Web Appendix

A Model Properties and Identification Analyses

A.1 Proof of Theorem T.1

We use the following auxiliary lemma to prove the result:

Lemma L.1. Let a linear system $\boldsymbol{y} = \boldsymbol{B}\boldsymbol{x}$, where $\boldsymbol{B}$ is a real-valued matrix with full row-rank and ξ be a vector with the same dimension of $\boldsymbol{x}$. Thus, $\xi'\boldsymbol{x}$ is point identified if and only if $\xi'\xi = \xi'\boldsymbol{B}'(\boldsymbol{B}\boldsymbol{B}')^{-1}\boldsymbol{B}\xi$.

Proof. The general solution for $\boldsymbol{x}$ in the system of linear equations represented by $\boldsymbol{y} = \boldsymbol{B}\boldsymbol{x}$ is:[24]

$$\boldsymbol{y} = \boldsymbol{B}\boldsymbol{x} \Rightarrow \boldsymbol{x} = \boldsymbol{B}^{+}\boldsymbol{y} + (\boldsymbol{I} - \boldsymbol{B}^{+}\boldsymbol{B})\boldsymbol{\lambda} \tag{47}$$

where $\boldsymbol{\lambda}$ is an arbitrary real-valued $|\boldsymbol{x}|$-dimension vector, $\boldsymbol{I}$ is an identity matrix of the same dimension and $\boldsymbol{B}^{+}$ is the Moore–Penrose Pseudoinverse of matrix $\boldsymbol{B}$.[25] It follows that a linear combination $\xi'\boldsymbol{x}$ is point identified if and only if $\xi'(\boldsymbol{I} - \boldsymbol{B}^{+}\boldsymbol{B}) = \boldsymbol{0}$. Note that $\boldsymbol{B}^{+}\boldsymbol{B}$ denotes an orthogonal projection since $(\boldsymbol{B}^{+}\boldsymbol{B})' = \boldsymbol{B}^{+}\boldsymbol{B}$ and $(\boldsymbol{B}^{+}\boldsymbol{B})\cdot(\boldsymbol{B}^{+}\boldsymbol{B}) = \boldsymbol{B}^{+}\boldsymbol{B}$ holds. Thus, it is also the case that $\boldsymbol{I} - \boldsymbol{B}^{+}\boldsymbol{B}$ is an orthogonal projection and therefore $(\boldsymbol{I} - \boldsymbol{B}^{+}\boldsymbol{B})(\boldsymbol{I} - (\boldsymbol{B}^{+}\boldsymbol{B}))' = \boldsymbol{I}' - (\boldsymbol{B}^{+}\boldsymbol{B})$. Combining these properties, we have that:

$$\xi'(\boldsymbol{I} - \boldsymbol{B}^{+}\boldsymbol{B}) = \boldsymbol{0} \quad \Leftrightarrow \quad \xi'(\boldsymbol{I} - \boldsymbol{B}^{+}\boldsymbol{B})(\xi'(\boldsymbol{I} - \boldsymbol{B}^{+}\boldsymbol{B}))' = 0 \quad \Leftrightarrow \quad \xi'(\boldsymbol{I} - \boldsymbol{B}^{+}\boldsymbol{B})\xi = 0. \quad \Leftrightarrow \quad \xi'\xi = \xi'\boldsymbol{B}^{+}\boldsymbol{B}\xi.$$

Note that if matrix $\boldsymbol{B}$ has full row rank, the pseudo-inverse matrix is given by $\boldsymbol{B}^{+} = \boldsymbol{B}'(\boldsymbol{B}\boldsymbol{B}')^{-1}$.[26] We can combine these properties to state that $\xi'\boldsymbol{x}$ is point identified if and only if $\xi'\xi = \xi'\boldsymbol{B}'(\boldsymbol{B}\boldsymbol{B}')^{-1}\boldsymbol{B}\xi$. □

Equation (6) establishes the following systems of linear equations:

$$\boldsymbol{Q}_Z(t) \odot \boldsymbol{P}_Z(t) = \boldsymbol{B}_t\left(\boldsymbol{Q}_S(t) \odot \boldsymbol{P}_S\right) \text{ and } \boldsymbol{P}_Z(t) = \boldsymbol{B}_t\boldsymbol{P}_S \text{ for all } t \in \mathcal{T}.$$

We seek to examine the identification of $E(Y(t)|\boldsymbol{S} \in \tilde{\mathcal{S}})$ for some response type set $\tilde{\mathcal{S}} \subset \mathcal{S}$. Let $\boldsymbol{b}(\tilde{\mathcal{S}})$ be the $N_S \times 1$ vector that indicates the types that belongs to set $\tilde{\mathcal{S}}$, namely:

$$\boldsymbol{b}(\tilde{\mathcal{S}}) = \left[\boldsymbol{1}[\boldsymbol{s}_1 \in \tilde{\mathcal{S}}], ..., \boldsymbol{1}[\boldsymbol{s}_{N_S} \in \tilde{\mathcal{S}}]\right]'.$$

Thus we can express $E(Y(t)|\boldsymbol{S} \in \tilde{\mathcal{S}})$ as:

$$E(Y(t)|\boldsymbol{S} \in \tilde{\mathcal{S}}) = \frac{\boldsymbol{b}(\tilde{\mathcal{S}})'\left(\boldsymbol{Q}_S(t) \odot \boldsymbol{P}_S\right)}{\boldsymbol{b}(\tilde{\mathcal{S}})'\boldsymbol{P}_S}. \tag{48}$$

According to Lemma **L.1**, the criteria for the identification of both the numerator and the denominator of the ratio in (48) is given by $\boldsymbol{b}(\tilde{\mathcal{S}})'\boldsymbol{b}(\tilde{\mathcal{S}}) = \boldsymbol{b}(\tilde{\mathcal{S}})'\boldsymbol{B}_t'(\boldsymbol{B}_t\boldsymbol{B}_t')^{-1}\boldsymbol{B}_t\boldsymbol{b}(\tilde{\mathcal{S}})$. Note that $\boldsymbol{b}(\tilde{\mathcal{S}})$ is an indicator vector. Thus, $\boldsymbol{b}(\tilde{\mathcal{S}})'\boldsymbol{b}(\tilde{\mathcal{S}})$ is simply the cardinality of $\tilde{\mathcal{S}}$, that is, $\boldsymbol{b}(\tilde{\mathcal{S}})'\boldsymbol{b}(\tilde{\mathcal{S}}) = |\tilde{\mathcal{S}}|$. The term $\boldsymbol{B}_t\boldsymbol{b}(\tilde{\mathcal{S}})$ is the sum of the columns of $\boldsymbol{B}_t$ corresponding to the types in $\tilde{\mathcal{S}}$, that is,

[24] See Magnus and Neudecker (1999) for a general discussion of linear systems.

[25] The Moore–Penrose Pseudoinverse $\boldsymbol{B}^{+}$ of matrix $\boldsymbol{B}$ is unique and defined by the following properties: (1) $\boldsymbol{B}\boldsymbol{B}^{+}\boldsymbol{B} = \boldsymbol{B}$; (2) $\boldsymbol{B}^{+}\boldsymbol{B}\boldsymbol{B}^{+} = \boldsymbol{B}^{+}$; (3) $\boldsymbol{B}^{+}\boldsymbol{B} = (\boldsymbol{B}^{+}\boldsymbol{B})'$; and (4) $\boldsymbol{B}\boldsymbol{B}^{+} = (\boldsymbol{B}\boldsymbol{B}^{+})'$.

[26] See Magnus and Neudecker (1999).

$\boldsymbol{B}_t\boldsymbol{b}(\tilde{\mathcal{S}}) = \sum_{\boldsymbol{s}\in\tilde{\mathcal{S}}} \boldsymbol{B}_t[\cdot,\boldsymbol{s}]$. Combining these results, we have the criteria:

$$\boldsymbol{b}(\tilde{\mathcal{S}})'\boldsymbol{b}(\tilde{\mathcal{S}}) = \boldsymbol{b}(\tilde{\mathcal{S}})'\boldsymbol{B}_t'(\boldsymbol{B}_t\boldsymbol{B}_t')^{-1}\boldsymbol{B}_t\boldsymbol{b}(\tilde{\mathcal{S}}) \Leftrightarrow \frac{\left(\sum_{\boldsymbol{s}\in\tilde{\mathcal{S}}} \boldsymbol{B}_t[\cdot,\boldsymbol{s}]\right)'\left(\boldsymbol{B}_t\boldsymbol{B}_t'\right)^{-1}\left(\sum_{\boldsymbol{s}\in\tilde{\mathcal{S}}} \boldsymbol{B}_t[\cdot,\boldsymbol{s}]\right)}{|\tilde{\mathcal{S}}|} = 1.$$

This result proves the first part of the theorem. The second part of the theorem employs the general solution of linear systems in (47). If the identification criteria holds, then $P(\boldsymbol{S}\in\tilde{\mathcal{S}})$ can be expressed as:

$$P(\boldsymbol{S}\in\tilde{\mathcal{S}}) = \boldsymbol{b}(\tilde{\mathcal{S}})'\boldsymbol{B}_t^{+}\boldsymbol{P}_Z(t) = \boldsymbol{b}(\tilde{\mathcal{S}})'\boldsymbol{B}_t\left(\boldsymbol{B}_t\boldsymbol{B}_t'\right)^{-1}\boldsymbol{P}_Z(t) = \Big(\sum_{\boldsymbol{s}\in\tilde{\mathcal{S}}} \boldsymbol{B}_t[\cdot,\boldsymbol{s}]\Big)'\left(\boldsymbol{B}_t\boldsymbol{B}_t'\right)^{-1}\boldsymbol{P}_Z(t).$$

In the same token, the identification of $E(Y(t)|\boldsymbol{S}\in\tilde{\mathcal{S}})P(\boldsymbol{S}\in\tilde{\mathcal{S}})$ is given that:

$$\begin{aligned} E(Y(t)|\boldsymbol{S}\in\tilde{\mathcal{S}})P(\boldsymbol{S}\in\tilde{\mathcal{S}}) = \boldsymbol{b}(\tilde{\mathcal{S}})'\boldsymbol{B}_t^{+}\left(\boldsymbol{Q}_Z(t)\odot\boldsymbol{P}_Z(t)\right) &= \boldsymbol{b}(\tilde{\mathcal{S}})'\boldsymbol{B}_t'\left(\boldsymbol{B}_t\boldsymbol{B}_t'\right)^{-1}\left(\boldsymbol{Q}_Z(t)\odot\boldsymbol{P}_Z(t)\right) \\ &= \Big(\sum_{\boldsymbol{s}\in\tilde{\mathcal{S}}} \boldsymbol{B}_t[\cdot,\boldsymbol{s}]\Big)'\left(\boldsymbol{B}_t\boldsymbol{B}_t'\right)^{-1}\left(\boldsymbol{Q}_Z(t)\odot\boldsymbol{P}_Z(t)\right). \end{aligned}$$

A.2 Applying Theorem T.1 to LATE

Consider the LATE model where $T\in\{t_0,t_1\}$, $Z\in\{z_0,z_1\}$, and the monotonicity condition $\mathbf{1}[T_i(z_0)=t_1]\le\mathbf{1}[T_i(z_1)=t_1]\forall i$ holds. This model admits three types: never-takers $\boldsymbol{s}_{\text{nt}}=[t_0,t_0]'$, compliers $\boldsymbol{s}_{\text{c}}=[t_0,t_1]'$, and always-takers $\boldsymbol{s}_{\text{at}}=[t_1,t_1]'$. The corresponding response matrix $\boldsymbol{R}$ and the binary matrices $\boldsymbol{B}_{t_0}\equiv\mathbf{1}[\boldsymbol{R}=t_0]$, $\boldsymbol{B}_{t_1}\equiv\mathbf{1}[\boldsymbol{R}=t_1]$ are:

$$\boldsymbol{R} = \begin{array}{c} \begin{array}{ccc} \boldsymbol{s}_{nt} & \boldsymbol{s}_{c} & \boldsymbol{s}_{at} \end{array} \\ \begin{bmatrix} t_0 & t_0 & t_1 \\ t_0 & t_1 & t_1 \end{bmatrix} \end{array} \begin{array}{l} T(z_0) \\ T(z_1) \end{array} \quad \therefore \quad \boldsymbol{B}_{t_0} = \begin{array}{c} \begin{array}{ccc} \boldsymbol{s}_{nt} & \boldsymbol{s}_{c} & \boldsymbol{s}_{at} \end{array} \\ \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix} \end{array} \begin{array}{l} T_i(z_0) \\ T_i(z_1) \end{array}, \quad \boldsymbol{B}_{t_1} = \begin{array}{c} \begin{array}{ccc} \boldsymbol{s}_{nt} & \boldsymbol{s}_{c} & \boldsymbol{s}_{at} \end{array} \\ \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix} \end{array}. \tag{49}$$

It is useful to define the identification criteria $\boldsymbol{H}[t,\boldsymbol{s}]\equiv\boldsymbol{B}_t[\cdot,\boldsymbol{s}]'\left(\boldsymbol{B}_t\boldsymbol{B}_t'\right)^{-1}\boldsymbol{B}_t[\cdot,\boldsymbol{s}]$. According to Theorem T.1, $E(Y(t)|\boldsymbol{S}=\boldsymbol{s})$ is identified if $\boldsymbol{H}[t,\boldsymbol{s}]=1$. The following equation computes the identification criteria $\boldsymbol{H}[t_1,\boldsymbol{s}_c]$ for the treated compliers $E(Y(t_1)|\boldsymbol{S}=\boldsymbol{s}_c)$ of the LATE model:

$$\boldsymbol{H}[t_1,\boldsymbol{s}_c] = \boldsymbol{B}_{t_1}[\cdot,\boldsymbol{s}_c]'\left(\boldsymbol{B}_{t_1}\boldsymbol{B}_{t_1}'\right)^{-1}\boldsymbol{B}_{t_1}[\cdot,\boldsymbol{s}_c] = \begin{bmatrix}0 & 1\end{bmatrix}\left(\begin{bmatrix}0&0&1\\0&1&1\end{bmatrix}\begin{bmatrix}0&0\\0&1\\1&1\end{bmatrix}\right)^{-1}\begin{bmatrix}0\\1\end{bmatrix} = 1,$$

$\boldsymbol{H}[t_1,\boldsymbol{s}_c]=1$ means that $E(Y(t_1)|\boldsymbol{S}=\boldsymbol{s}_c)$ is identified, and, according to Theorem **T.1**, the identification equation for $E(Y(t_1)|\boldsymbol{S}=\boldsymbol{s}_c)$ is given by:

$$E(Y(t_1)|\boldsymbol{S}=\boldsymbol{s}_c) = \frac{\left[\boldsymbol{B}_{t_1}[\cdot,\boldsymbol{s}_c]'\left(\boldsymbol{B}_{t_1}\boldsymbol{B}_{t_1}'\right)^{-1}\right]\cdot\left(\boldsymbol{Q}_Z(t_1)\odot\boldsymbol{P}_Z(t_1)\right)}{\left[\boldsymbol{B}_{t_1}[\cdot,\boldsymbol{s}_c]'\left(\boldsymbol{B}_{t_1}\boldsymbol{B}_{t_1}'\right)^{-1}\right]\cdot\left(\boldsymbol{P}_Z(t_1)\right)} \tag{50}$$

$$= \frac{\begin{bmatrix}-1 & 1\end{bmatrix}\cdot\begin{pmatrix}E(Y|T=t_1,Z=z_0)P(T=t_1|Z=z_0)\\ E(Y|T=t_1,Z=z_1)P(T=t_1|Z=z_1)\end{pmatrix}}{\begin{bmatrix}-1 & 1\end{bmatrix}\begin{pmatrix}P(T=t_1|Z=z_0)\\ P(T=t_1|Z=z_1)\end{pmatrix}} \tag{51}$$

$$= \frac{E(Y\cdot D_{t_1}|Z=z_1)-E(Y\cdot D_{t_1}|Z=z_0)}{E(D_{t_1}|Z=z_1)-E(D_{t_1}|Z=z_0),}, \tag{52}$$

where $D_{t_1} \equiv \mathbf{1}[T=t_1]$. The parameter can be estimated by a 2SLS regression that uses Z to instrument the effect of the endogenous choice indicator D_{t_1} on the outcome variable $Y\cdot D_{t_1}$. The the following *Identification Matrix* displays the value of the identification criteria $\boldsymbol{H}[t,\boldsymbol{s}]$ for all $(t,\boldsymbol{s})\in\{t_0,t_1\}\times\{\boldsymbol{s}_n,\boldsymbol{s}_c,\boldsymbol{s}_1\}$ of the LATE model:

$$\textbf{LATE Identification Matrix} \quad \boldsymbol{H} \quad = \quad \begin{array}{cc} \begin{array}{ccc} \boldsymbol{s}_{nt} & \boldsymbol{s}_c & \boldsymbol{s}_{at} \end{array} & \\ \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix} & \begin{array}{c} t_0 \\ t_1 \end{array} \end{array} \tag{53}$$

The matrix indicates that the identification status of six counterfactual outcomes. Four counterfactual outcomes are identified: $E(Y(t_0)|\boldsymbol{S}=\boldsymbol{s}_c)$ and $E(Y(t_1)|\boldsymbol{S}=\boldsymbol{s}_c)$ for compliers, $E(Y(t_0)|\boldsymbol{S}=\boldsymbol{s}_{nt})$ for never-takes, and $E(Y(t_1)|\boldsymbol{S}=\boldsymbol{s}_a)$ for always-taker. Neither $E(Y(t_1)|\boldsymbol{S}=\boldsymbol{s}_{nt})$ or $E(Y(t_0)|\boldsymbol{S}=\boldsymbol{s}_{at})$ are identified, indeed, they are not even defined. The expression that identifies the counterfactual outcome for treated compliers, $E(Y(t_1)|\boldsymbol{S}=\boldsymbol{s}_c)$ is presented in equations (50)–(52). The remaining expressions according to Theorem **T.1** are displayed below:

$$E(Y(t_0)|\boldsymbol{S}=\boldsymbol{s}_{nt}) = \frac{\boldsymbol{B}_{t_0}[\cdot,\boldsymbol{s}_{nt}]'\left(\boldsymbol{B}_{t_0}\boldsymbol{B}_{t_0}'\right)^{-1}\cdot\left(\boldsymbol{Q}_Z(t_0)\odot\boldsymbol{P}_Z(t_0)\right)}{\boldsymbol{B}_{t_0}[\cdot,\boldsymbol{s}_{nt}]'\left(\boldsymbol{B}_{t_0}\boldsymbol{B}_{t_0}'\right)^{-1}\cdot\boldsymbol{P}_Z(t_0)} = \frac{E(Y\cdot D_{t_0}|Z=z_1)}{E(D_{t_0}|Z=z_1),},$$

$$E(Y(t_1)|\boldsymbol{S}=\boldsymbol{s}_{at}) = \frac{\boldsymbol{B}_{t_1}[\cdot,\boldsymbol{s}_{at}]'\left(\boldsymbol{B}_{t_1}\boldsymbol{B}_{t_1}'\right)^{-1}\cdot\left(\boldsymbol{Q}_Z(t_1)\odot\boldsymbol{P}_Z(t_1)\right)}{\boldsymbol{B}_{t_1}[\cdot,\boldsymbol{s}_{at}]'\left(\boldsymbol{B}_{t_1}\boldsymbol{B}_{t_1}'\right)^{-1}\cdot\boldsymbol{P}_Z(t_1)} = \frac{E(Y\cdot D_{t_1}|Z=z_0)}{E(D_{t_1}|Z=z_0),},$$

$$E(Y(t_0)|\boldsymbol{S}=\boldsymbol{s}_c) = \frac{\boldsymbol{B}_{t_0}[\cdot,\boldsymbol{s}_c]'\left(\boldsymbol{B}_{t_0}\boldsymbol{B}_{t_0}'\right)^{-1}\cdot\left(\boldsymbol{Q}_Z(t_0)\odot\boldsymbol{P}_Z(t_0)\right)}{\boldsymbol{B}_{t_0}[\cdot,\boldsymbol{s}_c]'\left(\boldsymbol{B}_{t_0}\boldsymbol{B}_{t_0}'\right)^{-1}\cdot\boldsymbol{P}_Z(t_0)} = \frac{E(Y\cdot D_{t_0}|Z=z_0)-E(Y\cdot D_{t_0}|Z=z_1)}{E(D_{t_0}|Z=z_0)-E(D_{t_0}|Z=z_1),}.$$

We can combine the identification equations for the treated and untreated compliers to obtain the well-known LATE expression:

$$E(Y(t_1)-Y(t_0)) = \frac{E(Y|Z=z_1)-E(Y|Z=z_0)}{P(T=t_1|Z=z_1)-P(T=t_0|Z=z_1)}.$$

Now consider the LATE model in which we relax the monotonicity condition. In this case, the response matrix and the corresponding identification matrix are:

$$\boldsymbol{R} = \begin{array}{c} \\ \\ \end{array}\begin{array}{cccc|c} \boldsymbol{s}_{nt} & \boldsymbol{s}_c & \boldsymbol{s}_{at} & \boldsymbol{s}_d & \\ \end{array}$$

$$\boldsymbol{R} = \begin{array}{c} \begin{array}{cccc} \boldsymbol{s}_{nt} & \boldsymbol{s}_c & \boldsymbol{s}_{at} & \boldsymbol{s}_d \end{array} \\ \begin{bmatrix} t_0 & t_0 & t_1 & t_1 \\ t_0 & t_1 & t_1 & t_0 \end{bmatrix} \end{array} \begin{array}{c} \\ T_i(z_0) \\ T_i(z_1) \end{array} \Rightarrow \boldsymbol{H} = \begin{array}{c} \begin{array}{cccc} \boldsymbol{s}_{nt} & \boldsymbol{s}_c & \boldsymbol{s}_{at} & \boldsymbol{s}_d \end{array} \\ \begin{bmatrix} 2/3 & 2/3 & 0 & 2/3 \\ 0 & 2/3 & 2/3 & 2/3 \end{bmatrix} \end{array} \begin{array}{c} \\ t_0 \\ t_1 \end{array}$$

Note that none of the elements of the identification matrix are equal to one, which indicates that there are no point-identified counterfactuals when the monotonicity condition is relaxed.

A.3 Using Revealed Preference Analysis to Generate Choice Restrictions

Our choice model stems from a classical economic framework where the potential choice of agent i for a fixed IV-value z is characterized by the following utility maximization problem:

$$\text{Choice Equation :} \quad T_i(z) = \text{argmax}_{t \in \mathcal{T}} \left(\max_{\boldsymbol{g} \in \mathcal{B}_i(Z_i,t)} u_i(t, \boldsymbol{g}) \right). \tag{54}$$

The real-valued utility function $u_i : \mathcal{T} \times \mathbb{R}^K_+$ represents the rational preferences of agent i towards the bundle $(t, \boldsymbol{g})$ where t is the treatment status and $\boldsymbol{g}$ is a K-dimensional vector of unobserved consumption goods. The set $\mathcal{B}_i(z,t) \subset \mathbb{R}^K_+$ stands for the potential budget set of consumption goods $\boldsymbol{g}$ of agent i when the treatment is *fixed* to $t \in \mathcal{T}$ and the instrument is *fixed* to the value $z \in \mathcal{Z}$. The budget set is broadly interpreted to encompass various decisions extending beyond traditional consumption goods. It can include decisions regarding education attainment, neighborhood selection, and time allocation depending on the empirical setting under examination.

The incentive matrix $\boldsymbol{L}$ characterises budget set relationships in which bigger incentives correspond to larger budget sets for a given a choice t :

$$\text{Budget Relationships:} \quad \boldsymbol{L}[z,t] \leq \boldsymbol{L}[z',t] \Rightarrow \mathcal{B}_i(z,t) \subseteq \mathcal{B}_i(z',t). \tag{55}$$

To put in context, consider the LATE model where $T = t_1$ denotes college enrollment and $T = t_0$ denotes no college. $Z \in$ is a randomly assigned tuition discount, z_1 if the discount is granted and z_0 if not. The LATE incentive matrix yields the following budget set relations:

$$\underbrace{\boldsymbol{L} = \begin{array}{c} \begin{array}{cc} t_0 & t_1 \end{array} \\ \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \end{array} \begin{array}{c} \\ z_0 \\ z_1 \end{array}}_{\text{LATE Incentive Matrix}} \Rightarrow \underbrace{\begin{array}{c} \mathcal{B}_i(z_0,t_0) = \mathcal{B}_i(z_1,t_0) \\ \mathcal{B}_i(z_0,t_1) \subset \mathcal{B}_i(z_1,t_1) \end{array}}_{\text{Implied Budget Set Relationships}} \tag{56}$$

The budget set equality $\mathcal{B}_i(z_0,t_0) = \mathcal{B}_i(z_1,t_0)$ implies that when the choice is set to no college t_0, the tuition discount is irrelevant. Conversely, $\mathcal{B}_i(z_0,t_1) \subset \mathcal{B}_i(z_1,t_1)$ suggests that the tuition discount increases agent i's budget if they choose to attend college.

Budget relationships enable us to use the Weak Axiom of Revealed Preference (WARP) of Richter (1971). Bundles $(t, \boldsymbol{g})$ and $(t', \boldsymbol{g}')$ are said to be available given z if $\boldsymbol{g} \in \mathcal{B}_i(z,t)$ and $\boldsymbol{g}' \in \mathcal{B}_i(z,t')$. If bundle $(t, \boldsymbol{g})$ is chosen by agent i when $(t, \boldsymbol{g})$ and $(t', \boldsymbol{g}')$ are available, then $(t, \boldsymbol{g})$ is said to be directly and strictly revealed preferred to $(t', \boldsymbol{g}')$, that is, $(t, \boldsymbol{g}) \succ^d_{i,z} (t', \boldsymbol{g}')$. WARP states that if $(t, \boldsymbol{g})$ revealed preferred to $(t', \boldsymbol{g}')$ under $z \in \mathcal{Z}$, then $(t', \boldsymbol{g}')$ cannot be revealed preferred to $(t, \boldsymbol{g})$ under $z' \in \mathcal{Z} \setminus \{z\}$. Notationally, we write that:

$$\text{WARP:} \quad (t, \boldsymbol{g}) \succ^d_{i,z} (t', \boldsymbol{g}') \Rightarrow (t', \boldsymbol{g}') \not\succ^d_{i,z'} (t, \boldsymbol{g}). \tag{57}$$

The following lemma uses WARP and the budget relations (55) to translate incentives into choice restrictions.

Lemma L.2. Let a choice model with an incentive matrix $\boldsymbol{L}$. Under the budget relationships (55) and WARP (57), the following choice rule holds:

$$\text{WARP Rule:} \quad \text{If } T_i(z) = t, \text{ and } \boldsymbol{L}[z', t'] - \boldsymbol{L}[z, t'] \leq 0 \leq \boldsymbol{L}[z', t] - \boldsymbol{L}[z, t] \text{ then } T_i(z') \neq t'. \tag{58}$$

Proof. $T_i(z) = t$ implies that there exists $\boldsymbol{g} \in \mathcal{B}_i(z, t)$ such that $(t, \boldsymbol{g}) \succ^d_{i,z} (t', \boldsymbol{g}')$ for all $\boldsymbol{g}' \in \mathcal{B}_i(z, t')$. The inequality $0 \leq \boldsymbol{L}[z', t] - \boldsymbol{L}[z, t]$ implies that the budget set associated with t increases as we move from z to z', $\mathcal{B}_i(z, t) \subseteq \mathcal{B}_i(z', t)$. Thus the bundle $(t, \boldsymbol{g})$ remains available under z'. On the other hand, $\boldsymbol{L}[z', t'] - \boldsymbol{L}[z, t'] \leq 0$ implies that the budget set associated with t' decreases as we move from z to z', $\mathcal{B}_i(z', t') \subseteq \mathcal{B}_i(z, t')$. Thus any bundle $(t', \boldsymbol{g}'')$ that is available under z', $(t', \boldsymbol{g}''); g'' \in \mathcal{B}_i(z', t')$ were also available under z. Thus, according to WARP, agent i still prefers $(t, \boldsymbol{g})$ to any $(t', \boldsymbol{g}''); g'' \in \mathcal{B}_i(z', t')$, that is, $(t', \boldsymbol{g}'') \not\succ^d_{i,z'} (t, \boldsymbol{g})$. This implies that agent i does not choose t' under z', $T_i(z') \neq t'$. □

As mentioned in the main paper, applying the WARP Rule 58 to LATE incentives (56) yields the choice restriction $T_i(z_0) = t_1 \Rightarrow T_i(z_1) \neq t_0$, which means that if the student chooses college under no incentives, it will not choose otherwise when incentives to enroll in college are offered.

It is possible to exploit additional economic choice behaviors that enable us to enhance the WARP rule. For instance, consider the choice of a college student who debates between two majors: electrical or mechanical engineering. Suppose the student chooses electrical over mechanical engineering under no tuition discount. In that case, it is natural to assume that the student will maintain choice when granted a tuition discount that applies to both majors. This behavior is captured by the condition called *Normal Choice*:

$$\text{Normal Choice: } t \succ_{i,z} t' \text{ and } \boldsymbol{L}[z', t'] - \boldsymbol{L}[z, t'] = \boldsymbol{L}[z', t] - \boldsymbol{L}[z, t] \text{ then } t \succ_{i,z'} t' \text{ holds,} \tag{59}$$

where $t \succ_{i,z} t'$ means that there is $\boldsymbol{g} \in \mathcal{B}_i(z, t)$ such that $(t, \boldsymbol{g}) \succ_{i,z} (t', \boldsymbol{g}')$ for all $\boldsymbol{g}' \in \mathcal{B}_i(z, t')$. Normal Choice states that if an agent i prefers t instead of t' under z, and if the change in incentives for choosing either t or t' is the same under z', then agent i maintains its preference of t over t' under z'.[27] WARP and Normal Choice (59) yield the following choice rule:

[27] Normal Choice is a no-crossing condition on the ranking of choice preferences that maintains the relative rank of two choices that share the same incentives. The normal choice is related to the notion of normal goods. Consider an agent that debates between two goods a and b. Suppose a discount of d dollars is applied to both goods. This discount can be understood as an increase in income of d dollars since the agent will benefit from it regardless of his choice. An increase in income does not decrease the consumption of a normal good. If the agent decides to buy a under no discount, it will continue to consume one unit of good a when the discount is available.

Lemma L.3. Let a choice model with an incentive matrix $\boldsymbol{L}$. Under the budget relationships (55), WARP (57), and Normal Choice (59), the following choice rule holds:

$$\text{Choice Rule:} \quad \text{If } T_i(z) = t \text{ and } \boldsymbol{L}[z', t'] - \boldsymbol{L}[z, t'] \leq \boldsymbol{L}[z', t] - \boldsymbol{L}[z, t] \text{ then } T_i(z') \neq t'. \tag{60}$$

Proof. Let $\delta_t \equiv \boldsymbol{L}[z', t] - \boldsymbol{L}[z, t]$ and $\delta_{t'} \equiv \boldsymbol{L}[z', t'] - \boldsymbol{L}[z, t']$, where $\delta_{t'} \leq \delta_t$. Note that we can set $\boldsymbol{L}[z, t] = \boldsymbol{L}[z, t'] = 0$, $\boldsymbol{L}[z', t] = \delta_t$, and $\boldsymbol{L}[z', t'] = \delta_{t'}$ without loss of generality. Note also that $T_i(z) = t$ means that $t \succ_{i,z} t'$. Now consider an auxiliary instrument z^* that sets $\boldsymbol{L}[z^*, t] = \boldsymbol{L}[z^*, t'] = \delta_{t'}$. We first examine the change from z to z^*. In this case, we have that $\boldsymbol{L}[z^*, t'] - \boldsymbol{L}[z, t'] = \boldsymbol{L}[z^*, t] - \boldsymbol{L}[z, t] = \delta_{t'}$. According to Normal Choice (59), we have that $t \succ_{i,z^*} t'$. Now consider the change from z^* to z'. The inequality $\delta_{t'} \leq \delta_t$ implies that: $\boldsymbol{L}[z', t'] - \boldsymbol{L}[z^*, t'] = 0 \leq \boldsymbol{L}[z', t] - \boldsymbol{L}[z^*, t]$. By WARP Rule (58), we have that $t \succ_{i,z'} t'$ and therefore $T_i(z') \neq t'$. $\square$

As mentioned, the Choice Rule highlights a cornerstone principle of rational choice theory, which posits that an individual's preferences will remain consistent unless there is a compelling incentive to choose otherwise. Specifically, if an agent chooses t over t' when presented with z-incentives, and if z'-incentives are at least as persuasive for choice t as they are for t', then the agent will not choose t' over t.

A.4 Additional Analyses of the IV Model in Example E.3

The incentive matrix of Example **E.3** is:

$$\boldsymbol{L} = \begin{array}{c} \\ \\ \\ \\ \end{array} \begin{array}{c} \begin{array}{ccc} t_0 & t_1 & t_2 \end{array} \\ \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \end{array} \begin{array}{c} \\ z_0 \\ z_1 \\ z_2 \end{array} \tag{61}$$

The incentive matrix (61) justifies two monotonicity conditions:

$$\mathbf{1}[T_i(z_0) = t_1] \leq \mathbf{1}[T_i(z_1) = t_1] \tag{62}$$

$$\mathbf{1}[T_i(z_0) = t_2] \leq \mathbf{1}[T_i(z_2) = t_2]. \tag{63}$$

These monotonicity conditions eliminate 12 out of the 27 possible response types as described in Panel B of Table A.1.

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

$$\boldsymbol{R} = \begin{array}{c} \begin{array}{ccccccccccccccc} \boldsymbol{s}_1 & \boldsymbol{s}_2 & \boldsymbol{s}_3 & \boldsymbol{s}_4 & \boldsymbol{s}_5 & \boldsymbol{s}_6 & \boldsymbol{s}_7 & \boldsymbol{s}_8 & \boldsymbol{s}_9 & \boldsymbol{s}_{10} & \boldsymbol{s}_{11} & \boldsymbol{s}_{12} & \boldsymbol{s}_{13} & \boldsymbol{s}_{14} & \boldsymbol{s}_{15} \end{array} \\ \begin{bmatrix} t_0 & t_0 & t_0 & t_0 & t_0 & t_0 & t_0 & t_0 & t_0 & t_2 & t_1 & t_1 & t_2 & t_2 & t_2 \\ t_0 & t_0 & t_0 & t_1 & t_1 & t_1 & t_2 & t_2 & t_2 & t_2 & t_1 & t_1 & t_0 & t_1 & t_2 \\ t_0 & t_1 & t_2 & t_0 & t_1 & t_2 & t_0 & t_1 & t_2 & t_0 & t_1 & t_2 & t_2 & t_2 & t_2 \end{bmatrix} \end{array} \begin{array}{c} \\ T_i(z_0) \\ T_i(z_1) \\ T_i(z_2) \end{array} \tag{64}$$

The response matrix is then used as input to generate the identification matrix $\boldsymbol{H}$, which is a $N_{\times} N_T$-dimensional matrix whose elements are given by $\boldsymbol{H}[t, \boldsymbol{s}] \equiv \boldsymbol{B}_t[\cdot, \boldsymbol{s}]' \left(\boldsymbol{B}_t \boldsymbol{B}_t'\right)^{-1} \boldsymbol{B}_t[\cdot, \boldsymbol{s}]$.

Table A.1: Elimination of Response types of the IV Model in Example **E.3**

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
$T_1(z_0)$	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
$T_1(z_1)$	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
$T_1(z_2)$	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 (62)–(63)**																										
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 $S_1 = [T_1(z_0)\,,\, T_1(z_1)\,,\, T_1(z_2)]$ 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 $\text{supp}(T) = \{t_0, t_1, t_2\}$. Panel B describes an elimination process based on the two monotonicity conditions (62)–(63). 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.

$$
\begin{array}{cc}
 & \begin{array}{ccccccccccccccc} \boldsymbol{s}_1 & \boldsymbol{s}_2 & \boldsymbol{s}_3 & \boldsymbol{s}_4 & \boldsymbol{s}_5 & \boldsymbol{s}_6 & \boldsymbol{s}_7 & \boldsymbol{s}_8 & \boldsymbol{s}_9 & \boldsymbol{s}_{10} & \boldsymbol{s}_{11} & \boldsymbol{s}_{12} & \boldsymbol{s}_{13} & \boldsymbol{s}_{14} & \boldsymbol{s}_{15} \end{array} \\
\boldsymbol{H} = & \begin{bmatrix} 11/27 & 8/27 & 8/27 & 8/27 & 5/27 & 5/27 & 8/27 & 5/27 & 5/27 & 1/3 & 0 & 0 & 1/3 & 0 & 0 \\ 0 & 4/13 & 0 & 7/26 & 11/26 & 7/26 & 0 & 4/13 & 0 & 0 & 15/26 & 15/26 & 0 & 7/26 & 0 \\ 0 & 0 & 16/75 & 0 & 0 & 16/75 & 19/75 & 19/75 & 31/75 & 34/75 & 0 & 16/75 & 1/3 & 1/3 & 8/25 \end{bmatrix}
\end{array}
$$

Note that none of the entries of the identification matric is equal to one. According to **T.1**, this means that not a single counterfactual outcome of the type $E(Y(t)|\boldsymbol{S}\boldsymbol{s}); (t, \boldsymbol{s}) \in \{t_0, t_1, t_2\} \times \{\boldsymbol{s}_1, ..., \boldsymbol{s}_{15}\}$ is identified. We conclude that the elimination of response type due to the monotonicity conditions (62)–(63) is not sufficient to point-identify counterfactual outcome.

Revealed preference analysis is more effective in eliminating types than the monotonicity conditions. Table A.2 applies choice rule (9) to the incentive matrix (61). There are 22 binding restrictions. Table A.3 summarise these 22 choice restrictions of Table A.2 into the five restrictions,[28] and Panel C of Table (A.1) shows that these five restrictions eliminate 19 out of the 27 possible response types.

Table A.2: Choice Restrictions of Example **E.3** Due to Revealed Preference Analysis

#	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	$T_i(z_0) = t_0,$	$\boldsymbol{L}[z_2,t_1] - \boldsymbol{L}[z_0,t_1] = 0$	$\le$	$0 = \boldsymbol{L}[z_2,t_0] - \boldsymbol{L}[z_0,t_0]$	$T_i(z_2) \neq t_1$
2	$T_i(z_0) = t_0,$	$\boldsymbol{L}[z_1,t_2] - \boldsymbol{L}[z_0,t_2] = 0$	$\le$	$0 = \boldsymbol{L}[z_1,t_0] - \boldsymbol{L}[z_0,t_0]$	$T_i(z_1) \neq t_2$
3	$T_i(z_0) = t_1,$	$\boldsymbol{L}[z_1,t_0] - \boldsymbol{L}[z_0,t_0] = 0$	$\le$	$1 = \boldsymbol{L}[z_1,t_1] - \boldsymbol{L}[z_0,t_1]$	$T_i(z_1) \neq t_0$
4	$T_i(z_0) = t_1,$	$\boldsymbol{L}[z_2,t_0] - \boldsymbol{L}[z_0,t_0] = 0$	$\le$	$0 = \boldsymbol{L}[z_2,t_1] - \boldsymbol{L}[z_0,t_1]$	$T_i(z_2) \neq t_0$
5	$T_i(z_0) = t_1,$	$\boldsymbol{L}[z_1,t_2] - \boldsymbol{L}[z_0,t_2] = 0$	$\le$	$1 = \boldsymbol{L}[z_1,t_1] - \boldsymbol{L}[z_0,t_1]$	$T_i(z_1) \neq t_2$
6	$T_i(z_0) = t_2,$	$\boldsymbol{L}[z_1,t_0] - \boldsymbol{L}[z_0,t_0] = 0$	$\le$	$0 = \boldsymbol{L}[z_1,t_2] - \boldsymbol{L}[z_0,t_2]$	$T_i(z_1) \neq t_0$
7	$T_i(z_0) = t_2,$	$\boldsymbol{L}[z_2,t_0] - \boldsymbol{L}[z_0,t_0] = 0$	$\le$	$1 = \boldsymbol{L}[z_2,t_2] - \boldsymbol{L}[z_0,t_2]$	$T_i(z_2) \neq t_0$
8	$T_i(z_0) = t_2,$	$\boldsymbol{L}[z_2,t_1] - \boldsymbol{L}[z_0,t_1] = 0$	$\le$	$1 = \boldsymbol{L}[z_2,t_2] - \boldsymbol{L}[z_0,t_2]$	$T_i(z_2) \neq t_1$
9	$T_i(z_1) = t_0,$	$\boldsymbol{L}[z_0,t_1] - \boldsymbol{L}[z_1,t_1] = -1$	$\le$	$0 = \boldsymbol{L}[z_0,t_0] - \boldsymbol{L}[z_1,t_0]$	$T_i(z_0) \neq t_1$
10	$T_i(z_1) = t_0,$	$\boldsymbol{L}[z_2,t_1] - \boldsymbol{L}[z_1,t_1] = -1$	$\le$	$0 = \boldsymbol{L}[z_2,t_0] - \boldsymbol{L}[z_1,t_0]$	$T_i(z_2) \neq t_1$
11	$T_i(z_1) = t_0,$	$\boldsymbol{L}[z_0,t_2] - \boldsymbol{L}[z_1,t_2] = 0$	$\le$	$0 = \boldsymbol{L}[z_0,t_0] - \boldsymbol{L}[z_1,t_0]$	$T_i(z_0) \neq t_2$
12	$T_i(z_1) = t_2,$	$\boldsymbol{L}[z_0,t_0] - \boldsymbol{L}[z_1,t_0] = 0$	$\le$	$0 = \boldsymbol{L}[z_0,t_2] - \boldsymbol{L}[z_1,t_2]$	$T_i(z_0) \neq t_0$
13	$T_i(z_1) = t_2,$	$\boldsymbol{L}[z_2,t_0] - \boldsymbol{L}[z_1,t_0] = 0$	$\le$	$1 = \boldsymbol{L}[z_2,t_2] - \boldsymbol{L}[z_1,t_2]$	$T_i(z_2) \neq t_0$
14	$T_i(z_1) = t_2,$	$\boldsymbol{L}[z_0,t_1] - \boldsymbol{L}[z_1,t_1] = -1$	$\le$	$0 = \boldsymbol{L}[z_0,t_2] - \boldsymbol{L}[z_1,t_2]$	$T_i(z_0) \neq t_1$
15	$T_i(z_1) = t_2,$	$\boldsymbol{L}[z_2,t_1] - \boldsymbol{L}[z_1,t_1] = -1$	$\le$	$1 = \boldsymbol{L}[z_2,t_2] - \boldsymbol{L}[z_1,t_2]$	$T_i(z_2) \neq t_1$
16	$T_i(z_2) = t_0,$	$\boldsymbol{L}[z_0,t_1] - \boldsymbol{L}[z_2,t_1] = 0$	$\le$	$0 = \boldsymbol{L}[z_0,t_0] - \boldsymbol{L}[z_2,t_0]$	$T_i(z_0) \neq t_1$
17	$T_i(z_2) = t_0,$	$\boldsymbol{L}[z_0,t_2] - \boldsymbol{L}[z_2,t_2] = -1$	$\le$	$0 = \boldsymbol{L}[z_0,t_0] - \boldsymbol{L}[z_2,t_0]$	$T_i(z_0) \neq t_2$
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	$T_i(z_2) = t_1,$	$\boldsymbol{L}[z_0,t_0] - \boldsymbol{L}[z_2,t_0] = 0$	$\le$	$0 = \boldsymbol{L}[z_0,t_1] - \boldsymbol{L}[z_2,t_1]$	$T_i(z_0) \neq t_0$
20	$T_i(z_2) = t_1,$	$\boldsymbol{L}[z_1,t_0] - \boldsymbol{L}[z_2,t_0] = 0$	$\le$	$1 = \boldsymbol{L}[z_1,t_1] - \boldsymbol{L}[z_2,t_1]$	$T_i(z_1) \neq t_0$
21	$T_i(z_2) = t_1,$	$\boldsymbol{L}[z_0,t_2] - \boldsymbol{L}[z_2,t_2] = -1$	$\le$	$0 = \boldsymbol{L}[z_0,t_1] - \boldsymbol{L}[z_2,t_1]$	$T_i(z_0) \neq t_2$
22	$T_i(z_2) = t_1,$	$\boldsymbol{L}[z_1,t_2] - \boldsymbol{L}[z_2,t_2] = -1$	$\le$	$1 = \boldsymbol{L}[z_1,t_1] - \boldsymbol{L}[z_2,t_1]$	$T_i(z_1) \neq t_2$

This table displays the binding choice restrictions generated by choice rule (9) to the incentive matrix of Example **E.3**.

[28]The remaining restrictions do not eliminate any additional response types that is not already covered by these five restrictions.

Table A.3: Summary of Choice Restrictions generated by applying Choice Rule (9) to Example **E.3**

#	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 $T_i(z_1) = t_1$

The eight response types that survive the elimination process are displayed in the response matrix below:

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

The corresponding identification matrix is given by:

$$\begin{array}{cc} & \begin{array}{cccccccc} \boldsymbol{s}_1 & \boldsymbol{s}_2 & \boldsymbol{s}_3 & \boldsymbol{s}_4 & \boldsymbol{s}_5 & \boldsymbol{s}_6 & \boldsymbol{s}_7 & \boldsymbol{s}_8 \end{array} \\ \boldsymbol{H} = & \begin{bmatrix} 3/4 & 3/4 & 3/4 & 3/4 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1/3 & 1/3 & 1 & 1 & 1/3 & 0 \\ 0 & 1/3 & 0 & 1/3 & 0 & 1/3 & 1 & 1 \end{bmatrix} \begin{array}{l} t_0 \\ t_1 \\ t_2 \end{array} \end{array}$$

The entries of the identification matrix show that four counterfactual outcomes are point-identified, namely, $E(Y(t_1)|\boldsymbol{S} = \boldsymbol{s}_5)$, $E(Y(t_1)|\boldsymbol{S} = \boldsymbol{s}_6)$, $E(Y(t_2)|\boldsymbol{S} = \boldsymbol{s}_7)$, and $E(Y(t_2)|\boldsymbol{S} = \boldsymbol{s}_8)$.

A.5 Causal Interpretation of Angrist and Imbens (1995) Monotonicity

The main paper shows that response matrix above satisfies the monotonicity condition of Angrist and Imbens (1995). A celebrated result of Angrist and Imbens (1995) is that the monotonicity condition delivers a causal interpretation to standard 2SLS estimates. The LATE parameter that compares two IV-values z, z' evaluates a weighted average of the per-unit treatment effect among the compliers that change their choice as the instrument shifts from z to z'.

The general formula for the LATE parameter that compares any two IV-values z, z' where $T_i(z) \leq T_i(z')$ is:

$$LATE(z, z') = \frac{E(Y|Z = z') - E(Y|Z = z)}{E(T|Z = z') - E(T|Z = z)} = \sum_{t<t'} E(Y(t') - Y(t)|\boldsymbol{S} \in \mathcal{S}_{t'}(z') \cap \mathcal{S}_t(z))\omega_{t,t'},$$

$$\text{where } \omega_{t,t'} = \frac{P(\boldsymbol{S} \in \mathcal{S}_{t'}(z') \cap \mathcal{S}_t(z))}{\sum_{t<t'} (t' - t) \cdot P(\boldsymbol{S} \in \mathcal{S}_{t'}(z') \cap \mathcal{S}_t(z))}, \text{ and } \mathcal{S}_t(z) = \{\boldsymbol{s} \in \mathcal{S}; \boldsymbol{s}[z] = t\}.$$

The set $\mathcal{S}_t(z)$ comprise the response-types that takes value t when the instrument is set to z. Thus

$\mathcal{S}_{t'}(z') \cap \mathcal{S}_t(z)$ is the set of response types that take value t under z and t' under z'. The weights $\omega_{t,t'}$ are positive, but do not necessarily sum to one.

The LATE parameter corresponding to IV-values z_0, z_1 in the choice model given by response matrix (20) is:

$$
\begin{aligned}
LATE(z_1, z_0) &\equiv \frac{E(Y|Z = z_0) - E(Y|Z = z_1)}{E(T|Z = z_0) - E(T|Z = z_1)} \\
&= \frac{E(Y(t_2) - Y(t_1)|\boldsymbol{S} \in \{\boldsymbol{s}_4, \boldsymbol{s}_6\})P(\boldsymbol{S} \in \{\boldsymbol{s}_4, \boldsymbol{s}_6\}) + E(Y(t_3) - Y(t_1)|\boldsymbol{S} = \boldsymbol{s}_8)P(\boldsymbol{S} = \boldsymbol{s}_8)}{(t_2 - t_1) \cdot P(\boldsymbol{S} \in \{\boldsymbol{s}_4, \boldsymbol{s}_6\}) + (t_3 - t_1) \cdot P(\boldsymbol{S} = \boldsymbol{s}_8)}.
\end{aligned}
$$

If the treatment were to represent schooling years, then the LATE parameter can be interpreted as a weighted average of the causal effect of one additional year of education on the response types $(\boldsymbol{s}_4, \boldsymbol{s}_6, \boldsymbol{s}_8)$, which comprise the agents who alter their schooling choice as the instrument shifts.

A.6 Proof of Theorem T.2

The main statement of the theorem is that, under the choice rule (9), supermodular incentives imply and is implied by OMC.

We first prove that, given choice rule (9), supermodular incentives imply OMC.

Proof. Consider a sequence of IV-values $z_1, ..., z_{N_Z}$ and a sequence of treatment choices $t_1, ...t_{N_T}$ for which supermodularity holds. We first examine the choices t_{j+1} versus t_j for an IV-change from z_k to z_{k+1}. Let $T_i(z_k) = t_{j+1}$, under supermodular incentives, we have that $\boldsymbol{L}[z_{k+1}, t_j] - \boldsymbol{L}[z_k, t_j] \leq \boldsymbol{L}[z_{k+1}, t_{j+1}] - \boldsymbol{L}[z_k, t_{j+1}]$. Thus, according to the choice rule (9), $T_i(z_k) \neq t_j$. In summary, we have that $T_i(z_k) = t_{j+1} \Rightarrow T_i(z_{k+1}) \neq t_j$. We can extend this rationale to compare choice t_{j+1} versus t_κ for $\kappa = 1, ..., j$. This analysis generates the following choice restrictions: $T_i(z_k) = t_{j+1} \Rightarrow T_i(z_{k+1}) \notin \{t_1, ..., t_j\}$ for all $j = 1, ..., N_T - 1$. Otherwise states, we have that $T_i(z_k) = t_{j+1} \Rightarrow T_i(z_{k+1}) \in \{t_{j+1}, ..., t_{N_T}\}$ for all $j = 1, ..., N_T - 1$. This statement is equivalent to OMC, which states that higher ranks of z-values of the counterfactual choice $T_i(z)$ correspond to higher ranks of treatment choices. In particular, for treatment values $t_1 < ... < t_{N_T}$, the choice restriction implies $T_i(z_k) \leq T_i(z_{k+1})$. □

Next, we show that for OMC to hold, incentives must be supermodular.

Proof. Consider the IV-values z_k, z_{k+1} and the treatment values t_j, t_{j+1}, t_{j+2}. Let the incentive difference for choice t_j between z_{k+1} and z_{k+1} be denoted as $\Delta_j = \boldsymbol{L}[z_{k+1}, t_j] - \boldsymbol{L}[z_k, t_j]$ □

A.7 Proof of Theorem T.3

We seek to prove that, if the incentive matrix $\boldsymbol{L}$ is binary, then monotonic incentives imply UMC. Recall that UMC holds if and only if no 2×2 submatrix in $\boldsymbol{R}$ that exhibits the prohibit pattern which displays a choice t in one of its diagonals while displays no t in the other diagonal. Specifically, no 2×2 submatrix in $\boldsymbol{R}$ can take the form:

$$\begin{array}{ccl} \boldsymbol{s} & \boldsymbol{s}' & \\ \multicolumn{2}{c}{\begin{bmatrix} t & t'' \\ t' & t \end{bmatrix}} & \begin{array}{l} T_i(z) \\ T_i(z') \end{array} \end{array}, \tag{65}$$

where $t, t', t'' \in \mathcal{T}$ and $z, z' \in \mathcal{Z}$. It is useful to investigate how the prohibit patter (65) in light of the choice rule (9). Let $\boldsymbol{L}'$ be the 2×3 submatrix of the binary incentive matrix $\boldsymbol{L}$ corresponding to rows z, z' and columns t, t', t''.

Consider the first type in (65) $\boldsymbol{s} = [t, t']'$. For $\boldsymbol{s}$ to arise it must be the case that $T_i(z) = t \nRightarrow T_i(z') \neq t'$ According to the choice rule (9), this lack of choice restriction can only arise when:[29]

$$\boldsymbol{L}[z', t'] - \boldsymbol{L}[z, t'] > \boldsymbol{L}[z', t] - \boldsymbol{L}[z, t]. \tag{66}$$

Given that the incentive matrix is binary, this must the be case that:

1. $\boldsymbol{L}[z', t'] > \boldsymbol{L}[z, t']$ and $\boldsymbol{L}[z', t] \leq \boldsymbol{L}[z, t]$; or
2. $\boldsymbol{L}[z', t'] \geq \boldsymbol{L}[z, t']$ and $\boldsymbol{L}[z', t] < \boldsymbol{L}[z, t]$

Now consider the second type in (65) $\boldsymbol{s}; = [t'', t]'$. For $\boldsymbol{s}'$ to arise it must be the case that $T_i(z') = t \nRightarrow T_i(z) \neq t''$. This lack of choice restriction can only arise when

$$\boldsymbol{L}[z, t''] - \boldsymbol{L}[z', t''] > \boldsymbol{L}[z, t] - \boldsymbol{L}[z', t]. \tag{67}$$

Given that the incentive matrix is binary, this must the be case that:

1. $\boldsymbol{L}[z, t''] > \boldsymbol{L}[z', t'']$ and $\boldsymbol{L}[z, t] \leq \boldsymbol{L}[z', t]$; or
2. $\boldsymbol{L}[z, t''] = \boldsymbol{L}[z', t'']$ and $\boldsymbol{L}[z, t] < \boldsymbol{L}[z', t]$

It is clear that the only possibility to generate the prohibit parte in by combining the first item of the two lists, namely,

$$\boldsymbol{L}[z', t'] > \boldsymbol{L}[z, t'], \quad \boldsymbol{L}[z, t''] > \boldsymbol{L}[z', t''], \text{ and } \boldsymbol{L}[z, t] = \boldsymbol{L}[z', t].$$

In other words, the prohibit pattern requires the following pattern of incentives:

1. Incentives for t must be equal $\boldsymbol{L}[z, t] = \boldsymbol{L}[z', t]$
2. Incentives for t' must increase as Z changes from z to $z' : \boldsymbol{L}[z, t'] < \boldsymbol{L}[z', t']$.
3. Incentives for t'' must decrease as Z changes from z to $z' : \boldsymbol{L}[z, t''] > \boldsymbol{L}[z', t'']$.

The pattern of incentives for t and t' violate the monotonic incentive condition, which proves the theorem.

[29] Alternatively, one can state that the type only arises when $T_i(z') = t' \nRightarrow T_i(z) \neq t$. According to the choice rule (9), this lack of choice restriction can only arise when $\boldsymbol{L}[z, t] - \boldsymbol{L}[z', t] > \boldsymbol{L}[z, t'] - \boldsymbol{L}[z', t']$. It turns out that the incentive relationship above is equivalent to the incentive relationship in (67).

A.8 Proof of Theorem T.4

We first seek to prove that t-monotonic incentives (30) implies the monotonicity condition (29). To do so, it suffices to prove that t-monotonic incentives prevents the advent of the prohibit pattern in the response matrix $\boldsymbol{R}$. Specifically, no 2×2 submatrix in $\boldsymbol{R}$ can take the form:

$$\begin{array}{cc} \boldsymbol{s} & \boldsymbol{s}' \end{array} \qquad \begin{bmatrix} t & t'' \\ t' & t \end{bmatrix} \begin{array}{c} T_i(z) \\ T_i(z') \end{array}, \tag{68}$$

where $t, t', t'' \in \mathcal{T}$ and $z, z' \in \mathcal{Z}$. Consider the IV-values $z, z' \in \mathcal{Z}$. If t-monotonic incentives hold, there are two cases to consider.

The first case consists the instance where

$$\boldsymbol{L}[z', t] - \boldsymbol{L}[z, t] \leq \boldsymbol{L}[z', t'] - \boldsymbol{L}[z, t'] \,\forall\, t' \in \mathcal{T} \setminus \{t\}$$

holds. According to the choice rule (9), it must be the case that $T_i(z) = t \Rightarrow T_i(z') \neq t' \forall t' \in \mathcal{T} \setminus \{t\}$, which is equivalent to state that $T_i(z) = t \Rightarrow T_i(z') = t$ which prevents the prohibit pattern.

The second case is where the following condition holds:

$$\boldsymbol{L}[z', t] - \boldsymbol{L}[z, t] \geq \boldsymbol{L}[z', t'] - \boldsymbol{L}[z, t'] \,\forall\, t' \in \mathcal{T} \setminus \{t\}$$

This condition can be equivalently stated as:

$$\boldsymbol{L}[z, t] - \boldsymbol{L}[z', t] \leq \boldsymbol{L}[z, t'] - \boldsymbol{L}[z', t'] \,\forall\, t' \in \mathcal{T} \setminus \{t\}.$$

Applying the same rationale of the first case, we have that $T_i(z') = t \Rightarrow T_i(z) = t$ which also prevents the prohibit pattern.

Next we seek to prove that if the monotonicity condition (29) holds, than t-monotonic incentives (30) must be satisfied. For the monotonicity condition (29) to hold, the prohibit pattern (68) cannot occur. The prohibit pattern requires two conditions to occur:

1. $T_i(z) = t$ must not imply $T_i(z') = t'$ for some $t' \in \mathcal{T} \setminus \{t\}$; and
2. $T_i(z') = t$ must not imply $T_i(z) = t''$ for some $t'' \in \mathcal{T} \setminus \{t\}$.

According to the choice rule (9), these two conditions require the following incentive relationships:

1. $\boldsymbol{L}[z', t] - \boldsymbol{L}[z, t] < \boldsymbol{L}[z', t'] - \boldsymbol{L}[z, t']$ for some $t' \in \mathcal{T} \setminus \{t\}$, and
2. $\boldsymbol{L}[z, t] - \boldsymbol{L}[z', t] < \boldsymbol{L}[z, t''] - \boldsymbol{L}[z', t'']$ for some $t'' \in \mathcal{T} \setminus \{t\}$.

These conditions imply that the prohibit patters requires the following incentive scheme:

$$\boldsymbol{L}[z', t''] - \boldsymbol{L}[z, t''] < \boldsymbol{L}[z', t] - \boldsymbol{L}[z, t] < \boldsymbol{L}[z', t'] - \boldsymbol{L}[z, t'] \text{ for some } t', t'' \in \mathcal{T} \setminus \{t\}.$$

Otherwise stated, the prohibit pattern requires that the incentive difference for choice t be strictly larger than the minimum difference among the choices and strictly smaller than the maximum difference among the treatment choices. Consequently if the prohibit pattern does not occur, then it must be the case that:

$$\boldsymbol{L}[z', t] - \boldsymbol{L}[z, t] = \max_{t' \in \mathcal{T}} \boldsymbol{L}[z', t'] - \boldsymbol{L}[z, t'] \quad \text{or} \quad \boldsymbol{L}[z', t] - \boldsymbol{L}[z, t] = \min_{t' \in \mathcal{T}} \boldsymbol{L}[z', t'] - \boldsymbol{L}[z, t'].$$

This condition is equivalent to t-monotonic incentives (30).

A.9 Examples of Incentive IV Models where UMC Holds

A.10 Proof of Theorem C.3

A.11 Proof of Theorem T.5

A.12 Proof of Theorem T.6

A.13 Doubly Robust Estimation Algorithm for Response-type Probabilities

Step 1. Partition the sample index $\mathcal{I} = \{1, \ldots, n\}$ into K subsets such that $\cup_{k=1}^{K}\{\mathcal{I}_k\} = \mathcal{I}$, where the number of partitions K is commonly fixed to five. Let $\mathcal{I}_k^c = \mathcal{I} \setminus \mathcal{I}_k$ be the complement of $\mathcal{I}_k$

Step 2. For each value $t \in \{1, 2, 3\}$ and each partition k, compute the estimator $\hat{\gamma}_{t,k,\boldsymbol{s}}$ associated with the kappa function $\kappa_{\boldsymbol{s}}(t, Z, X)$ by minimizing the following expression:

$$\hat{\boldsymbol{\gamma}}_{\boldsymbol{s},t,k} \in \arg\min_{\boldsymbol{\gamma}\in\mathbf{R}^p} \sum_{i\in\mathcal{I}_k^c} \left(\frac{1}{2}(\boldsymbol{h}(Z_i, X_i)'\boldsymbol{\gamma})^2 + \sum_{z\in\mathcal{Z}} \nu_{\boldsymbol{s}}(t, z)\boldsymbol{h}(z, X_i)'\boldsymbol{\gamma} \right) + \alpha_{\boldsymbol{\gamma}}\|\boldsymbol{\gamma}\|_1, \tag{69}$$

where $\hat{\boldsymbol{\gamma}}_{\boldsymbol{s},t,k}$ is evaluated using all data that is not in $\mathcal{I}_k$, while $\alpha_{\boldsymbol{\gamma}}$ is the penalty parameter determined by a cross-validation procedure employing all sampling data.

Step 3. For each value $t \in \{1, 2, 3\}$ and each partition I_k, compute the estimator $\hat{\boldsymbol{\beta}}_{t,k}$ associated with the propensity score $P(T = t|Z, X)$ via the least absolute shrinkage and selection operator (lasso) procedure that minimizes the following expression:

$$\hat{\boldsymbol{\beta}}_{t,k} \in \arg\min_{\boldsymbol{\beta}\in\mathbf{R}^p} \sum_{i\in\mathcal{I}_k^c} (1[T_i = t] - \boldsymbol{h}(Z_i, X_i)'\boldsymbol{\beta})^2 + \alpha_{\boldsymbol{\beta}}\|\boldsymbol{\beta}\|_1,$$

where $\alpha_{\boldsymbol{\beta}}$ is the penalty parameter also determined by via cross-validation procedure.[30]

Step 4. Given $\hat{\boldsymbol{\gamma}}_{\boldsymbol{s},t,k}$ and $\hat{\boldsymbol{\beta}}_{t,k}$, we compute the orthogonal score estimator $\hat{\psi}_{\boldsymbol{s},i,k}$ for each participant $i \in \mathcal{I}_k$ and for each partition k :

$$\hat{\psi}_{\boldsymbol{s},k,i} \equiv \sum_{t\in\mathcal{T}} \left(\boldsymbol{h}(Z_i, X_i)'\hat{\boldsymbol{\gamma}}_{\boldsymbol{s},t,k} \cdot \left(1[T_i = t] - \boldsymbol{h}(Z_i, X_i)'\hat{\boldsymbol{\beta}}_{t,k}\right) + \sum_{z\in\mathcal{Z}} \nu_{\boldsymbol{s}}(t, z)\boldsymbol{h}(z, X_i)'\hat{\boldsymbol{\beta}}_{t,k} \right).$$

Step 5. The estimator for the propensity score $P(\boldsymbol{S} = \boldsymbol{s})$ is the average of the orthogonal scores within partition, that is, $\hat{\psi}_{\boldsymbol{s},k} = |\mathcal{I}_k|^{-1} \sum_{i\in\mathcal{I}_k} \sum_{t\in\mathcal{T}} \hat{\psi}_{\boldsymbol{s},k,i}$. The final estimate is the average of the orthogonal scores across partitions, namely, $\hat{\psi}_{\boldsymbol{s}} = n^{-1} \sum_{k=1}^{K} \hat{\psi}_{\boldsymbol{s},k} \cdot |\mathcal{I}_k|$.

Step 6. Inference is performed via the bootstrap multiplier method. For each partition k, we draw B samples $\{W_i^{(b)}\}_{i\in\mathcal{I}_k}$ of i.i.d. standard normals to compute:

$$\hat{\psi}_{\boldsymbol{s},k}^{(b)} = \hat{\psi}_{\boldsymbol{s},k} + \frac{1}{n} \sum_{i\in\mathcal{I}_k}^{n} W_i^{(b)} \left(\hat{\psi}_{\boldsymbol{s},k,i} - \hat{\psi}_{\boldsymbol{s},k}\right), \text{ and } \hat{\psi}_{\boldsymbol{s}}^{(b)} = n^{-1} \sum_{k=1}^{K} \hat{\psi}_{\boldsymbol{s},k}^{(b)} \cdot |\mathcal{I}_k|.$$

[30] Note that the penalty parameters $\alpha_{\boldsymbol{\beta}}$ and $\alpha_{\boldsymbol{\gamma}}$ do not need to be the same, but the functions $\boldsymbol{h}(Z, X)$ are the same in steps 2 and 3.

We use the distribution of $\hat{\psi}_{\boldsymbol{s}}^{(b)}$ to compute the standard error of the estimator for the type probability.

A few notes on the estimation method are in order. The sample splitting in Step 1 is not necessary to secure normality of the estimator and can be voided. The estimators in Steps 2 and 3 allow for some degree of flexibility. In our setup, $(Z,X)'\hat{\boldsymbol{\beta}}_{t,k}$ estimates the propensity score and $\boldsymbol{h}(Z,X)'\hat{\gamma}_{\boldsymbol{s},t,k}$ estimates the kappa function. These estimates can be obtained by suitable alternative machine learning estimators. For instance, it is possible to transform the minimization that evaluates $\hat{\boldsymbol{\gamma}}_{\boldsymbol{s},t,k}$ in Step 2 into a standard lasso-type estimator.

Let $\boldsymbol{H}_k(z) \equiv \boldsymbol{h}(z,\boldsymbol{X})$ denotes the $|\mathcal{I}_k^c| \times p$ matrices that stack $\boldsymbol{h}(z,X_i)'$ across participants $i \in \mathcal{I}_k^c$. In the same token, let $\boldsymbol{H}_k \equiv \boldsymbol{h}(\boldsymbol{Z},\boldsymbol{X})$ be the matrix that stakes $\boldsymbol{h}(Z_i,X_i)'$ across $i \in \mathcal{I}_k^c$, and let $\boldsymbol{\iota}_k$ be the $|\mathcal{I}_k^c|$-dimensional vector of ones. In this notation, the minimization of Step 2 can be equivalently expressed as:[31]

$$\hat{\boldsymbol{\gamma}}_{\boldsymbol{s},t,k} \in \arg\min_{\boldsymbol{\gamma}\in\mathbf{R}^p} \sum_{i\in\mathcal{I}_k^c} (\boldsymbol{h}(Z_i,X_i)'\boldsymbol{\theta} - \boldsymbol{h}(Z_i,X_i)'\boldsymbol{\gamma})^2 + \alpha_{\boldsymbol{\gamma}}\|\boldsymbol{\gamma}\|_1, \text{ where } \boldsymbol{\theta} \quad \equiv (\boldsymbol{H}_k'\boldsymbol{H}_k)^{-1}\left(\sum_{z\in\mathcal{Z}} \nu_{\boldsymbol{s}}(t,z)\boldsymbol{H}_k(z)'\boldsymbol{\iota}_k\right).$$

The term $\boldsymbol{h}(Z_i,X_i)'\boldsymbol{\theta}$ can be roughly understood as the projection of the function $\sum_{z\in\mathcal{Z}} \nu_{\boldsymbol{s}}(t,z)\boldsymbol{h}(z,X_i)$ into the space generated by $\boldsymbol{h}(Z_i,X_i)$. Finally, we use the leave-one-out sampling scheme in all cross-validation methods.

A.14 Doubly Robust Estimation Algorithm for Identified Counterfactual Outcomes

Step 1. Partition $\mathcal{I}$ into $\cup_{k=1}^K\{\mathcal{I}_k\} = \mathcal{I}$, where $\mathcal{I}_k^c = \mathcal{I} \setminus \mathcal{I}_k$.

Step 2. For each k, compute the estimator $\hat{\gamma}_{t,k,\boldsymbol{s}}$ as:

$$\hat{\boldsymbol{\gamma}}_{\boldsymbol{s},t,k} \in \arg\min_{\boldsymbol{\gamma}\in\mathbf{R}^p} \sum_{i\in\mathcal{I}_k^c} \left(\frac{1}{2}(\boldsymbol{h}(Z_i,X_i)'\boldsymbol{\gamma})^2 + \sum_{z\in\mathcal{Z}} \nu_{\boldsymbol{s},t}(z)\boldsymbol{h}(z,X_i)'\boldsymbol{\gamma}\right) + \alpha_{\boldsymbol{\gamma}}\|\boldsymbol{\gamma}\|_1, \tag{70}$$

where $\alpha_{\boldsymbol{\gamma}}$ is the penalty parameter determined by a cross-validation (leave-one-out) procedure.

Step 3. For each partition k, compute the estimators $\hat{\boldsymbol{\beta}}_{t,k}$, and $\hat{\boldsymbol{\theta}}_{t,k}$ via lasso:

$$\hat{\boldsymbol{\theta}}_{t,k} \in \arg\min_{\boldsymbol{\theta}\in\mathbf{R}^p} \sum_{i\in\mathcal{I}_k^c} \left(Y \cdot 1[T_i = t] - \boldsymbol{h}(Z_i,X_i)'\boldsymbol{\theta}\right)^2 + \alpha_{\boldsymbol{\theta}}\|\boldsymbol{\theta}\|_1,$$

$$\hat{\boldsymbol{\beta}}_{t,k} \in \arg\min_{\boldsymbol{\beta}\in\mathbf{R}^p} \sum_{i\in\mathcal{I}_k^c} \left(1[T_i = t] - \boldsymbol{h}(Z_i,X_i)'\boldsymbol{\beta}\right)^2 + \alpha_{\boldsymbol{\beta}}\|\boldsymbol{\beta}\|_1,$$

where $\alpha_{\boldsymbol{\beta}}, \alpha_{\boldsymbol{\theta}}$ are the penalty parameters determined by cross-validation.

Step 4. Given $\hat{\boldsymbol{\gamma}}_{\boldsymbol{s},t,k}, \hat{\boldsymbol{\beta}}_{t,k}$, and $\hat{\boldsymbol{\theta}}_{t,k}$, for each agent $i \in \mathcal{I}_k$ and each partition k, compute the

[31] The estimator is numerically equivalent to evaluating the minimum of the function in Step 2. The equivalence is easy be shown when expressing the minimization using matrix notation.

orthogonal score $\hat{\psi}_{\boldsymbol{s},i,k}$ for $P(\boldsymbol{S}=\boldsymbol{s})$ and $\hat{\varphi}_{\boldsymbol{s},i,k}$ for $E(Y\mathbf{1}[\boldsymbol{S}=\boldsymbol{s}])$

$$\hat{\psi}_{\boldsymbol{s},k,i} \equiv \left(\boldsymbol{h}(Z_i,X_i)'\hat{\boldsymbol{\gamma}}_{\boldsymbol{s},t,k} \cdot \Big(1[T_i=t] - \boldsymbol{h}(Z_i,X_i)'\hat{\boldsymbol{\beta}}_{t,k} \Big) + \sum_{z\in\mathcal{Z}} \nu_{\boldsymbol{s}}(t,z)\boldsymbol{h}(z,X_i)'\hat{\boldsymbol{\beta}}_{t,k} \right),$$

$$\hat{\varphi}_{\boldsymbol{s},k,i} \equiv \left(\boldsymbol{h}(Z_i,X_i)'\hat{\boldsymbol{\gamma}}_{\boldsymbol{s},t,k} \cdot \Big(Y\cdot 1[T_i=t] - \boldsymbol{h}(Z_i,X_i)'\hat{\boldsymbol{\theta}}_{t,k} \Big) + \sum_{z\in\mathcal{Z}} \nu_{\boldsymbol{s}}(t,z)\boldsymbol{h}(z,X_i)'\hat{\boldsymbol{\theta}}_{t,k} \right).$$

Step 5. The estimator for $P(\boldsymbol{S}=\boldsymbol{s})$ is the average of the orthogonal scores $\hat{\psi}_{\boldsymbol{s}} = n^{-1}\sum_{k=1}^{K}\hat{\psi}_{\boldsymbol{s},k}\cdot|\mathcal{I}_k|$, where $\hat{\psi}_{\boldsymbol{s},k} = |\mathcal{I}_k|^{-1}\sum_{i\in\mathcal{I}_k}\sum_{t\in\mathcal{T}}\hat{\psi}_{\boldsymbol{s},k,i}$. The estimator for $E(Y(t)\mathbf{1}[\boldsymbol{S}=\boldsymbol{s}])$ is also the average of the orthogonal scores $\hat{\varphi}_{\boldsymbol{s}} = n^{-1}\sum_{k=1}^{K}\hat{\varphi}_{\boldsymbol{s},k}\cdot|\mathcal{I}_k|$, where $\hat{\varphi}_{\boldsymbol{s},k} = |\mathcal{I}_k|^{-1}\sum_{i\in\mathcal{I}_k}\sum_{t\in\mathcal{T}}\hat{\varphi}_{\boldsymbol{s},k,i}$. The final estimator for $E(Y(t)|\boldsymbol{S}=\boldsymbol{s})$ is the ratio $\hat{\varphi}_{\boldsymbol{s}}/\hat{\psi}_{\boldsymbol{s}}$.

Step 6. Our inference uses a multiplier bootstrap that draw B samples $\{W_i^{(b)}\}_{i\in\mathcal{I}_k}$ of i.i.d. standard normals for each partition k. We then compute both scores:

$$\hat{\psi}_{\boldsymbol{s},k}^{(b)} = \hat{\psi}_{\boldsymbol{s},k} + \frac{1}{n}\sum_{i\in\mathcal{I}_k}^{n} W_i^{(b)}\big(\hat{\psi}_{\boldsymbol{s},k,i} - \hat{\psi}_{\boldsymbol{s},k}\big), \text{ and } \hat{\psi}_{\boldsymbol{s}}^{(b)} = n^{-1}\sum_{k=1}^{K}\hat{\psi}_{\boldsymbol{s},k}^{(b)}\cdot|\mathcal{I}_k|,$$

$$\hat{\varphi}_{\boldsymbol{s},k}^{(b)} = \hat{\varphi}_{\boldsymbol{s},k} + \frac{1}{n}\sum_{i\in\mathcal{I}_k}^{n} W_i^{(b)}\big(\hat{\varphi}_{\boldsymbol{s},k,i} - \hat{\varphi}_{\boldsymbol{s},k}\big), \text{ and } \hat{\varphi}_{\boldsymbol{s}}^{(b)} = n^{-1}\sum_{k=1}^{K}\hat{\varphi}_{\boldsymbol{s},k}^{(b)}\cdot|\mathcal{I}_k|.$$

We use the joint distribution $\{\hat{\psi}_{\boldsymbol{s}}^{(b)},\hat{\varphi}_{\boldsymbol{s}}^{(b)}\}_{b=1}^{B}$ to estimate the variance matrix of the orthogonal scores denoted by $\hat{\boldsymbol{V}}(\hat{\psi}_{\boldsymbol{s}},\hat{\varphi}_{\boldsymbol{s}})$. We compute the standard error for the ratio $\hat{\varphi}_{\boldsymbol{s}}/\hat{\psi}_{\boldsymbol{s}}$ using the Delta method, namely, $\hat{\sigma} = \big(n^{-1}\boldsymbol{\omega}'\hat{\boldsymbol{V}}(\hat{\psi}_{\boldsymbol{s}},\hat{\varphi}_{\boldsymbol{s}})\boldsymbol{\omega}\big)^{1/2}$ where $\boldsymbol{\omega} = [-(\hat{\varphi}_{\boldsymbol{s}}/\hat{\psi}_{\boldsymbol{s}}^2), 1/\hat{\psi}_{\boldsymbol{s}}]'$.

The steps above differ from the estimation of type probabilities in a few instances. Step 2 uses the function $\nu_{\boldsymbol{s},\boldsymbol{t}}(Z)$ instead of $\nu_{\boldsymbol{s}}(T,Z)$. Steps 3 computes an additional parameter $\boldsymbol{\theta}$ while Step 4 computes two orthogonal scores. Steps 5 states that our estimator is a ratio of orthogonal scores means and Step 6 uses bootstrap and the delta method to evaluate the standard error of the ratio.

A.15 Doubly Robust Estimation Algorithm for Counterfactuals Using Comparison Compliers

We first consider the task of evaluating $E(Y(1)|\boldsymbol{S}=\boldsymbol{s}_{12})$. Steps 1, 5 and 6 of the previous procedure remain the same. Steps 2–4 are modified as following.

Step 2'. For each k, compute the estimator $\hat{\gamma}_{t,k,\boldsymbol{s}}$ as:

$$\hat{\boldsymbol{\gamma}}_{1,k} \in \arg\min_{\boldsymbol{\gamma}\in\mathbf{R}^p} \sum_{i\in\mathcal{I}_k^c} \left(\frac{1}{2}(\boldsymbol{h}(Z_i,X_i)'\boldsymbol{\gamma})^2 + -(\boldsymbol{h}(1,X_i)' - \boldsymbol{h}(0,X_i))'\boldsymbol{\gamma} \right) + \alpha_{\boldsymbol{\gamma}}\|\boldsymbol{\gamma}\|_1. \tag{71}$$

Step 3'. For each partition k, compute the parameters $\hat{\boldsymbol{\beta}}_{1,k}$, $\hat{\boldsymbol{\beta}}_{2,k}$, $\hat{\boldsymbol{\theta}}_{1,k}$, $\hat{\boldsymbol{\pi}}_{1,k}$, and $\hat{\boldsymbol{\pi}}_{2,k}$, via the following lasso estimations:

$$\hat{\boldsymbol{\theta}}_{1,k} \in \arg\min_{\boldsymbol{\theta}\in\mathbf{R}^p} \sum_{i\in\mathcal{I}_k^c} \left(Y \cdot 1[T_i = 1] - \boldsymbol{h}(Z_i, X_i)'\boldsymbol{\theta}\right)^2 + \alpha_{\boldsymbol{\theta}}\|\boldsymbol{\theta}\|_1,$$

$$\hat{\boldsymbol{\beta}}_{1,k} \in \arg\min_{\boldsymbol{\beta}\in\mathbf{R}^p} \sum_{i\in\mathcal{I}_k^c} \left(1[T_i = 1] - \boldsymbol{h}(Z_i, X_i)'\boldsymbol{\beta}\right)^2 + \alpha_{\boldsymbol{\beta},1}\|\boldsymbol{\beta}\|_1,$$

$$\hat{\boldsymbol{\beta}}_{2,k} \in \arg\min_{\boldsymbol{\beta}\in\mathbf{R}^p} \sum_{i\in\mathcal{I}_k^c} \left(1[T_i = 2] - \boldsymbol{h}(Z_i, X_i)'\boldsymbol{\beta}\right)^2 + \alpha_{\boldsymbol{\beta},\mathbf{2}}\|\boldsymbol{\beta}\|_1,$$

$$\hat{\boldsymbol{\pi}}_{1,k} \in \arg\min_{\boldsymbol{\pi}\in\mathbf{R}^p} \sum_{i\in\mathcal{I}_k^c} \left(1[T_i = 1]\boldsymbol{h}(Z_i, X_i)'\hat{\boldsymbol{\gamma}}_{1,k} - \boldsymbol{g}(X_i)'\boldsymbol{\pi}\right)^2 + \alpha_{\boldsymbol{\pi},1}\|\boldsymbol{\pi}\|_1,$$

$$\hat{\boldsymbol{\pi}}_{2,k} \in \arg\min_{\boldsymbol{\pi}\in\mathbf{R}^p} \sum_{i\in\mathcal{I}_k^c} \left(1[T_i = 2]\boldsymbol{h}(Z_i, X_i)'\hat{\boldsymbol{\gamma}}_{1,k} - \boldsymbol{f}(X_i)'\boldsymbol{\pi}\right)^2 + \alpha_{\boldsymbol{\pi},2}\|\boldsymbol{\pi}\|_1,$$

where $\boldsymbol{f}(X) \equiv (f_1(X), \ldots, f_q(X))'$ denote a q-dimensional vector of functions of baseline variable.

Step 4'. Given $\hat{\boldsymbol{\gamma}}_{\boldsymbol{s},t,k}, \hat{\boldsymbol{\beta}}_{t,k}$, and $\hat{\boldsymbol{\theta}}_{t,k}$, we can compute the orthogonal score $\hat{\psi}_{\boldsymbol{s},i,k}$ regarding $P(\boldsymbol{S} = \boldsymbol{s}_{21})$ for each agent $i \in \mathcal{I}_k$ and each partition k :

$$\hat{\psi}_{\boldsymbol{s},k,i} \equiv \left(\boldsymbol{h}(Z_i, X_i)'\hat{\boldsymbol{\gamma}}_{\boldsymbol{s},t,k} \cdot \left(1[T_i = 2] - \boldsymbol{h}(Z_i, X_i)'\hat{\boldsymbol{\beta}}_{t,k}\right) + \boldsymbol{h}(1, X_i)' - \boldsymbol{h}(0, X_i)'\hat{\boldsymbol{\beta}}_{t,k}\right).$$

The orthogonal score for $E(Y(1)\mathbf{1}[\boldsymbol{S} = \boldsymbol{s}_{21}])$ is cumbersome. We define the following terms to facilitate notation: $\Theta_i \equiv \boldsymbol{h}(Z_i, X_i)'\hat{\boldsymbol{\theta}}_{1,k}$, $\Lambda_{1,i} \equiv \boldsymbol{h}(Z_i, X_i)'\hat{\boldsymbol{\beta}}_{1,k}$, $\Lambda_{2,i} \equiv \boldsymbol{h}(Z_i, X_i)'\hat{\boldsymbol{\beta}}_{2,k}$, $\Delta_i \equiv \boldsymbol{h}(0, X_i)' - \boldsymbol{h}(1, X_i)$, $\kappa_i \equiv \boldsymbol{h}(Z_i, X_i)'\hat{\boldsymbol{\gamma}}_{1,k}$. $U_i \equiv \boldsymbol{f}(X_i)'\hat{\boldsymbol{\pi}}_{1,k}$, and $C_i \equiv \boldsymbol{f}(X_i)'\hat{\boldsymbol{\pi}}_{2,k}/U_i$. In this notation, we can define the orthogonal score for $E(Y(1)\mathbf{1}[\boldsymbol{S} = \boldsymbol{s}_{21}])$ associated to agent $i \in \mathcal{I}_k$ and each partition k as:

$$\begin{aligned}\hat{\varphi}_{\boldsymbol{s},k,i} \equiv{}& ((Y_i .* 1[T_i = 1] - \Theta_i)\kappa_i + (\Delta_i\Theta_i))C_i - (((1[T_i = 2] - \Lambda_{2,i})\kappa_i) \cdot (\Delta_i\Theta_i))\frac{1}{U_i} \\ & - (((1[T_i = 1] - \Lambda_{1,i})\kappa_i) \cdot (\Delta_i\Theta_i))\frac{C}{U_i} - (((\Delta_i\Lambda_{2,i})) \cdot (\Delta_i\Theta_i))\frac{1}{U_i} - (((\Delta_i\Lambda_{1,i})) \cdot (\Delta_i\Theta_i))\frac{C}{U_i}.\end{aligned}$$

As mentioned, the Steps 5–6 remains the same. This estimator evaluates $E(Y(1)|\boldsymbol{S} = \boldsymbol{s}_{12})$ which enable us to estimate the causal effect $E(Y(2) - Y(1)|\boldsymbol{S} = \boldsymbol{s}_{12})$ since $E(Y(2)|\boldsymbol{S} = \boldsymbol{s}_{12})$ was already estimated. The standard error of the causal effect is obtained via the multiplier bootstrap. The counterfactual outcome $E(Y(1)|\boldsymbol{S} = \boldsymbol{s}_{13})$ is obtained by replacing the choice 2 in Steps 3' and Step 4' by the treatment choice 3.