s} \in supp(\boldsymbol{S})$. The numerator in (160) is identified by:

$$\begin{aligned}
&\int_{P_t(z)}^{P_t(z')} E(Y(t)|U_t=u)du = \\
&\quad = E(Y(t)\mathbf{1}[P_t(z) \leq U_t \leq P_t(z')]) \\
&\quad = E(Y(t)\mathbf{1}[U_t \leq P_t(z')] - \mathbf{1}[U_t \geq P_t(z)]) \\
&\quad = E(Y(t)\mathbf{1}[U_t \leq P_t(Z)]|P_t(Z)=P_t(z')) - E(Y(t)\mathbf{1}[U_t \geq P_t(Z)]|P_t(Z)=P_t(z)) \\
&\quad = E(YD_t|P_t(Z)=P_t(z')) - E(YD_t|P_t(Z)=P_t(z))
\end{aligned} \tag{161}$$

This section seeks to identify (160) as a function of propensity scores conditional of $\boldsymbol{X}$, that is $P_t(z,\boldsymbol{x}) \equiv P(T=t|Z=z,\boldsymbol{X}=\boldsymbol{x})$. The conditional version of (161) is given by:

$$\begin{aligned}
E(Y(t)|\boldsymbol{S}=\boldsymbol{s},\boldsymbol{X}=\boldsymbol{x}) &= \frac{\int_{P_t(z',\boldsymbol{x})}^{P_t(z,\boldsymbol{x})} E(Y(t)|U_t=u,\boldsymbol{X}=\boldsymbol{x})du}{P_t(z,\boldsymbol{x})-P_t(z',\boldsymbol{x})} \\
&= \frac{E(YD_t|P_t(Z)=P_t(z',\boldsymbol{x}),\boldsymbol{X}=\boldsymbol{x}) - E(YD_t|P_t(Z)=P_t(z,\boldsymbol{x}),\boldsymbol{X}=\boldsymbol{x})}{P_t(z',\boldsymbol{x})-P_t(z,\boldsymbol{x})}
\end{aligned} \tag{162}$$

Integrating $E(Y(t)|\boldsymbol{S}=\boldsymbol{s},\boldsymbol{X}=\boldsymbol{x})$ over $\boldsymbol{X}$ generates the following equation:

$$\begin{aligned}
\int E(Y(t)|\boldsymbol{S}=\boldsymbol{s},\boldsymbol{X}=\boldsymbol{x})dF_{\boldsymbol{X}|\boldsymbol{S}=\boldsymbol{s}}(\boldsymbol{x}) &= \int E(Y(t)|\boldsymbol{S}=\boldsymbol{s},\boldsymbol{X}=\boldsymbol{x})dF_{\boldsymbol{X}|\boldsymbol{S}=\boldsymbol{s}}(\boldsymbol{x}) \\
&= \int E(Y(t)|\boldsymbol{S}=\boldsymbol{s},\boldsymbol{X}=\boldsymbol{x})\frac{P(\boldsymbol{S}=\boldsymbol{s}|\boldsymbol{X}=\boldsymbol{x})}{P(\boldsymbol{S}=\boldsymbol{s})}dF_{\boldsymbol{X}}(\boldsymbol{x}),
\end{aligned} \tag{163}$$

where the second equality is due to Bayes' theorem. Recall that $P(\boldsymbol{S}=\boldsymbol{s}|\boldsymbol{X}=\boldsymbol{x}) = P_t(z',\boldsymbol{x}) - P_t(z,\boldsymbol{x})$ and thereby $P(\boldsymbol{S}=\boldsymbol{s}) = \int P_t(z',\boldsymbol{x}) - P_t(z,\boldsymbol{x})dF_{\boldsymbol{X}}(\boldsymbol{x})$. Inserting (162) into (163) and using the above results generates:

$$\begin{aligned}
&E(Y(t)|\boldsymbol{S}=\boldsymbol{s}) = \\
&\quad = \frac{\int \Big(E(YD_t|P_t(Z)=P_t(z',\boldsymbol{x}),\boldsymbol{X}=\boldsymbol{x}) - E(YD_t|P_t(Z)=P_t(z,\boldsymbol{x}),\boldsymbol{X}=\boldsymbol{x})\Big)dF_{\boldsymbol{X}}(\boldsymbol{x})}{\int P_t(z',\boldsymbol{x}) - P_t(z,\boldsymbol{x})dF_{\boldsymbol{X}}(\boldsymbol{x})}
\end{aligned} \tag{164}$$

G.4 Estimation Formulas for Counterfactual Outcomes

Let $P_t(z,\boldsymbol{x}) = P(T=t|Z=z,\boldsymbol{X}=\boldsymbol{x})$ be the propensity score for choice $t \in \{t_h, t_m, t_l\}$ conditioned on the baseline variables $\boldsymbol{X}=\boldsymbol{x}$ given the instrumental value $z \in \{z_c, z_8, z_e\}$. Let $M_t(p,\boldsymbol{x}) = E(Y \cdot D_t|P_t=p,\boldsymbol{X}=\boldsymbol{x})$ be the expected value of the interaction $Y \cdot D_t$ for $t \in \{t_h, t_m, t_l\}$ conditioned on the propensity score value $P_t = p$ and baseline variables $\boldsymbol{X}=\boldsymbol{x}$. Under this notation, the counterfactual outcome means for t_h can be estimated by the empirical counterpart of the following expressions:

$$E(Y(t_h)|\boldsymbol{S}=\boldsymbol{s}_{ah})=\frac{\int \big(M_{t_h}(P_{t_h}(z_8,\boldsymbol{x}),\boldsymbol{x})\big)dF_{\boldsymbol{X}}(\boldsymbol{x})}{\int P_{t_h}(z_8,\boldsymbol{x})dF_{\boldsymbol{X}}(\boldsymbol{x})},$$
$$E(Y(t_h)|\boldsymbol{S}=\boldsymbol{s}_{ph})=\frac{\int \big(M_{t_h}(P_{t_h}(z_e,\boldsymbol{x}),\boldsymbol{x})-M_{t_h}(P_{t_h}(z_8,\boldsymbol{x}),\boldsymbol{x})\big)dF_{\boldsymbol{X}}(\boldsymbol{x})}{\int \big(P_{t_h}(z_e,\boldsymbol{x})-P_{t_h}(z_8,\boldsymbol{x})\big)dF_{\boldsymbol{X}}(\boldsymbol{x})},$$
$$E(Y(t_h)|\boldsymbol{S}=\boldsymbol{s}_{fc})=\frac{\int \big(M_{t_h}(P_{t_h}(z_e,\boldsymbol{x})+P_{fc}(\boldsymbol{x}),\boldsymbol{x})-M_{t_h}(P_{t_h}(z_e,\boldsymbol{x}),\boldsymbol{x})\big)dF_{\boldsymbol{X}}(\boldsymbol{x})}{\int P_{fc}(\boldsymbol{x})dF_{\boldsymbol{X}}(\boldsymbol{x})},$$
$$E(Y(t_h)|\boldsymbol{S}=\boldsymbol{s}_{pl})=\frac{\int \big(M_{t_h}(P_{t_h}(z_c,\boldsymbol{x}),\boldsymbol{x})-M_{t_h}(P_{t_h}(z_e,\boldsymbol{x})+P_{fc}(\boldsymbol{x}),\boldsymbol{x})\big)dF_{\boldsymbol{X}}(\boldsymbol{x})}{\int \big(P_{t_h}(z_c,\boldsymbol{x})-P_{t_h}(z_e,\boldsymbol{x})-P_{fc}(\boldsymbol{x})\big)dF_{\boldsymbol{X}}(\boldsymbol{x})},$$
$$\text{where } P_{fc}(\boldsymbol{x})=\big(P_{t_h}(z_8,\boldsymbol{x})-P_{t_h}(z_e,\boldsymbol{x})\big)-\big(P_{t_m}(z_c,\boldsymbol{x})-P_{t_m}(z_8,\boldsymbol{x})\big).$$

The counterfactual outcome means for t_m can be estimated by the empirical counterpart of the following expressions:

$$E(Y(t_m)|\boldsymbol{S}=\boldsymbol{s}_{am})=\frac{\int \big(M_{t_m}(P_{t_m}(z_e,\boldsymbol{x}),\boldsymbol{x})\big)dF_{\boldsymbol{X}}(\boldsymbol{x})}{\int P_{t_m}(z_e,\boldsymbol{x})dF_{\boldsymbol{X}}(\boldsymbol{x})},$$
$$E(Y(t_m)|\boldsymbol{S}=\boldsymbol{s}_{pm})=\frac{\int \big(M_{t_m}(P_{t_m}(z_c,\boldsymbol{x}),\boldsymbol{x})-M_{t_m}(P_{t_m}(z_e,\boldsymbol{x}),\boldsymbol{x})\big)dF_{\boldsymbol{X}}(\boldsymbol{x})}{\int \big(P_{t_m}(z_c,\boldsymbol{x})-P_{t_m}(z_e,\boldsymbol{x})\big)dF_{\boldsymbol{X}}(\boldsymbol{x})},$$
$$E(Y(t_m)|\boldsymbol{S}=\boldsymbol{s}_{fc})=\frac{\int \big(M_{t_m}(P_{t_m}(z_c,\boldsymbol{x})+P_{fc}(\boldsymbol{x}),\boldsymbol{x})-M_{t_m}(P_{t_m}(z_c,\boldsymbol{x}),\boldsymbol{x})\big)dF_{\boldsymbol{X}}(\boldsymbol{x})}{\int P_{fc}(\boldsymbol{x})dF_{\boldsymbol{X}}(\boldsymbol{x})},$$
$$E(Y(t_m)|\boldsymbol{S}=\boldsymbol{s}_{ph})=\frac{\int \big(M_{t_m}(P_{t_m}(z_8,\boldsymbol{x}),\boldsymbol{x})-M_{t_m}(P_{t_m}(z_e,\boldsymbol{x})+P_{fc}(\boldsymbol{x}),\boldsymbol{x})\big)dF_{\boldsymbol{X}}(\boldsymbol{x})}{\int \big(P_{t_m}(z_8,\boldsymbol{x})-P_{t_m}(z_c,\boldsymbol{x})-P_{fc}(\boldsymbol{x})\big)dF_{\boldsymbol{X}}(\boldsymbol{x})},$$
$$\text{where } P_{fc}(\boldsymbol{x})=\big(P_{t_m}(z_8,\boldsymbol{x})-P_{t_m}(z_e,\boldsymbol{x})\big)-\big(P_{t_m}(z_c,\boldsymbol{x})-P_{t_m}(z_8,\boldsymbol{x})\big).$$

The counterfactual outcome means for t_l can be estimated by the empirical counterpart of the following expressions:

$$E(Y(t_l)|\boldsymbol{S}=\boldsymbol{s}_{am})=\frac{\int \big(M_{t_l}(P_{t_l}(z_c,\boldsymbol{x}),\boldsymbol{x})\big)dF_{\boldsymbol{X}}(\boldsymbol{x})}{\int P_{t_l}(z_c,\boldsymbol{x})dF_{\boldsymbol{X}}(\boldsymbol{x})},$$
$$E(Y(t_l)|\boldsymbol{S}=\boldsymbol{s}_{pm})=\frac{\int \big(M_{t_l}(P_{t_l}(z_8,\boldsymbol{x}),\boldsymbol{x})-M_{t_l}(P_{t_l}(z_c,\boldsymbol{x}),\boldsymbol{x})\big)dF_{\boldsymbol{X}}(\boldsymbol{x})}{\int \big(P_{t_l}(z_8,\boldsymbol{x})-P_{t_l}(z_c,\boldsymbol{x})\big)dF_{\boldsymbol{X}}(\boldsymbol{x})},$$
$$E(Y(t_l)|\boldsymbol{S}=\boldsymbol{s}_{fc})=\frac{\int \big(M_{t_l}(P_{t_l}(z_8,\boldsymbol{x})+P_{fc}(\boldsymbol{x}),\boldsymbol{x})-M_{t_l}(P_{t_l}(z_8,\boldsymbol{x}),\boldsymbol{x})\big)dF_{\boldsymbol{X}}(\boldsymbol{x})}{\int P_{fc}(\boldsymbol{x})dF_{\boldsymbol{X}}(\boldsymbol{x})},$$
$$E(Y(t_l)|\boldsymbol{S}=\boldsymbol{s}_{ph})=\frac{\int \big(M_{t_l}(P_{t_l}(z_e,\boldsymbol{x}),\boldsymbol{x})-M_{t_l}(P_{t_l}(z_8,\boldsymbol{x})+P_{fc}(\boldsymbol{x}),\boldsymbol{x})\big)dF_{\boldsymbol{X}}(\boldsymbol{x})}{\int \big(P_{t_l}(z_e,\boldsymbol{x})-P_{t_l}(z_8,\boldsymbol{x})-P_{fc}(\boldsymbol{x})\big)dF_{\boldsymbol{X}}(\boldsymbol{x})},$$
$$\text{where } P_{fc}(\boldsymbol{x})=\big(P_{t_l}(z_8,\boldsymbol{x})-P_{t_l}(z_e,\boldsymbol{x})\big)-\big(P_{t_l}(z_c,\boldsymbol{x})-P_{t_l}(z_8,\boldsymbol{x})\big).$$

G.5 Estimation Procedure of Counterfactual Outcomes

The estimation of counterfactual outcomes is based on a standard application of the propensity score estimator. We use the ingredients (60) and (62) to estimate expressions such as (49)–(51). The method consists of the following steps:

1. Estimate the conditional propensity score $P_{t,i}(z) \equiv P(T=t|Z=z,\boldsymbol{X}=\boldsymbol{X}_i,\boldsymbol{K}=\boldsymbol{K}_i)$ for $(t,z)\in\{t_h,t_l,t_m\}\times\{z_c,z_8,z_e\}$, given the baseline characteristics $\boldsymbol{X}_i,\boldsymbol{K}_i$ of family i;
2. Estimate the conditional expected value of the interaction $M_{t,i}(p)=E(Y\cdot D_t|P_t=p,\boldsymbol{X}=$

$\boldsymbol{X}_i, \boldsymbol{K} = \boldsymbol{K}_i$) as a function of the propensity scores P_t for choice t, given the baseline characteristics $\boldsymbol{X}_i, \boldsymbol{K}_i$ of family i;

3. Estimate the counterfactual outcome means $E(Y(t)|\boldsymbol{S} = \boldsymbol{s})$ using the empirical counterpart of the propensity score estimator discussed in (49).

The first step estimates the propensity scores using the following linear probability model:

$$D_{t,i} = \sum_{z \in \{z_c, z_8, z_e\}} \mathbf{1}[Z_i = z] \cdot \Big(\alpha_{t,z} + \boldsymbol{X}_i \boldsymbol{\theta}_{t,z} + \boldsymbol{K}_i \boldsymbol{\gamma}_{t,z}\Big) + \epsilon_{t,i}; t \in \{t_l, t_m, t_h\}. \tag{165}$$

The estimate for the propensity score for a family i and IV-value z is given by:

$$\hat{P}_{t,i}(z) = \hat{\alpha}_{t,z} + \boldsymbol{X}_i \hat{\boldsymbol{\theta}}_{t,z} + \boldsymbol{K}_i \hat{\boldsymbol{\gamma}}_{t,z}; \text{ for } (t, z) \in \{t_h, t_l, t_m\} \times \{z_c, z_8, z_e\}.$$

In particular, the estimate for the full-complier probability conditioned on the baseline characteristics of family i is $\hat{P}_i(\boldsymbol{s}_{fc}) = \big(\hat{P}_{t_h,i}(z_8) - \hat{P}_{t_h,i}(z_e)\big) - \big(\hat{P}_{t_m,i}(z_c) - \hat{P}_{t_m,i}(z_8)\big)$. The fact that baseline variables $\boldsymbol{X}, \boldsymbol{K}$ are standardized to have mean zero assures that the estimates for propensity scores $\hat{P}_{t,i}(z_c), \hat{P}_{t,i}(z_8), \hat{P}_{t,i}(z_e)$ sum to one for each family i. The linear probability model does not impose positive probabilities.

The second step evaluates the conditional expectation of the interaction YD_t as a local polynomial of propensity the score estimates:

$$Y_i \cdot D_{t,i} = \sum_{k=0}^{3} \alpha_k \cdot \big(\hat{P}_{t,i}\big)^k + \Big(\hat{P}_{t,i} \cdot \boldsymbol{K}_i\Big)\boldsymbol{\xi}_t + \Big(\hat{P}_{t,i} \cdot \boldsymbol{X}_i\Big)\boldsymbol{\psi}_t + \boldsymbol{K}_i \boldsymbol{\gamma}_t + \boldsymbol{X}_i \boldsymbol{\theta}_t + \epsilon_{t,i}, \tag{166}$$

where $\hat{P}_{t,i} \equiv \hat{P}_{t,i}(Z_i)$ is the propensity score of family i. Appendix H evaluates the propensity scores using the multinomial logistic regression. The empirical results using the logistic model are closely related to the ones presented in the main paper. The estimate for $M_{t,i}(p)$ is $\hat{M}_{t,i}(p) = \sum_{k=0}^{3} \hat{\alpha}_k \cdot p^k + p(\boldsymbol{K}_i \hat{\boldsymbol{\xi}}_t + \boldsymbol{X}_i \hat{\boldsymbol{\psi}}_t) + \boldsymbol{K}_i \hat{\boldsymbol{\gamma}}_t + \boldsymbol{X}_i \hat{\boldsymbol{\theta}}_t$.

Start to evaluate the conditional expectations of the counterfactual outcomes. For instance, $E(Y(t_l)|\boldsymbol{S} = \boldsymbol{s}_{fc})$ in (49), we use the previous to establish $[P_{t_l}(z_8), P_{t_l}(z_8) + P(\boldsymbol{s}_{fc})]$ to form the empirical counterpart of equation (52):

$$\hat{E}(Y(t_l)|\boldsymbol{S} = \boldsymbol{s}_{fc}) = \frac{\sum_i \Big(\hat{M}_{t_l,i}\big(\hat{P}_{t_l,i}(z_8) + \hat{P}_i(\boldsymbol{s}_{fc})\big) - \hat{M}_{t_l,i}\big(\hat{P}_{t_l,i}(z_8)\big)\Big) \cdot W_i}{\sum_i \hat{P}_i(\boldsymbol{s}_{fc}) \cdot W_i} \tag{167}$$

where W_i denotes the MTO weights.

G.6 Connection between TSLS and Propensity Score Estimations

$E(Y(t_l)|\boldsymbol{S} = \boldsymbol{s}_{fc})$ and $E(Y(t_l)|\boldsymbol{S} = \boldsymbol{s}_{pm})$ are estimated by interpolation.[72] It is instructive that explain the connection with this interpolation approach with the standard 2SLS regression.

[72] Recently, Brinch, Mogstad, and Wiswall (2017); Kline and Walters (2017); Mogstad, Andres, and Torgovitsky (2017); Mogstad and Torgovitsky (2018) have studied the problem of assessing ATE in binary choice models with discrete instruments by extrapolating the Marginal Treatment Effect MTE parameter of Heckman and Vytlacil (2005). The method described here differs from this literature in three instances: it investigates a multiple choice model instead of the binary case, it seeks to identify a LATE parameter instead of ATE and it employs interpolation instead of extrapolation.

A standard empirical approach is to estimate $P_t(z) = P(T = t|Z = z)$ and then evaluate D_{t_l} and YD_{t_l} as a polynomial of propensity scores:

$$D_{t_l,i} = \theta_0 + \theta_1 P_{t_l,i} + \theta_2 P_{t_l,i}^2 + \epsilon_{i,D} = \boldsymbol{\Lambda}(P_{t_l,i})\boldsymbol{\theta} + \epsilon_{i,D}, \tag{168}$$

$$Y_i D_{t_l,i} = \beta_0 + \beta_1 P_{t_l,i} + \beta_2 P_{t_l,i}^2 + \epsilon_{i,Y} = \boldsymbol{\Lambda}(P_{t_l,i})\boldsymbol{\beta} + \epsilon_{i,D}, \tag{169}$$

where $P_{t_l,i} = P_{t_l}(Z_i)$ denotes the propensity score of t_l for family i, and $\boldsymbol{\Lambda}(p) = [1, p, p^2]$ simply stacks the polynomial as a vector. The estimator for $E(Y(t_l)|\boldsymbol{S} \in \{\boldsymbol{s}_{fc}, \boldsymbol{s}_{pm}\})$ in (170) is numerically the same as the TSLS estimator of Table A.8. Moreover, this equivalence holds for any choice of linearly independent functions in $\boldsymbol{\Lambda}(z)$ and for any method that estimates the propensity scores. Kline and Walters (2019) provides a recent discussion on numerical equivalence among estimators for the LATE model.

The key feature that is at the core of such equivalences is the space spanned by suing the indicators of the IV-values versus the polynomials of the propensity scores. To gain intuition, suppose that the instrument Z takes three values z_1, z_2, z_3 and let $\hat{p}_1, \hat{p}_2, \hat{p}_3$ the propensity scores associated with a treatment indicator D_t. Now consider the linear regression that uses the interaction $Y \cdot D_t$ as dependent variable. Using the indicator of the IV-values or a second degree polynomial generates the same fitted values of the regression. Indeed, the matrix of explanatory variables in both models span the same space. The two matrices below represent the values of the explanatory variables as Z ranges in z_1, z_2, z_3 for each model. Both matrices have full rank and span the same space.

$$\begin{matrix} z_1 \\ z_2 \\ z_3 \end{matrix} \quad \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \begin{bmatrix} 1 & p_1 & p_1^2 \\ 1 & p_2 & p_2^2 \\ 1 & p_3 & p_3^2 \end{bmatrix}.$$

Returning to model (168)–(169), the estimation of the counterfactual outcome means is given by:

$$\hat{E}(Y(t_l)|\boldsymbol{S} \in \{\boldsymbol{s}_{fc}, \boldsymbol{s}_{pm}\}) = \frac{\Big(\boldsymbol{\Lambda}(P_{t_l}(z_e)) - \boldsymbol{\Lambda}(P_{t_l}(z_8))\Big)'\widehat{\boldsymbol{\beta}}_t}{\Big(\boldsymbol{\Lambda}(P_{t_l}(z_e)) - \boldsymbol{\Lambda}(P_{t_l}(z_8))\Big)'\widehat{\boldsymbol{\theta}}_t}, \tag{170}$$

The response type probability is estimated as $P(\boldsymbol{S} = \boldsymbol{s}_{fc}) = \big(P_{t_h}(z_8) - P_{t_h}(z_e)\big) - \big(P_{t_m}(z_c) - P_{t_m}(z_8)\big)$. We can then evaluate the probability $P^* = P_{t_l}(z_8) + P(\boldsymbol{S} = \boldsymbol{s}_{fc})$ and disentangle $E(Y(t_l)|\boldsymbol{S} \in \{\boldsymbol{s}_{fc}, \boldsymbol{s}_{pm}\})$ in (170) via:

$$\hat{E}(Y(t_l)|\boldsymbol{S} = \boldsymbol{s}_{fc}) = \frac{\Big(\boldsymbol{\Lambda}(P^*) - \boldsymbol{\Lambda}(P_{t_l}(z_8))\Big)'\widehat{\boldsymbol{\beta}}_t}{\Big(\boldsymbol{\Lambda}(P^*) - \boldsymbol{\Lambda}(P_{t_l}(z_8))\Big)'\widehat{\boldsymbol{\theta}}_t}, \quad \hat{E}(Y(t_l)|\boldsymbol{S} = \boldsymbol{s}_{pm}) = \frac{\Big(\boldsymbol{\Lambda}(P_{t_l}(z_e)) - \boldsymbol{\Lambda}(P^*)\Big)'\widehat{\boldsymbol{\beta}}_t}{\Big(\boldsymbol{\Lambda}(P_{t_l}(z_e)) - \boldsymbol{\Lambda}(P^*)\Big)'\widehat{\boldsymbol{\theta}}_t}. \tag{171}$$

H Sensitivity Analyses

This section present additional evaluations that check the robustness of these findings under modifications of the baseline model.

Tables A.10–A.12 presents results based on variations of the original model that generates Table 7 of the main text.

Table A.10 suppresses the interaction of site fixed effects and the propensity scores. Suppressing the interaction between propensity scores and site fixed effects forces cities to shift the the mean potential outcomes for a given neighborhood choice in parallel across all response types.

Table A.11 suppresses the interaction of baseline variables and propensity scores. This forces that the family baseline characteristics to shift the mean potential outcomes for a given neighborhood choice in parallel.

Table A.12 estimates the same outcome equation displayed in the main text. The model however uses a multinomial logit model to estimate propensity scores, instead of the linear probability model in (165).

Estimates in Tables A.10–A.12 are very close to those presented in Table 7.

Table A.10: Causal Effects for Full Compliers $\boldsymbol{S} = \boldsymbol{s}_{fc}$ (No Site Interaction)

	$E(Y(t_l) - Y(t_h)\|\boldsymbol{s}_{fc})$		$E(Y(t_l) - Y(t_m)\|\boldsymbol{s}_{fc})$		$E(Y(t_m) - Y(t_h)\|\boldsymbol{s}_{fc})$	
Income of Family Head	2.030	***	0.110		1.919	**
(s.e.)	0.761		0.896		0.902	
(p-value)	0.005		0.905		0.032	
Income of Head and Spouse	0.710		0.486		0.224	
(s.e.)	0.826		0.899		1.019	
(p-value)	0.393		0.598		0.800	
Total household income	1.421		1.083		0.337	
(s.e.)	0.871		0.981		1.060	
(p-value)	0.117		0.277		0.752	
Above Poverty Line	0.089	**	0.038		0.051	
(s.e.)	0.039		0.050		0.053	
(p-value)	0.018		0.453		0.348	
Employed without welfare	0.102	**	0.034		0.068	
(s.e.)	0.044		0.057		0.059	
(p-value)	0.027		0.570		0.245	
Currently on welfare	−0.111	***	0.061		−0.172	***
(s.e.)	0.041		0.055		0.057	
(p-value)	0.008		0.270		0.010	
Job tenure	0.073		0.028		0.044	
(s.e.)	0.044		0.052		0.053	
(p-value)	0.102		0.607		0.393	
Economic self-sufficiency	0.062	*	−0.023		0.085	*
(s.e.)	0.033		0.045		0.045	
(p-value)	0.062		0.592		0.055	
Neighborhood Poverty	−32.843	***	−20.999	***	−11.844	***
(s.e.)	0.996		1.690		1.914	
(p-value)	0.000		0.000		0.000	

This table evaluates the neighborhood effects for full compliers $\boldsymbol{s}_{fc}$ across several outcomes. The first column lists the outcome variables. The second column evaluates the causal effect between the neighborhood types of low and high poverty. The third column compares low versus medium poverty neighborhoods and the last column evaluates the neighborhood effects between medium versus high poverty types. The results are based on a semi-parametric method that evaluates propensity scores and response type probabilities using a linear probability model. All estimates are conditioned on the site of intervention and account for the person-level weight for adult survey of the interim analyses (Interim Impacts Evaluation manual, 2003, Appendix B). Inference is obtained by a bootstrap method that employs a weighted sampling scheme. The p-values are associated with the double-tailed inference that tests if the estimates are equal to zero. Asterisks indicate the typical p-value thresholds: *** for p-value < 0.01, ** for $0.01 \leq p$-value < 0.05, * for $0.05 \leq p$-value < 0.1.

Table A.11: Causal Effects for Full Compliers $\boldsymbol{S} = \boldsymbol{s}_{fc}$ (No Covariate Interaction)

	$E(Y(t_l) - Y(t_h) \mid \boldsymbol{s}_{fc})$		$E(Y(t_l) - Y(t_m) \mid \boldsymbol{s}_{fc})$		$E(Y(t_m) - Y(t_h) \mid \boldsymbol{s}_{fc})$	
Income of Family Head	2.188	***	−0.289		2.477	
(s.e.)	0.822		1.516		1.469	
(p-value)	0.005		0.847		0.112	
Income of Head and Spouse	0.865		0.176		0.689	
(s.e.)	0.862		1.642		1.711	
(p-value)	0.335		0.910		0.663	
Total household income	2.040	**	0.693		1.347	
(s.e.)	0.908		1.621		1.623	
(p-value)	0.035		0.647		0.420	
Above Poverty Line	0.122	***	−0.050		0.173	*
(s.e.)	0.042		0.086		0.084	
(p-value)	0.008		0.607		0.082	
Employed without welfare	0.114	**	0.140		−0.026	
(s.e.)	0.046		0.095		0.097	
(p-value)	0.022		0.157		0.803	
Currently on welfare	−0.130	***	−0.048		−0.081	
(s.e.)	0.044		0.090		0.089	
(p-value)	0.007		0.565		0.343	
Job tenure	0.094	**	0.061		0.033	
(s.e.)	0.047		0.096		0.094	
(p-value)	0.048		0.533		0.723	
Economic self-sufficiency	0.077	**	−0.133		0.210	**
(s.e.)	0.034		0.087		0.084	
(p-value)	0.027		0.177		0.030	
Neighborhood Poverty	−33.283	***	−21.631	***	−11.652	***
(s.e.)	1.008		2.216		2.279	
(p-value)	0.000		0.000		0.000	

This table evaluates the neighborhood effects for full compliers $\boldsymbol{s}_{fc}$ across several outcomes. The first column lists the outcome variables. The second column evaluates the causal effect between the neighborhood types of low and high poverty. The third column compares low versus medium poverty neighborhoods and the last column evaluates the neighborhood effects between medium versus high poverty types. The results are based on a semi-parametric method that evaluates propensity scores and response type probabilities using a linear probability model. All estimates are conditioned on the site of intervention and account for the person-level weight for adult survey of the interim analyses (Interim Impacts Evaluation manual, 2003, Appendix B). Inference is obtained by a bootstrap method that employs a weighted sampling scheme. The p-values are associated with the double-tailed inference that tests if the estimates are equal to zero. Asterisks indicate the typical p-value thresholds: *** for p-value < 0.01, ** for $0.01 \leq p$-value < 0.05, * for $0.05 \leq p$-value < 0.1.

Table A.12: Causal Effects for Full Compliers $\boldsymbol{S} = \boldsymbol{s}_{fc}$ (Using Multinomial Logit)

	$E(Y(t_l) - Y(t_h) \mid \boldsymbol{s}_{fc})$		$E(Y(t_l) - Y(t_m) \mid \boldsymbol{s}_{fc})$		$E(Y(t_m) - Y(t_h) \mid \boldsymbol{s}_{fc})$	
Income of Family Head	2.471	***	0.887		1.585	
(s.e.)	0.821		1.179		1.225	
(p-value)	0.005		0.453		0.220	
Income of Head and Spouse	1.298		1.278		0.021	
(s.e.)	0.923		1.371		1.399	
(p-value)	0.195		0.350		0.993	
Total household income	2.161	**	2.780	**	−0.619	
(s.e.)	0.954		1.379		1.423	
(p-value)	0.032		0.048		0.668	
Above Poverty Line	0.133	***	0.077		0.057	
(s.e.)	0.048		0.069		0.065	
(p-value)	0.010		0.275		0.408	
Employed without welfare	0.117	**	0.096		0.021	
(s.e.)	0.051		0.079		0.079	
(p-value)	0.023		0.223		0.808	
Currently on welfare	−0.108	**	−0.051		−0.057	
(s.e.)	0.046		0.069		0.068	
(p-value)	0.028		0.462		0.372	
Job tenure	0.120	**	0.053		0.067	
(s.e.)	0.050		0.077		0.079	
(p-value)	0.028		0.457		0.398	
Economic self-sufficiency	0.076	*	−0.044		0.119	*
(s.e.)	0.042		0.059		0.057	
(p-value)	0.087		0.485		0.067	
Neighborhood Poverty	−32.893	***	−22.737	***	−10.157	***
(s.e.)	1.150		1.725		1.995	
(p-value)	0.000		0.000		0.000	

This table evaluates the neighborhood effects for full compliers $\boldsymbol{s}_{fc}$ across several outcomes. The first column lists the outcome variables. The second column evaluates the causal effect between the neighborhood types of low and high poverty. The third column compares low versus medium poverty neighborhoods and the last column evaluates the neighborhood effects between medium versus high poverty types. The results are based on a semi-parametric method that evaluates propensity scores and response type probabilities using a multinomial logit model. All estimates are conditioned on the site of intervention and account for the person-level weight for adult survey of the interim analyses (Interim Impacts Evaluation manual, 2003, Appendix B). Inference is obtained by a bootstrap method that employs a weighted sampling scheme. The p-values are associated with the double-tailed inference that tests if the estimates are equal to zero. Asterisks indicate the typical p-value thresholds: *** for p-value < 0.01, ** for $0.01 \leq p$-value < 0.05, * for $0.05 \leq p$-value < 0.1.