Noncompliance as a Rational Choice:

A Framework that Exploits Compromises in Social Experiments to Identify Causal Effects.

Rodrigo Pinto *

January 9, 2019

Abstract

Noncompliance is a pervasive problem in social experiments which hinges the identification of causal effects. This paper offers a framework in which noncompliance is not portrayed as a drawback, but a key ingredient of identification analysis. The method uses revealed preference analysis to exploit the incentives generated by the design of social experiments in order to nonparametrically identify causal parameters. The framework is used to evaluate the Moving to Opportunity, the largest housing experiment in the US. Moving to Opportunity was designed to investigate the casual effect of relocating disadvantaged families from high-poverty neighborhoods to low-poverty communities. Substantial noncompliance prevents the evaluation of *neighborhood effects*, that is the causal effect of residing in different neighborhoods types. Nevertheless, noncompliance still allows for the evaluation of *voucher effects*, that is the causal effect of being offered a voucher. Previous literature shows that voucher effects on labor market outcomes are not statistically significant. This paper exploits the incentives of the MTO intervention to identify neighborhood effects. It shows that neighborhood effects are statistically significant even though voucher effects aren't. The result reconciles MTO with a growing literature attesting the impact of neighborhood quality on economic well-being. The framework can be broadly applied to exploit economic incentives in multiple choice models with heterogeneous agents and categorical instrumental variables. I show it is possible to evaluate all causal parameters using 2SLS regressions using particular transformations of the observed data.
Keywords: Moving to Opportunity, Randomization, Selection Bias, Social Experiment; Causal Inference.
JEL codes: H43, I18, I38. J38.

*Rodrigo Pinto is an assistant Professor at the Economics Department of UCLA. I am most grateful for extensive interaction with Professors Steven Durlauf, Paul Gertler and James Heckman. I also benefit from productive interactions with Moshe Buchinsky at UCLA. I thank many Professors for useful suggestions. Among those, I cite: Stephane Bonhomme, Magne Mogstad, Azeem Shaikh, Melissa Tartari, Thibaut Lamadon, Christopher Taber, Jens Ludwig, Bernard Salanie, Yuichi Kitamura, Adriana Lleras-Muney, Dora Costa, Flavio Cunha, Maurizio Mazzocco, Orazio Attanasio, Richard Blundell, Andrew Chesher, and Amanda Agan. I also benefit from useful comments on presentations at Berkeley, Columbia, Getulio Vargas Foundation (EPGE), Emory, Northwestern University, University of California at Los Angeles (UCLA), University of Chicago (Economics and Public Policy Departments), University College London, University of Princeton, University of Pennsylvania, University of Southern California (Schaeffer and CSSR Institutes). University of Wisconsin-Madison, Yale University, and Washington University. All errors are my own.

1 Introduction

Randomized Control Trials (RCTs) were formalized by Sir Ronald Fisher (1890-1962) and are often considered the gold standard for evaluating causal effects in social experiments. Fisher (1935) worked originally in agricultural experiments. He explains that perfectly implemented RCTs enable the evaluation of average treatment effects, but experiments that depart from its original random assignments lead to bias and inaccurate interpretation of data: *"If the design of the experiment is faulty, any method of interpretation that makes it out to be decisive must be faulty too."*

Unfortunately, perfectly implemented RCTs are rare in social experiments. Most experiments suffer from some degree of *noncompliance*, when agents choose to depart from its original treatment assignments.[1] Noncompliance induces selection bias which prevents the evaluation of the causal effects intended by the RCT. Faced by this caveat, experimental economists seek strategies to prevent noncompliance while econometricians have developed statistical methods to correct for it. These efforts share the mindset that the original assignments of the RCT is a desirable benchmark and deviations from it ought to be avoided. This paper reverts this mindset by exploring a simple insight: while a departure from random assignments in an agricultural experiment is a failure of the experiment, a departure from random assignments in a social experiment is realisation of a rational choice and thereby a useful source of information.

This paper offers a framework that merges experimental design with classical economic behavior. The method enables researchers to exploit the information on the incentives induced by the design of a social experiment to identify causal effects. It uses a simple economic model that employs revealed preference analysis to characterize the set of counterfactual choices that are economically justified. The economic model is then embedded into a casual model suitable to the study of treatment effects. Depending on the design of incentives, noncompliance is not perceived as an econometric problem, but an essential tool for the identification of causal effects.

The method is used to evaluate the Moving to Opportunity (MTO) Intervention, the most influential housing experiment in the US. As of 2018, the US spends over \$ 45 billion on Housing and Urban Development. About half of this budget is used to finance the Section 8 Program, a policy that offers house-subsidizing vouchers to five million people and enables over two million low-income families to move from poor neighborhoods to better ones (Chetty et al., 2016).

The Moving to Opportunity targeted over 4,000 households living in high-poverty housing projects across five U.S. cities during the years of 1994 to 1997. It was designed to investigate the casual effect of relocating disadvantaged families from high-poverty neighborhoods to low-poverty communities (Orr et al., 2003). Families were randomly assigned into three groups: *experimental*, which granted a rent-subsidizing voucher that incentivized families to relocate from the high-poverty neighborhoods targeted by the intervention to low-poverty neighborhoods,[2], *Section 8*,

[1]Examples of works that describe the problem of noncompliance in social experiments on early childhood education are: Conti, Heckman, and Pinto (2016) (Abecedarian project), Feller et al. (2014); Kline and Walters (2016) (Head Start), Yazejian and Bryant (2012) (Educare program), Heckman et al. (2010, 2013) (Perry Program).

[2]Low-poverty neighborhoods are defined as those whose share of poor residents is below 10% according to the

whose voucher incentivized relocation to either low or medium poverty neighborhoods,[3] and *control*, which had no voucher. The experiment did not enforce relocation and MTO noncompliance was substantial, but not unusual. Nearly 50% of experimental families *did not* use the voucher to relocate while 21% of control families living in high-poverty neighborhoods relocated to low-poverty areas.

It is useful to distinguish *neighborhood effects*, the causal effect of residing in different neighborhoods types, from *voucher effects*, that is the causal effect of being offered a voucher. Noncompliance induces selection bias that hinges the evaluation of *neighborhood effects*. Families who comply with the voucher incentives differ from those who don't. Noncompliance still enables the evaluation of *voucher effects* such as the intention-to-treat and the treatment-on-the-treated.[4] Voucher effects quantify the impact of the housing policy itself and have been evaluated by a prominent MTO literature (Gennetian et al., 2012; Hanratty et al., 2003; Katz et al., 2001, 2003; Kling et al., 2007, 2005; Ladd and Ludwig, 2003; Leventhal and Brooks-Gunn, 2003; Ludwig et al., 2005, 2001).

This paper adds to MTO literature by addressing a long-standing question: *how to exploit the random assignments of vouchers to identify neighborhood effects instead of voucher effects?*[5] It moves beyond voucher effects by exploiting the information on the incentives induced by the MTO experiment.

I devise two matrices that play a primary role in this analysis: the *incentive matrix* and the *response matrix*. The incentive matrix (Table 1) encodes the incentives associated with the design of the experiment. Revealed preference analysis translates incentives into choice restrictions. Those restrictions generate the response matrix (Table 2), which conveniently describes the counterfactual choices that MTO families may take. The response matrix contains all the necessary information to examine the nonparametric identification of causal parameters. It also enables to express voucher effects in terms of neighborhood effects.

The contributions of this paper are of two types: empirical contributions to the MTO literature and theoretical contributions to the literature on policy evaluation. On the empirical realm, this paper shows that neighborhood effects on adult labor market outcomes are statistically significant even though voucher effects aren't. The causal effects of moving from a high-poverty neighborhoods to low-poverty neighborhoods for families that respond to voucher incentives are: 14% increase in income, a 20% increase in employment and an increase of 38% on the likelihood of breaking out of poverty. Moreover, greater gaps in neighborhood poverty levels correspond to larger causal effects. This result reconciles MTO with a recent literature that attests the influence of neighborhood quality in the economic well-being of its residents (Aliprantis and Richter, 2014; Chetty et al., 2017, 2016; Chyn, 2016). I also evaluate the share of the families that belong to each of the

1990 sensus Orr et al. (2003).

[3]Medium-poverty neighborhoods are defined by exclusion. Those are the neighborhoods that are not the high-poverty neighborhoods targeted by the intervention nor the ones classified as low-poverty neighborhoods.

[4]Intention-to-treat is the outcome difference-in-mean between voucher assignments while treatment-on-the-treated is the intention-to-treat divided by the voucher compliance rate. See Appendix B for a detailed discussion.

[5]See Clampet-Lundquist and Massey (2008); Ludwig et al. (2008); Sampson (2008) for a debate whether neighborhood effects can be assessed from voucher effects.

Table 1: MTO Incentive Matrix

Group Assignment	Choices: Neighborhood Poverty Level High (t_h)	Medium (t_m)	Low (t_l)
Control (z_c)	0	0	0
Section 8 (z_8)	0	1	1
Experimental (z_e)	0	0	1

This table summarizes the incentives of the MTO vouchers (rows) associated with each neighborhood type (columns). Control families (z_c) received no voucher and its incentives are indicated by the elements zero in the first row of the incentive matrix. The Section 8 voucher (z_8) incentivizes relocation to either medium (t_m) or low-poverty (t_l) neighborhoods as indicated by two elements one in the second row. The experimental voucher (z_e) incentivizes low-poverty (t_l) neighborhood relocation as indicated the element one in the last row of the incentive matrix.

response-types $\boldsymbol{s}_1, ..., \boldsymbol{s}_7$ in Table 2 and the expected value of family pre-program variables for each response-type. These analyses are particularly useful in designing more efficient interventions.

On the theoretical realm, this paper provides a range of identification and estimation results that apply to any social experiment characterised by *Monotonic Incentives* (Pinto, 2016). These consist of experimental designs in which a change in the instrument (i.e. vouchers) induces changes in incentives toward the same direction *for all choices*. Otherwise stated, a change in the instrument that increases incentives for a choice t cannot decrease incentives for another choice t'. MTO complies with this criteria: as the voucher ranges along $z_c \rightarrow z_e \rightarrow z_8$, the incentives in Table 1 are weakly increasing *for all neighborhood choices*.

Monotonic incentives attributes a useful decomposition and four non-trivial properties to the response matrix. These features render a range of identification and estimation results. For instance, I present general close-form solutions for the nonparametric identification of counterfactual outcomes and show that each counterfactual outcome can be estimated by a Two-stage Least Square (2SLS) regression using particular transformations of the observed data. Monotonic incentives also imply the *unordered monotonicity* criteria of Heckman and Pinto (2018). The criteria is used to develop a specification test on model assumptions. Unordered monotonicity also entails a separability condition that is used to extend the method of Local Instrumental Variables of Heckman and Vytlacil (1999) for the case of multiple choices. This extension is used to address a problem of partial identification of counterfactual outcomes that is common in multiple choice models with discrete instruments.

This paper proceeds as follows. Section 2 explains the identification problem in MTO. Section 3 describes the MTO intervention. Section 4 uses a simple economic model to characterize neighborhood choices in MTO. The section combines the MTO experimental design and revealed preference analysis to generate the response matrix. Section 5 is devoted to the study of the response matrix properties. All identification results and estimation methods stem from these properties. Section 6 exploits unordered monotonicity to assess model assumptions. Section 7 merges the economic model

Table 2: MTO Response Matrix

Group Assignment	Counterfactual Choices	Economically Justifiable Response-types $\boldsymbol{s}_1$	$\boldsymbol{s}_2$	$\boldsymbol{s}_3$	$\boldsymbol{s}_4$	$\boldsymbol{s}_5$	$\boldsymbol{s}_6$	$\boldsymbol{s}_7$
Control (z_c)	$T_\omega(z_c)$	t_h	t_m	t_l	t_h	t_h	t_m	t_h
Section 8 (z_8)	$T_\omega(z_8)$	t_h	t_m	t_l	t_m	t_l	t_m	t_m
Experimental (z_e)	$T_\omega(z_e)$	t_h	t_m	t_l	t_l	t_l	t_l	t_h

Each column of the MTO response matrix displays a response-type, which is the unobserved 3-dimensional vector of counterfactual choices $\boldsymbol{S}_w = [T_\omega(z_c), T_\omega(z_8), T_\omega(z_e)]'$ that a family ω would take if it were assigned to control (z_c), Section 8 (z_8), and experimental (z_e) vouchers respectively. The response matrix consists of seven response-types $\boldsymbol{s}_1, \ldots, \boldsymbol{s}_7$. Response-type $\boldsymbol{s}_1 = [t_h, t_h, t_h]'$, corresponds to families that choose high-poverty neighborhoods regardless of the voucher assignment. Families of type $\boldsymbol{s}_2$ and $\boldsymbol{s}_3$ always choose medium and low-poverty neighborhoods respectively. Families of response-type $\boldsymbol{s}_4 = [t_h, t_m, t_l]'$ choose high, medium and low-poverty neighborhoods when assigned to control, Section 8, and experimental vouchers respectively. $\boldsymbol{s}_5$-families choose a low-poverty neighborhood whenever subsidy is available but remain in high-poverty areas otherwise; $\boldsymbol{s}_6$-families choose a low-poverty neighborhood if the available subsidy applies only for this neighborhood type, and choose medium-poverty otherwise; $\boldsymbol{s}_7$-families chose a medium-poverty neighborhood if subsidy is available and remain in high-poverty neighborhoods otherwise.

into a causal model that is suitable to investigate the identification of causal effects. Section 7.1 presents general identification results for social experiments characterized by monotonic incentives. Section 7.2 expresses the treatment-on-the-treated parameter in terms of neighborhood effects. Section 8 shows that counterfactual outcome means can be estimated by standard Two-Stage Least Squares upon a suitable data transformation. Section 9 addresses the problem of partial identification of counterfactual outcomes. Section 10 summarizes the theoretical contributions. Section 11 presents empirical results and Section 12 concludes.

2 Understanding the MTO Identification Problem

I use a familiar binary choice model of to introduce the MTO identification problem. Consider a simplification of MTO design in which families choose between a low-poverty neighborhood (t_l) or a high-poverty neighborhood (t_h) and are randomly assigned to two groups: a control group (z_c) that does not grant vouchers and an experimental group (z_e) that awards a rent-subsiding voucher that applies to dwellings in low-poverty neighborhoods. Let $T_\omega \in \{t_l, t_h\}$ denotes the neighborhood choice of family ω, $Z_\omega \in \{z_c, z_e\}$ denotes the group assignment that plays the role of an instrumental variable and $T_\omega(z)$ be the counterfactual (potential) choice that family ω would choose if it were assigned to a voucher $z \in \{z_c, z_e\}$.

The response variable $\boldsymbol{S}_\omega = [T_\omega(z_c)\ ,\ T_\omega(z_e)]'$ is the unobserved two-dimensional vector of counterfactual choices of family ω if assigned to z_c and z_e, respectively. Vector $\boldsymbol{S}_\omega$ can take four values termed *response-types*. Angrist, Imbens, and Rubin (1996) name these response-types as:

never takers $\boldsymbol{S}_\omega = [t_h, t_h]'$, compliers $\boldsymbol{S}_\omega = [t_h, t_l]'$, always takers $\boldsymbol{S}_\omega = [t_l, t_l]'$, and defiers $\boldsymbol{S}_\omega = [t_l, t_h]'$ as in Table 3 below.

Table 3: Possible Response-types for the Binary Neighborhood Choice with Binary Voucher

		Response-types			
Voucher Assignment	Counterfactual Choices	Never Takers	Compliers	Always Takers	Defiers
Control (z_c)	$T_\omega(z_c)$	t_h	t_h	t_l	t_l
Experimental ($z_e c$)	$T_\omega(z_e)$	t_h	t_l	t_l	t_h
	Response variable	$\boldsymbol{S}_\omega = [t_h, t_h]'$	$\boldsymbol{S}_\omega = [t_h, t_l]'$	$\boldsymbol{S}_\omega = [t_l, t_l]'$	$\boldsymbol{S}_\omega = [t_l, t_h]'$

Voucher z_e incentivizes low-poverty neighborhood relocation while z_c does not. Thus it is intuitive to assume that the instrument change from z_c to z_e may influence family choices towards only low-poverty neighborhood (t_l). This choice behavior is termed *monotonicity* by Imbens and Angrist (1994) and can be equivalently declared as either a choice inequality or a choice restriction:

$$\underbrace{\mathbf{1}[T_\omega(z_c) = t_l] \leq \mathbf{1}[T_\omega(z_e) = t_l]}_{\text{Choice Inequality}} \quad \equiv \quad \underbrace{T_\omega(z_c) = t_l \Rightarrow T_\omega(z_e) = t_l}_{\text{Choice Restriction}} \text{ for all families } \omega, \qquad (1)$$

The choice inequality in (1) uses an indicator function $\mathbf{1}[\,\cdot\,]$ to state that the instrument change $z_c \to z_e$ induces families to choose low-poverty neighborhoods t_l. The choice restriction in (1) employs an implication and states that if a family ω chooses t_l under voucher z_c then it must be the case that family ω also chooses t_l under z_e. Both expressions capture the notion that the impact of incentives is coherent across all families. The monotonicity criteria (1) eliminates the defiers of Table 3. This elimination enables the identification the Local Average Treatment Effect (LATE), that is the causal effect of low versus high-poverty neighborhood for compliers: $E(Y(t_l) - Y(t_h)|\boldsymbol{S} = [t_h, t_l]')$, where $Y(t_l), Y(t_h)$ denotes counterfactual outcomes when choice is fixed at t_l and t_h.

that is

MTO differs from the binary LATE model of Table 3 as vouchers may take *three* values – control (z_c), experimental (z_e), or Section 8 (z_8) – and families decide among *three* choices – high (t_h), medium (t_m) or low-poverty (t_l) neighborhoods. Let $Z_\omega \in \{z_c, z_8, z_e\}$ denotes the voucher assignment of family ω which plays the role of an instrumental variable and $T_\omega \in \{t_h, t_m, t_l\}$ be neighborhood choice of family ω. The response variable $\boldsymbol{S}_\omega = [T_\omega(z_c)\ ,\ T_\omega(z_8)\ ,\ T_\omega(z_e)]'$ is the unobserved three-dimensional vector of counterfactual choices $T_\omega(z)$ that family ω would choose if assigned to z_c, z_8 and z_e respectively. $\boldsymbol{S}_\omega = [t_h\ ,\ t_m\ ,\ t_l]'$ means that family ω chooses high-poverty neighborhood if assigned to control ($T_\omega(z_c) = t_h$), medium-poverty under Section 8 ($T_\omega(z_8) = t_m$), and low-poverty under the experimental voucher ($T_\omega(z_e) = t_l$).[6]

Remark 2.1. While the number of possible response-types in the binary LATE model of Table 3 is 4, the total number of possible response-types in MTO is 27. If family ω were assigned to the control

[6] If all families were of this response-type, then the comparison between experimental and control families would identify the causal effect of low-poverty versus high-poverty neighborhoods on the outcome.

group, the family could choose either one of the three neighborhood types, i.e., $T_\omega(z_c) \in \{t_h, t_m, t_l\}$. If family ω were assigned to Section 8, the family may also choose among the three neighborhood types, i.e., $T_\omega(z_8) \in \{t_h, t_m, t_l\}$. The combination of these possibilities generates $3\times3 = 9$ possible patterns. Under the experimental voucher, family ω still has three possible choices $T_\omega(z_e) \in \{t_h, t_m, t_l\}$. Thus the number of possible response-types that the response vector $\boldsymbol{S}_\omega = [T_\omega(z_c), T_\omega(z_8), T_\omega(z_8)]$, can take totals $3 \times 3 \times 3 = 27$. The 27 possible response-types in MTO are presented in Panel A of Table 7. The identification of causal parameters depends on the elimination of some response-types in the same fashion that the monotonicity assumption (1) eliminates defiers in the binary LATE model.

I exploit the information on the incentives induced by the experimental design of MTO to eliminate response-types. Table 1 explains that the experimental voucher (z_e) incentivizes the choice of low-poverty neighborhood (t_l), Section 8 (z_8) incentivizes both low and medium-poverty neighborhoods (t_l, t_m) while the control group z_c has no incentives. Remark *2.2* investigates two response-types in light of MTO incentives.

Remark 2.2. MTO incentives imply that some response-types are unlikely. Consider the response-type $\boldsymbol{S}_\omega = [t_h, t_l, t_h]'$ which means that family ω chooses a low-poverty neighborhood under Section 8 ($T_\omega(z_8) = t_l$) but switches to a high-poverty under the experimental voucher ($T_\omega(z_e) = t_h$). The switch is incoherent as both vouchers subsidize low-poverty neighborhood. Another unlikely response-type is $\boldsymbol{S}_\omega = [t_m, t_m, t_h]'$, in which family ω chooses a high-poverty neighborhood under the experimental voucher ($T_\omega(z_e) = t_h$), but switches to medium-poverty under no voucher ($T_\omega(z_c) = t_m$). The switch lacks justification as these vouchers (z_e, z_c) do not subsidize neither high (t_h) nor medium-poverty (t_m) neighborhoods.

Response-types could be systematically eliminated by extending the monotonicity assumption (1) to the multiple choices of MTO. To do so, we can use the incentive matrix (Table 1) to examine the choice incentives induced by changes in instrumental values. There are three possibilities: **(1)** the change $z_c \to z_e$ induces families to choose low-poverty (t_l) neighborhoods; **(2)** the change $z_c \to z_8$ induces families to choose low (t_l) or medium-poverty (t_m) neighborhoods; **(3)** the change $z_e \to z_8$ induces families to choose medium-poverty (t_m) neighborhoods; These incentives are translated into the three monotonicity criteria of Table 4.

Table 4: Monotonicity Argument : Choice Inequalities and Equivalent Choice Restrictions

	Choice Inequalities	Equivalent Choice Restrictions
Monotonicity Criteria 1	$\mathbf{1}[T_\omega(z_c) = t_l] \leq \mathbf{1}[T_\omega(z_e) = t_l]$	$T_\omega(z_c) = t_l \;\Rightarrow\; T_\omega(z_e) = t_l$
Monotonicity Criteria 2	$\mathbf{1}[T_\omega(z_c) \in \{t_m, t_l\}] \leq \mathbf{1}[T_\omega(z_8) \in \{t_m, t_l\}]$	$T_\omega(z_c) \neq t_h \;\Rightarrow\; T_\omega(z_8) \neq t_h$
Monotonicity Criteria 3	$\mathbf{1}[T_\omega(z_e) = t_m] \leq \mathbf{1}[T_\omega(z_8) = t_m]$	$T_\omega(z_e) = t_m \;\Rightarrow\; T_\omega(z_8) = t_m$

First row of Table 4 states that if a family chooses a low-poverty neighborhood under control ($T_\omega(z_c) = t_l$) then it must chooses low-poverty neighborhood under the experimental voucher $T_\omega(z_e) = t_l$. This monotonicity criteria eliminates all the six response-types in which $T_\omega(z_c) = t_l$ but $T_\omega(z_e) \neq t_l$, namely, $[t_l, t_m, t_h]', [t_l, t_h, t_h]', [t_l, t_l, t_h]'$ and $[t_l, t_m, t_m]', [t_l, t_h, t_m]', [t_l, t_l, t_m]'$. Panel B of Table 7 lists the response-types that are eliminated by each monotonicity criteria of Table 4.

Overall, the three monotonicity criteria in Table 4 eliminate 13 out of 27 possible response-types. Unfortunately, this elimination is insufficient to render the identification of causal parameters.

I make the case that revealed preference analysis is more powerful and intuitive than monotonicity criteria of Table 4. For instance, the monotonicity criteria of Table 4 are not able to eliminate the response-types discussed in Remark *2.2*. Nevertheless those response-types can be eliminated by a simple revealed preference argument. Consider response-type $[t_m, t_m, t_h]'$. If family ω chooses a medium-poverty neighborhood under no voucher ($T_\omega(z_c) = t_m$). This means that medium-poverty is revealed preferred to high-poverty neighborhoods under no subsidy. The experimental voucher does not subsidize high-poverty neighborhoods. Thus, by the Weak Axiom of Revealed Preference(WARP), high-poverty neighborhood cannot be revealed preferred to medium-poverty under the experimental voucher. To summarize, MTO incentives combined with WARP generates the choice restriction $T_\omega(z_c) = t_m \;\Rightarrow T_\omega(z_e) \neq t_h$ which eliminates the response-type $[t_m, t_m, t_h]'$ of Remark *2.2*.

In Section 4, I use a simple economic choice model to examine the incentives induced by the design of an experiment. I use choice axioms commonly evoked in economics to translate the incentive matrix into choice restrictions that eliminate response-types. I show that the MTO incentive matrix generates seven choice restrictions that subsume the three monotonicity criteria of Table 4. These choice restrictions eliminate 20 out of 27 response-types. The seven response-types that survive this elimination process constitute the response matrix displayed in Table 2. All identification results stem from the properties of this response matrix.

3 The MTO Experiment: Data and Design

MTO is a housing experiment that targeted poor families living in high-poverty housing projects across five U.S. cities – Baltimore, Boston, Chicago, Los Angeles, and New York – between June 1994 and July 1998. The MTO sample totals 4,248 disadvantaged families, two-thirds of whom were African-American, three-quarters were on welfare, 92% of the households were headed by a female and less than 50% of those completed high school. I refer to Orr et al. (2003); Sanbonmatsu et al. (2011) for an detailed description of the experiment. Appendix C presents statistical description of selected variables of the MTO intervention regarding neighborhood choice, poverty levels and voucher compliance.

Families were randomly allocated into three groups: 28% to the Section 8 group (z_8), 41% to the experimental group (z_e) and 31% to control (z_c). Section 8 families were offered a rent-subsidizing voucher that could be used if a family agreed to relocate from the original housing projects to eligible private-market dwellings. Experimental families were offered a voucher similar to the Section 8 voucher but could be used only in low-poverty neighborhoods. Experimental families also received some counseling from local nonprofit organizations to search for houses. Control families were offered no voucher.

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

Figure 1 summarizes the relocation patterns of MTO families. A sizeable share of families did not use the voucher to relocate. The compliance rate for the experimental voucher was 47% while the compliance rate for the Section 8 voucher was 59 %. Table A.5 presents compliance rates by site.

Table 5: Compliance Rates by Site

Site	All Sites	Baltimore	Boston	Chicago	Los Angeles	New York
Experimental Compliance Rate	47 %	58 %	46 %	34 %	67 %	45%
Section 8 Compliance Rate	59 %	72 %	48 %	66 %	77 %	49%

This tables presents the fraction of voucher recipients that used the voucher (compliance rate) to relocate by site.

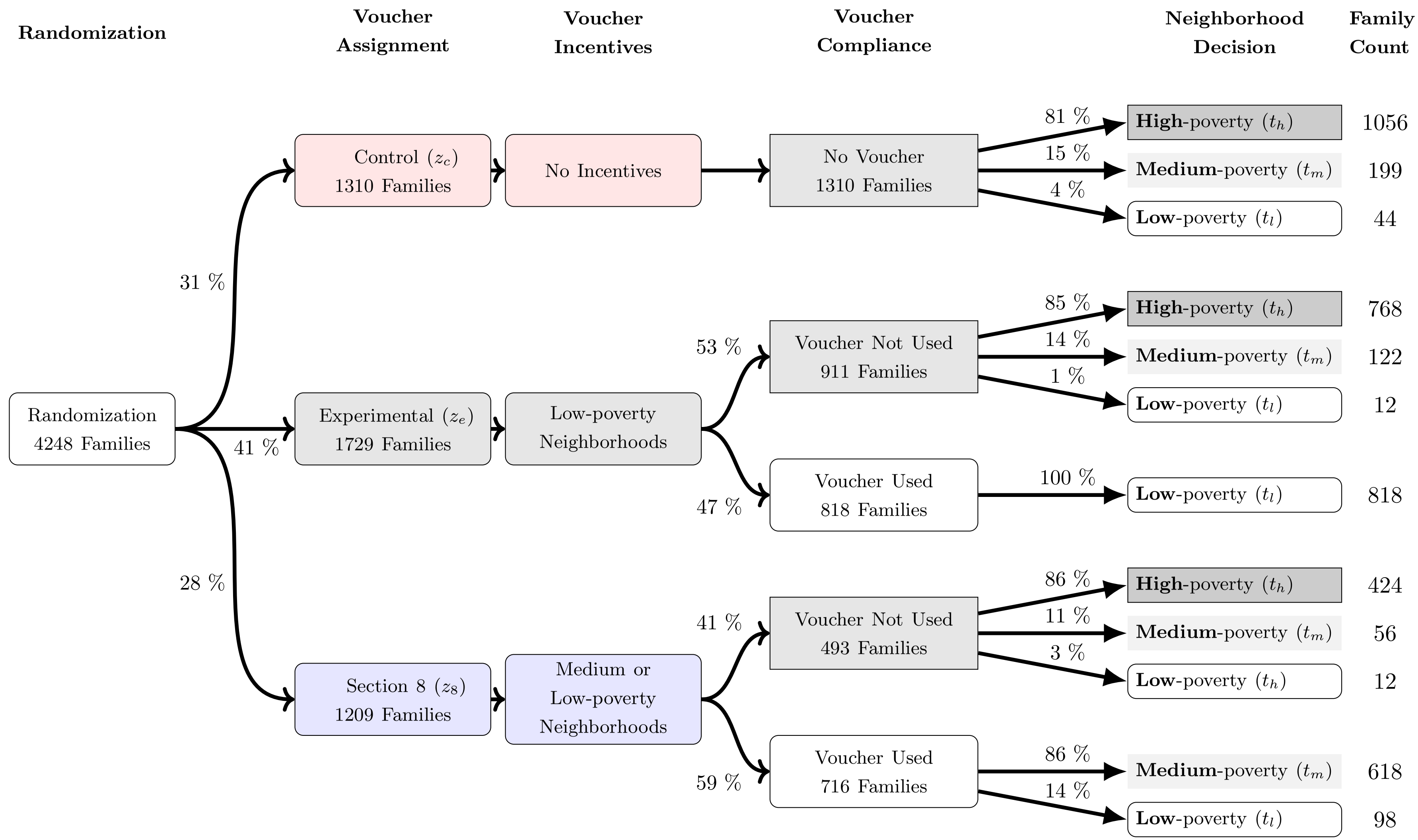
A baseline survey was conducted at the onset of the intervention. An impact interim evaluation conducted in 2002 (four to seven years after enrollment) collected data on employment, household income and public assistance.[7]

Table 6 describes these baseline variables prior to neighborhood decision. Columns 2–6 shows that variables are reasonably balanced across the voucher assignments apart some sampling variations. The remaining columns show evidence of selection bias regarding voucher usage. Columns 7–9 compare the baseline variables for experimental families that used and did not use the voucher. Columns 10–12 present a similar analysis for Section 8 families.

Columns 10–12 of Table 6 shows that families who complied with the Section 8 voucher differ from those who did not. Families who used the voucher were smaller, had fewer teenage siblings, fewer years of residency, were more likely to feel unsafe and more likely to be victimized in the neighborhood. In the same token, Columns 7–9 of Table 6 shows that families who complied with the experimental voucher differ from those who did not. Families who used the voucher had fewer social connections. Those were less likely to chat with neighbors, have family in the neighborhood or watch for neighborhood children.

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

Figure 1: **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. Families assigned to the control group could decide among three neighborhood choices:

1. High-poverty neighborhood stands for the choice of not relocating and continue living in the housing projects targeted at the onset of the experiment.

2. Medium-poverty neighborhood stands for the choice of relocating to a neighborhood other than the ones targeted by the experimental voucher;

3. Low-poverty neighborhood

Families assigned to experimental or Section 8 groups that did not use the voucher to relocate faced the same choices. These families could still move to medium or low-poverty neighborhoods without using the subsidy rendered by the voucher. Experimental families who used the voucher could move only to low-poverty neighborhoods. Section 8 families who used the voucher could move to either medium or low-poverty neighborhoods.

Table 6: Baseline Variables of MTO by Voucher Assignment and Compliance

	Full Sample					Experimental Group			Section 8 Group		
	Control Group	Experimental vs. Control		Section 8 vs. Control		Experimental Compliers	Comparison Compliers vs. Not		Section 8 Compliers	Comparison Compliers vs. Not	
	Mean	*Diff*	*p-val*	*Diff*	*p-val*	*Mean*	*Diff*	*p-val*	*Mean*	*Diff*	*p-val*
Variable	2	3	4	5	6	7	8	9	10	11	12
Family											
Disable Household Member	0.15	0.01	0.31	0.00	0.82	0.15	-0.04	0.34	0.13	-0.06	0.23
No teens (ages 13-17) at baseline	0.63	-0.03	0.12	-0.01	0.55	0.65	0.10	**0.00**	0.66	0.11	**0.00**
Household size is 2 or smaller	0.21	0.01	0.48	0.01	0.39	0.26	0.08	**0.04**	0.23	0.03	0.56
Sociability											
No family in the neighborhood	0.65	-0.02	0.35	0.00	1.00	0.65	0.03	**0.06**	0.65	0.01	0.57
Respondent reported no friends	0.41	-0.00	0.78	-0.01	0.56	0.44	0.06	**0.02**	0.41	0.02	0.48
Chat with neighbor	0.53	-0.01	0.60	-0.03	0.19	0.50	-0.05	**0.04**	0.51	0.01	0.77
Watch for neighbor children	0.57	-0.02	0.31	-0.03	0.16	0.51	-0.07	**0.00**	0.55	0.03	0.36
Neighborhood											
Victim last 6 months (baseline)	0.41	0.01	0.41	0.01	0.45	0.45	0.05	**0.08**	0.45	0.06	**0.09**
Living in neighborhood $>$ 5 yrs.	0.60	0.00	0.97	0.02	0.28	0.59	-0.03	0.17	0.59	-0.08	**0.00**
Unsafe at night (baseline)	0.50	-0.02	0.27	-0.00	1.00	0.52	0.08	**0.00**	0.54	0.10	**0.00**
Moved due to gangs	0.78	-0.01	0.52	-0.02	0.24	0.79	0.04	**0.00**	0.78	0.04	**0.00**
Schooling											
Has a GED (baseline)	0.20	-0.03	**0.04**	0.00	0.81	0.18	0.03	0.46	0.20	0.00	0.98
Completed high school	0.35	0.04	**0.01**	0.01	0.47	0.41	0.02	0.57	0.39	0.06	0.10
Enrolled in school (baseline)	0.16	0.00	0.95	0.02	0.22	0.19	0.07	0.10	0.19	0.04	0.45
Never married (baseline)	0.62	-0.00	0.97	-0.02	0.36	0.66	0.06	**0.00**	0.63	0.05	**0.02**
Teen pregnancy	0.25	0.01	0.41	0.01	0.69	0.27	0.02	0.49	0.29	-0.09	**0.05**
Missing GED and H.S. diploma	0.07	-0.01	0.12	-0.01	0.54	0.04	-0.03	0.50	0.06	-0.01	0.80
Welfare/economics											
AFDC/TANF Recipient	0.74	0.02	0.34	0.00	0.85	0.78	0.04	**0.00**	0.78	0.08	**0.00**
Car Owner	0.17	-0.01	0.65	-0.01	0.43	0.19	0.04	0.26	0.17	0.04	0.48
Adult Employed (baseline)	0.25	0.02	0.28	0.01	0.75	0.26	-0.01	0.84	0.27	0.03	0.51

This table presents a statistical description of MTO baseline variables by group assignment and compliance decision. By baseline variables I mean pre-program variables surveyed at the onset of the intervention before neighborhood relocation. Columns 2–6 present the arithmetic means for selected baseline variables conditional on Voucher assignments. Column 2 presents the control mean. Columns 3 presents the difference in means between the Experimental and Control groups. Columns 4 shows the double-sided single-hypothesis p-value associated with the equality in means test. Inference is based on the bootstrap method. Columns 5–6 compare the Section 8 group with the control group in the same fashion as columns 3–4. Columns 7–9 examine baseline variables for the experimental group conditional on the choice of voucher compliance. Column 7 presents the variable mean conditioned on voucher compliance. Column 8 gives the difference in means between the families assigned to the Experimental voucher that used the voucher and the ones that did not use the voucher for relocation. Columns 9 shows the double sided p-value associated with the equality in means test. Columns 10–12 analyze the families assigned to the Section 8 group in the same fashion as columns 7–9.

4 From Incentives to Counterfactual Choices

This section describes the steps that yield the MTO response matrix in Table 2. The matrix arises by using revealed preferences to analyze the incentives of the MTO intervention.

The incentives induced by the design of the MTO intervention can be are characterized by the incentive matrix $\boldsymbol{L}$ in (2) which displays a ranking of choice incentives (columns) across voucher assignments (rows). Let $\boldsymbol{L}[z,t]$ denotes the value of $\boldsymbol{L}$ for instrumental value (i.e. vouchers) $z \in \{z_c, z_8, z_e\}$ and choice $t \in \{t_l, t_m, t_h\}$. Thus $\boldsymbol{L}[z,t] < \boldsymbol{L}[z',t]$ means that the incentive for choice t increases when instrument changes from z to z'. The incentive matrix is ordinal, any strictly increasing transformation of the values in $\boldsymbol{L}$ characterizes equivalent incentives and delivers identical analysis.

$$\boldsymbol{L} = \begin{array}{c} \\ \\ \\ \end{array}\begin{matrix} t_h & t_m & t_l & \\ \end{matrix}$$

$$\boldsymbol{L} = \overset{\begin{matrix} t_h & t_m & t_l \end{matrix}}{\begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}} \begin{matrix} z_c \\ z_8 \\ z_e \end{matrix} \tag{2}$$

Remark 4.1. The MTO incentive matrix (2) is a case of monotonic incentives in which a change in the instrument induce incentives towards the same direction for all choices. Notationally, it means that for any values $z, z' \in \text{supp}(Z)$, the binary matrix $\boldsymbol{L}$ satisfies the following criteria:

$$\textbf{\textit{Monotonic Incentives:}}\ \boldsymbol{L}[z,t] \leq \boldsymbol{L}[z',t]\ \forall t \in \text{supp}(T)\ \textit{or}\ \boldsymbol{L}[z,t] \geq \boldsymbol{L}[z',t] \forall t \in \text{supp}(T). \tag{3}$$

This incentive criteria renders a range of identification and estimation results discussed further in this text.

I use a simple economic model to describe the neighborhood choices of MTO families. Let the real-valued function $u_\omega(t,g)$ represents the rational preferences of a family ω over the neighborhood types $t \in \{t_h, t_m, t_l\}$ and consumption goods $g \in \mathcal{G}$. Let $\mathcal{B}_\omega(z,t) \subset \mathcal{G}$ be the unobserved budget set of consumption goods for family ω when the neighborhood choice is *fixed* at $t \in \{t_h, t_m, t_l\}$ and the voucher assignment is *fixed* at $z \in \{z_c, z_8, z_e\}$. The budget set $\mathcal{B}_\omega(z,t)$ must be understood broadly. It comprises typical items such as food, clothing, leisure, but also housing characteristics. Equation (4) expresses the neighborhood choice of family ω when the voucher is fixed at z as a utility maximization problem:

$$T_\omega(z) = \underset{t \in \{t_l, t_m, t_h\}}{\arg\max} \left(\max_{g \in \mathcal{B}_\omega(z,t)} u_\omega(t,g) \right) \tag{4}$$

Incentives enter this economic model as inclusion relations among the budget sets. For instance, Section 8 subsidizes medium-poverty neighborhoods while the experimental voucher z_e does not. Vouchers subsidies allow families to afford consumption bundles that exceed the family's own resources. Thus if the neighborhood choice were *fixed* at medium-poverty (t_m), then a family ω would face a larger budget set if assigned to Section 8 (z_8) instead of the experimental voucher (z_e). Notationally, this means that $\mathcal{B}_\omega(z_e, t_m) \subset \mathcal{B}_\omega(z_8, t_m)$. On the other hand, neither z_8 nor z_e subsidize high-poverty neighborhood (t_h). Thus if the neighborhood choice were fixed at t_h, the budget sets of family ω would remain the same, that is, $\mathcal{B}_\omega(z_e, t_h) = \mathcal{B}_\omega(z_8, t_h)$. Assumption **A-1** lists all budget set relations generated by the voucher subsidies.

Assumption A-1. For any family ω, the following inclusion relations among budget sets hold:

Poverty Level		Control z_c		Experimental z_e		Section 8 z_8
High	t_h	$\mathcal{B}_\omega(z_c, t_h)$	$=$	$\mathcal{B}_\omega(z_e, t_h)$	$=$	$\mathcal{B}_\omega(z_8, t_h)$
Medium	t_m	$\mathcal{B}_\omega(z_c, t_m)$	$=$	$\mathcal{B}_\omega(z_e, t_m)$	$\subset$	$\mathcal{B}_\omega(z_8, t_m)$
Low	t_l	$\mathcal{B}_\omega(z_c, t_l)$	$\subset$	$\mathcal{B}_\omega(z_e, t_l)$	$=$	$\mathcal{B}_\omega(z_8, t_l)$

Each row presets relations among budget sets of family ω, $\mathcal{B}_\omega(z,t)$, across voucher assignments $z \in \{z_c, z_8, z_e\}$ for a fixed neighborhood choice $t \in \{t_h, t_m, t_l\}$.

Remark 4.2. Budget set relations in ***A-1*** *stem from the incentive matrix* (2) *and can be generically characterized by the following the rule:*

$$\boldsymbol{L}[z,t] \leq \boldsymbol{L}[z',t] \Rightarrow \mathcal{B}_\omega(z,t) \subset \mathcal{B}_\omega(z',t) \,\forall\, \omega. \quad (5)$$

The budget set relations in ***A-1*** *arise by applying* (5) *for each column* $t \in \{t_l, t_m, t_h\}$ *and across all rows* $z, z' \in \{z_c, z_8, z_e\}$ *of the incentive matrix* $\boldsymbol{L}$ *in* (2).

4.1 Choice Restrictions and The Response Matrix

Section 2 explains that the identification of causal parameters hinges on the elimination of response-types that are unlikely to occur. The elimination is based on a set of choice restrictions that arise by using revealed preference analysis to exploit the information on the incentives induced by a social experiment. Lemma **L-1** states a general rule that uses the Weak Axiom of Revealed Preferences (WARP)[8] to translate the content of the incentive matrix $\boldsymbol{L}$ into choice restrictions:

Lemma L-1. WARP implies that choice rule (6) holds for any instrumental values of $z, z' \in \text{supp}(Z)$ and any choices $t, t' \in \text{supp}(T)$:

$$\text{If} \quad T_\omega(z) = t \quad \text{and} \quad \boldsymbol{L}[z',t'] - \boldsymbol{L}[z,t'] \leq 0 \leq \boldsymbol{L}[z',t] - \boldsymbol{L}[z,t] \quad \text{then} \quad T_\omega(z') \neq t'. \quad (6)$$

Proof. See Appendix A.1. □

The choice rule in **L-1** is intuitive. Suppose a family ω chooses t when assigned to z, i.e. $T_\omega(z) = t$. This means that choice t is preferred to t' under z. As the instrument changes from z to z', the incentives to choose t increase $(0 \leq \boldsymbol{L}[z',t] - \boldsymbol{L}[z,t])$ and the incentives to choose t' decrease $(\boldsymbol{L}[z',t'] - \boldsymbol{L}[z,t'] \leq 0)$. As a consequence, family ω does not choose t' under z'.[9] Lemma **L-2** lists the choice restrictions generated by applying **L-1** to the incentive matrix (2).

[8]This paper adopts the WARP criteria suggested by Richter (1971) which is based on the interpretation of Samuelson (1938) seminal paper . Specifically, if bundle (t,g) is chosen over (t',g') when both were available in the budget set $\mathcal{B}$, then bundle (t,g) is said to be *(directly) strictly revealed preferred* to (t',g'), that is, $(t,g) \succ^d_\omega (t',g')$. WARP means that if (t,g) is revealed preferred to (t',g') then it cannot be the case that (t',g') is revealed preferred to (t,g), in short, $(t,g) \succ^d_\omega (t',g') \Rightarrow (t',g') \not\succ^d_\omega (t,g)$.

[9]WARP operates in the following manner. If $T_\omega(z) = t$, then there must exists an (unobserved) bundle (t, g^*); $g^* \in \mathcal{B}_\omega(z,t)$ that is strictly revealed preferred to any bundle (t',g'); $g' \in \mathcal{B}_\omega(z,t')$. As the instrument shifts $z \to z'$, the

Lemma L-2. WARP and Budget Relations **A-1** generate the following choice restrictions:

	Choice Restrictions generated by WARP		
1	$T_\omega(z_c) = t_l$	$\Rightarrow$	$T_\omega(z_e) = t_l$ and $T_\omega(z_8) \neq t_h$
2	$T_\omega(z_c) = t_m$	$\Rightarrow$	$T_\omega(z_e) \neq t_h$ and $T_\omega(z_8) \neq t_h$
3	$T_\omega(z_e) = t_m$	$\Rightarrow$	$T_\omega(z_c) = t_m$ and $T_\omega(z_8) = t_m$
4	$T_\omega(z_e) = t_h$	$\Rightarrow$	$T_\omega(z_c) = t_h$ and $T_\omega(z_8) \neq t_l$
5	$T_\omega(z_8) = t_h$	$\Rightarrow$	$T_\omega(z_c) = t_h$ and $T_\omega(z_e) = t_h$
6	$T_\omega(z_8) = t_l$	$\Rightarrow$	$T_\omega(z_e) = t_l$

Proof. See Appendix A.2. □

Choice restrictions in **L-2** have clear interpretations. The first restriction states that if a family chooses low-poverty under control group – no subsidy – then this family should also choose low-poverty under the experimental voucher, which subsidizes low-poverty neighborhoods. Moreover, this family does not choose high-poverty under Section 8, which subsidizes both low and medium-poverty neighborhoods.

Remark 4.3. Choice restrictions in ***L-2*** *emerges from the combining WARP with the budget set relations* ***A-1*** *that arise from the incentive design of the experiment. Those restrictions hold for each family ω regardless whether budget sets $\mathcal{B}_\omega(z,t); t \in \{t_h, t_m, t_l\}, z \in \{z_c, z_8, z_e\}$ are observed. The restrictions also hold regardless of family's voucher assignment Z_ω, neighborhood choice T_ω, or whether the family uses its assigned voucher to relocate.*

If a family decides to relocate to a low-poverty neighborhood under no subsidy, then this family reveals a preference for low-poverty instead of medium-poverty neighborhoods. It is reasonable to assume that this family would maintain its decision if a subsidy that applies to both low and medium-poverty neighborhoods were offered. This rationale can be understood as assuming that choices are normal goods. Consider a family that debates between low and medium poverty neighborhoods. Suppose that this family chooses a low-poverty neighborhood under no subsidies. This decision can interpret as the consumption of one unit of the good called low-poverty neighborhood. Now suppose a subsidy is offered to *both* neighborhood types. The subsidy can be interpreted as an increase in income as it applies to both of the choices being considered. If the neighborhood choice is a normal good, then an increase in income cannot decrease its consumption. Thus the family would still consume one unit of the low-poverty neighborhood good. Assumption (**A-2**) formalizes this rationale in terms of incentives $\boldsymbol{L}$:

Assumption A-2. Normal Choice: Let $z, z' \in \text{supp}(Z)$ and $t, t' \in \text{supp}(T)$ such that

$$\text{If} \quad T_\omega(z) = t \quad \text{and} \quad \boldsymbol{L}[z,t'] = \boldsymbol{L}[z,t] < \boldsymbol{L}[z',t'] = \boldsymbol{L}[z',t] \quad \text{then} \quad T_\omega(z') \neq t'. \tag{7}$$

Normal Choice **A-2** adds another choice restriction to the six restrictions generated by WARP (**L-2**). The new restriction can be stated as $T_\omega(z_c) \neq t_h \Rightarrow T_\omega(z_8) = T_\omega(z_c)$. Panel C′ of Table 7 displays the seven choice restrictions. Panel C of Table 7 shows that these restrictions eliminate 20

consumption budget set associated with t increases as $\mathcal{B}_\omega(z,t) \subseteq \mathcal{B}_\omega(z',t)$ due to $\boldsymbol{L}[z,t] \leq \boldsymbol{L}[z',t]$ while the budget for t' shrinks $\mathcal{B}_\omega(z,t') \supseteq \mathcal{B}_\omega(z',t')$ due to $\boldsymbol{L}[z,t'] \geq \boldsymbol{L}[z',t,]$. Thus, by WARP, no bundle $(t',g'); g' \in \mathcal{B}_\omega(z',t')$ can be reveled preferred to (t,g^*) and thereby $T_\omega(z') \neq t'$.

of the 27 possible response-types.[10] The seven response-types that survive the elimination process are arranged into the Response Matrix $\boldsymbol{R}$ displayed in Lemma **L-3**.

Lemma L-3. Let $T_\omega(z)$ be defined as in (4) such that Budget Relations **A-1** hold. Then WARP **A-1** and Normal Choice **A-2** generate the following response matrix $\boldsymbol{R}$:

$$\text{MTO Response Matrix:} \quad \boldsymbol{R} = \begin{array}{c} \begin{matrix} \boldsymbol{s}_1 & \boldsymbol{s}_2 & \boldsymbol{s}_3 & \boldsymbol{s}_4 & \boldsymbol{s}_5 & \boldsymbol{s}_6 & \boldsymbol{s}_7 \end{matrix} \\ \begin{bmatrix} t_h & t_m & t_l & t_h & t_h & t_m & t_h \\ t_h & t_m & t_l & t_m & t_l & t_m & t_m \\ t_h & t_m & t_l & t_l & t_l & t_l & t_h \end{bmatrix} \end{array} \begin{matrix} T_\omega(z_c) \\ T_\omega(z_8) \\ T_\omega(z_e) \end{matrix} \tag{8}$$

Proof. See Panel C of Table 7. □

The MTO response matrix $\boldsymbol{R}$ in (8) displays counterfactual choices by voucher assignments. Its columns correspond the seven response-types $\boldsymbol{s}_1, \ldots, \boldsymbol{s}_7$ that survive the elimination process while rows correspond to each voucher assignments z_c, z_e, z_8.

Each response-type characterizes a particular choice behavior. Response-types $\boldsymbol{s}_1, \boldsymbol{s}_2, \boldsymbol{s}_3$ are called always-takers as those refer to families who choose the same neighborhood type regardless of the voucher assignment. The voucher incentives are insufficient to instigate relocation for families of type $\boldsymbol{s}_1$, who always choose a low-poverty neighborhood. Families of type $\boldsymbol{s}_2$ always choose a medium-poverty neighborhood while $\boldsymbol{s}_3$-families always choose low-poverty.

Response-types $\boldsymbol{s}_4, \boldsymbol{s}_5, \boldsymbol{s}_6, \boldsymbol{s}_7$ are called compliers (or switchers) and refer to families that change their neighborhood choice according to voucher assignments.[11] $\boldsymbol{s}_4$-families are the most responsive. They chose high-poverty under no voucher, medium-poverty under Section 8 and low-poverty under the experimental voucher. $\boldsymbol{s}_5$-families relocate to low-poverty neighborhood whenever subsidy is available but remain in high-poverty areas under no subsidy. $\boldsymbol{s}_6$-families choose low-poverty neighborhoods if this is the only available subsidy. Otherwise, these families choose medium-poverty neighborhood. $\boldsymbol{s}_7$-families chose medium-poverty whenever subsidy is available and remain in high-poverty neighborhood otherwise. Section 6 uses observed data on propensity scores to test if the MTO response matrix is empirically sound.

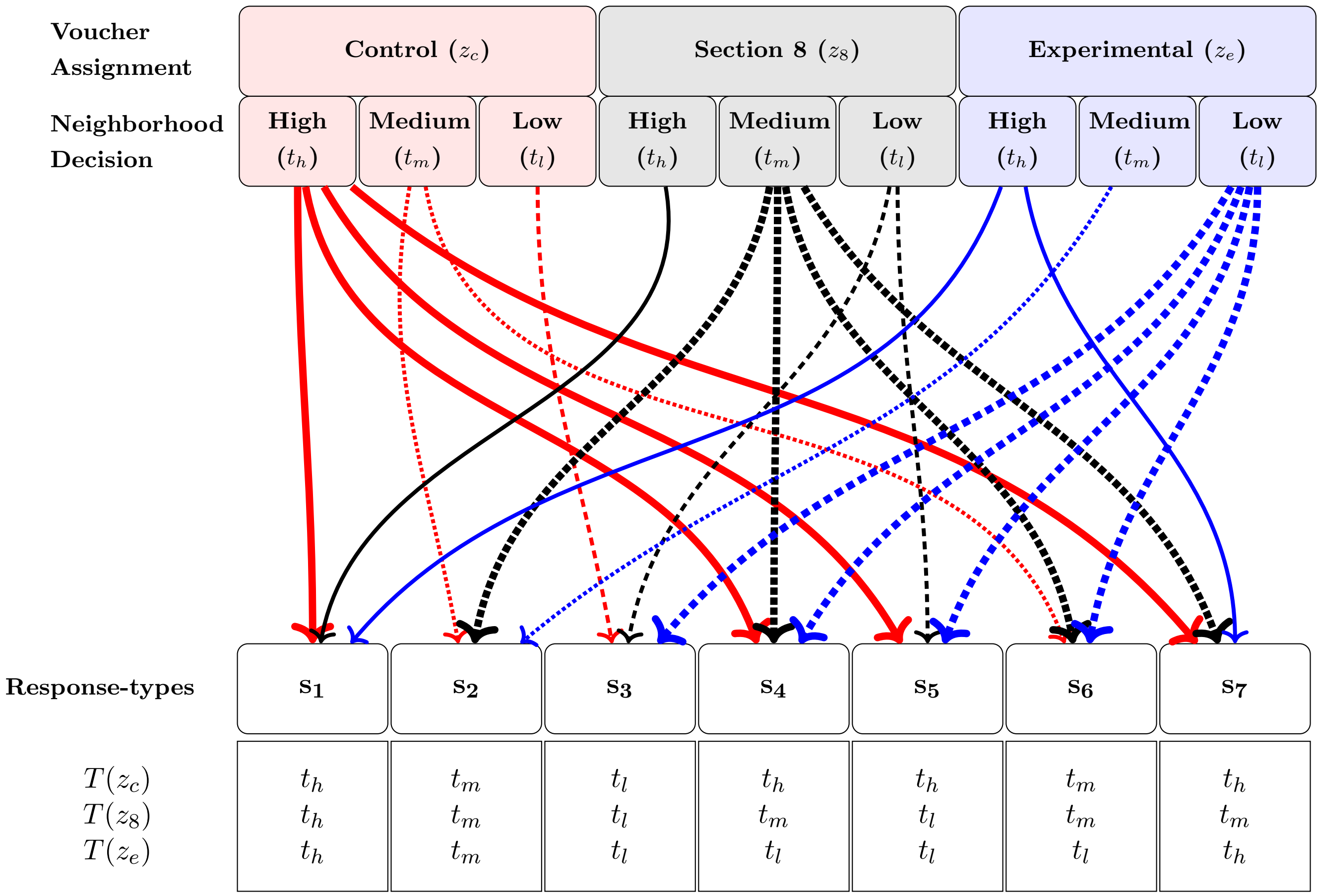
The response-types are not observed. Nevertheless, the response matrix determines the set of possible response-type that a family can take given its voucher assignment and its neighborhood choice. If family ω is assigned to Section 8 ($Z_\omega = z_8$) and chooses a low-poverty neighborhood ($T_\omega = t_l$), then, according to the second row of the response matrix $\boldsymbol{R}$, this family must be of type $\boldsymbol{s}_3$ or $\boldsymbol{s}_5$. Figure 2 displays a diagram that maps observed data on voucher assignments and

[10] WARP alone eliminates 18 response-types while Normal Choice **A-2** eliminates two.

[11] Appendix I examines a modification of the incentive matrix that increases the incentives of choosing a low-poverty neighborhood when the family is assigned to the experimental voucher. This increase in incentives can be interpreted as the counseling component of the experimental voucher. This modification adds one response-type to the response matrix which refers to the fraction of $\boldsymbol{s}_1$ families (t_h-always-takers) that are persuaded to move to a low-poverty *only* due to counseling but are *not* persuaded to move to a low-poverty neighborhood by the rent subsidy. The behavior is unlikely.

neighborhood choice into the seven unobserved response-types of the MTO response matrix.

Figure 2: **From Observed Vouchers Assignments and Neighborhood Choices to Unobserved Response-types**



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 (8). These response-types are denoted by $\boldsymbol{s}_1, ..., \boldsymbol{s}_7$. 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.

Table 7: Elimination of MTO Response-types

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

Panel A lists the 27 possible response-types that the response variable $S_\omega = [T_\omega(z_c)\,,\, T_\omega(z_8)\,,\, T_\omega(z_e)]$ can take. Rows present the counterfactual neighborhood choices that a family ω could choose if it were assigned to control group (z_c), Section 8 (z_8), and experimental group (z_e) respectively. Columns present all the values of response-type as choices range over $\text{supp}(T) = \{t_h, t_m, t_l\}$. Panel B describes an elimination process based on the three monotonicity criteria of Table 4 in Section 2. 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 of Section 4. These choice restrictions are also displayed in Panel C′ below.

Panel B′ – Monotonicity Relations

Monotonicity Criteria 1	$\mathbf{1}[T_\omega(z_c) = t_l]$	$\leq$	$\mathbf{1}[T_\omega(z_e) = t_l]$
Monotonicity Criteria 2	$\mathbf{1}[T_\omega(z_c) \in \{t_m, t_l\}]$	$\leq$	$\mathbf{1}[T_\omega(z_8) \in \{t_m, t_l\}]$
Monotonicity Criteria 3	$\mathbf{1}[T_\omega(z_e) = t_m]$	$\leq$	$\mathbf{1}[T_\omega(z_8) = t_m]$

Panel C′ – Choice Restrictions

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

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.

5 Properties of MTO Response Matrix

This section is dedicated to the study of the properties of the response matrix $\boldsymbol{R}$. All theoretical results regarding identification and estimation of causal parameters stem these properties. Section 5.1 presents a decomposition that expresses the response matrix as a sum of binary matrices. Section 5.2 exploits this decomposition to establish a necessary and sufficient condition for the unordered monotonicity criteria to hold. Section 5.3 builds upon 5.1 and 5.2 to state four properties of the MTO response matrix termed as (1) Monotonicity; (2) Nested Choices; (3) Lonesum; and (4) Separability.

5.1 Response Matrix Decomposition

Let $\boldsymbol{B}_t = \mathbf{1}[\boldsymbol{R} = t]; t \in \{t_l, t_m, t_h\}$ be the binary matrix that indicates which elements in $\boldsymbol{R}$ take value t. Each binary matrix $\boldsymbol{B}_t; t \in \{t_h, t_m, t_l\}$ has three rows corresponding to the instrumental values (z_c, z_8, z_e) and seven columns corresponding to the response-types $(\boldsymbol{s}_1, \ldots, \boldsymbol{s}_7)$. I use $\boldsymbol{R}[z, \boldsymbol{s}]$ to denote the element in matrix $\boldsymbol{R}$ associated with instrumental value z and response-type $\boldsymbol{s}$ and $\boldsymbol{B}_t[z, \boldsymbol{s}]$ for the respective element of $\boldsymbol{B}_t$. Therefore $\boldsymbol{R}[z, \boldsymbol{s}] = t \Leftrightarrow \boldsymbol{B}_t[z, \boldsymbol{s}] = 1$.

Equations (10),(11) and (12) display $\boldsymbol{B}_{t_h}, \boldsymbol{B}_{t_m}$ and $\boldsymbol{B}_{t_l}$ respectively. Each binary matrix $\boldsymbol{B}_t$ is decomposed 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

$$\textbf{Binary Decomposition:} \quad \boldsymbol{R} \equiv \sum_{t \in \text{supp}(T)} t \cdot \boldsymbol{B}_t = \sum_{t \in \text{supp}(T)} t \cdot \boldsymbol{C}_t \boldsymbol{A}_t, \tag{9}$$

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

$$\boldsymbol{B}_{t_h} = \overset{\begin{matrix}\boldsymbol{s}_1 & \boldsymbol{s}_2 & \boldsymbol{s}_3 & \boldsymbol{s}_4 & \boldsymbol{s}_5 & \boldsymbol{s}_6 & \boldsymbol{s}_7\end{matrix}}{\begin{bmatrix} 1 & 0 & 0 & 1 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}} = \underbrace{\overset{\begin{matrix}\boldsymbol{s}_4,\boldsymbol{s}_5 & \boldsymbol{s}_7 & \boldsymbol{s}_1\end{matrix}}{\begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}}}_{\boldsymbol{C}_{t_h}} \cdot \underbrace{\overset{\begin{matrix}\boldsymbol{s}_1 & \boldsymbol{s}_2 & \boldsymbol{s}_3 & \boldsymbol{s}_4 & \boldsymbol{s}_5 & \boldsymbol{s}_6 & \boldsymbol{s}_7\end{matrix}}{\begin{bmatrix} 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}}}_{\boldsymbol{A}_{t_h}} \tag{10}$$

$$\boldsymbol{B}_{t_m} = \overset{\begin{matrix}\boldsymbol{s}_1 & \boldsymbol{s}_2 & \boldsymbol{s}_3 & \boldsymbol{s}_4 & \boldsymbol{s}_5 & \boldsymbol{s}_6 & \boldsymbol{s}_7\end{matrix}}{\begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}} = \underbrace{\overset{\begin{matrix}\boldsymbol{s}_4,\boldsymbol{s}_7 & \boldsymbol{s}_6 & \boldsymbol{s}_2\end{matrix}}{\begin{bmatrix} 0 & 1 & 1 \\ 1 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}}}_{\boldsymbol{C}_{t_m}} \cdot \underbrace{\overset{\begin{matrix}\boldsymbol{s}_1 & \boldsymbol{s}_2 & \boldsymbol{s}_3 & \boldsymbol{s}_4 & \boldsymbol{s}_5 & \boldsymbol{s}_6 & \boldsymbol{s}_7\end{matrix}}{\begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}}}_{\boldsymbol{A}_{t_m}} \tag{11}$$

$$\boldsymbol{B}_{t_l} = \overset{\begin{matrix}\boldsymbol{s}_1 & \boldsymbol{s}_2 & \boldsymbol{s}_3 & \boldsymbol{s}_4 & \boldsymbol{s}_5 & \boldsymbol{s}_6 & \boldsymbol{s}_7\end{matrix}}{\begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 1 & 0 \end{bmatrix}} = \underbrace{\overset{\begin{matrix}\boldsymbol{s}_4,\boldsymbol{s}_6 & \boldsymbol{s}_5 & \boldsymbol{s}_3\end{matrix}}{\begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}}}_{\boldsymbol{C}_{t_l}} \cdot \underbrace{\overset{\begin{matrix}\boldsymbol{s}_1 & \boldsymbol{s}_2 & \boldsymbol{s}_3 & \boldsymbol{s}_4 & \boldsymbol{s}_5 & \boldsymbol{s}_6 & \boldsymbol{s}_7\end{matrix}}{\begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \end{bmatrix}}}_{\boldsymbol{A}_{t_l}} \tag{12}$$

Matrices $\boldsymbol{A}_t, \boldsymbol{B}_t, \boldsymbol{C}_t$ play a primary role in both identification and estimation of causal parameters. Section 7.1 shows that the identification of causal parameters depends only on the properties of binary matrices $\boldsymbol{B}_t$. The section employs matrices $\boldsymbol{C}_t$ and $\boldsymbol{A}_t$ to develop a closed-form solution for the nonparametric identification of all counterfactual outcomes. The solution applies to all social experiments that present monotonic incentives. Section 8.2, employes matrices $\boldsymbol{B}_t$ in the estimation of response-type probabilities (Lemma **L-7**). These matrices are also used in the estimation of the distribution of pre-program variables conditioned on response-types (Appendix A.13). Next section employs matrices $\boldsymbol{C}_t$ to generate a a necessary and sufficient condition for the unordered monotonicity criteria to hold.

5.2 Unordered Monotonicity: A Necessary and Sufficient Condition

Heckman and Pinto (2018) introduce a monotonicity criteria that applies to unordered choice models. The criteria is formalized in Equation (13) and is useful to investigate the properties of the MTO response matrix.

Unordered Monotonicity: *for any choice* $t \in \text{supp}(T)$, *and for any* $z, z' \in \text{supp}(Z)$,

$$\mathbf{1}[T_\omega(z) = t] \geq \mathbf{1}[T_\omega(z') = t] \text{ for all } \omega \in \Omega \text{ or } \mathbf{1}[T_\omega(z) = t] \leq \mathbf{1}[T_\omega(z') = t] \text{ for all } \omega \in \Omega, \quad (13)$$

where $\text{supp}(T)$, $\text{supp}(Z)$ *stand for the support of the choices and the instrumental variable and* Ω *denotes the sample space.*

Unordered monotonicity (13) is characterized by a choice inequality for each choice $t \in \text{supp}(T)$ and each pair of instrumental values $z, z' \in \text{supp}(Z)$. Each inequality employs choice indicators that capture the notion that all agents move against or towards a choice value as the instrument varies. For instance, $\mathbf{1}[T_\omega(z) = t] \geq \mathbf{1}[T_\omega(z') = t] \, \forall \, \omega$ in (13) means that any family ω that chooses t under z, i.e. $T_\omega(z) = t$, also chooses t under z', i.e. $T_\omega(z') = t$.[12]

Theorem **T-1** exploits the decomposition of the previous section to present a necessary and sufficient condition for unordered monotonicity.

Theorem T-1. Unordered monotonicity (13) holds if and only if Verifying Condition (14) is zero.

$$\sum_{t \in \{t_l, t_m, t_h\}} \boldsymbol{\iota}' \Big((\boldsymbol{C}_t' \bar{\boldsymbol{C}}_t) \, \odot \, (\bar{\boldsymbol{C}}_t' \boldsymbol{C}_t) \Big) \boldsymbol{\iota} = 0, \quad (14)$$

where $\boldsymbol{C}_t$ comes from decompositions (10)–(12), $\boldsymbol{\iota}$ stand for vectors of element ones,[13] $\bar{\boldsymbol{C}}_t \equiv (\boldsymbol{\iota}\boldsymbol{\iota}' - \boldsymbol{C}_t)$ is the complement of binary matrix $\boldsymbol{C}_t$, and $\odot$ denotes the Hadamard (element-wise) multiplication.

Proof. See Appendix A.3. □

The verifying condition (14) rests on a multiplication of binary matrices that is easily computed. It only depends on matrices $\boldsymbol{C}_t; t \in \text{supp}(T)$ that arises from the decomposition (9) of the response

[12] Unordered Monotonicity (13) does not require choice values to be ordered and it is useful to model unordered choice models. In contrast, the monotonicity criteria of Angrist and Imbens (1995) requires choice values to be ordered. Unordered monotonicity does not imply or is implied by the monotonicity criteria of Angrist and Imbens (1995).

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

matrix $\boldsymbol{R}$. Appendix G shows that condition (14) applies to the response matrix $\boldsymbol{R}$ in (8), therefore unordered monotonicity (13) holds.[14]

*Remark 5.1. Appendix G evaluates the condition of Theorem **T-1** for the MTO response matrix $\boldsymbol{R}$ in **L-3**. The condition is fulfilled, therefore Unordered monotonicity (13) holds. Unordered monotonicity traces back to the property of monotonic incentives of the MTO incentive matrix Pinto (2016).*

5.3 Examining the Properties of MTO Response Matrix

In the case of MTO, unordered monotonicity comprises nine inequalities, one for each of the combination of three choice values (t_l, t_m, t_h) and three pairs of instrumental values $(z_c, z_8), (z_c, z_e), (z_e, z_8)$. *Monotonicity Property* **P-1** presents an example of nine monotonicity inequalities that generate the MTO response matrix.

Property P-1. **(Monotonicity)** *The nine unordered monotonicity inequalities below generate the response matrix $\boldsymbol{R}$ of Lemma **L-3**.*

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

Proof. See Appendix A.4. □

The monotonicity relations in **P-1** are equivalent to the MTO choice restrictions (Panel C′ of Table 7) as both generate the same response matrix. In addition, Lemma **L-4** states that the monotonicity relations in **P-1** are unique. This means that a change in the direction of any of the nine monotonicity inequalities in **P-1** generates a response matrix that differs from $\boldsymbol{R}$ in **L-3**. Section 6 exploit this feature to do inference on the behavior assumptions that yield the MTO response matrix.

Lemma L-4. There is no other set of monotonicity relations other than **P-1** that generates the response matrix $\boldsymbol{R}$ in **L-3**.

Proof. See Appendix A.5. □

Heckman and Pinto (2018) show that unordered monotonicity can be traced to a lower-triangular property of binary matrices $\boldsymbol{B}_t$. Specifically, unordered monotonicity holds if and only if each binary matrix $\boldsymbol{B}_t; t \in \{t_h, t_m, t_h\}$ is equivalent to a lower triangular matrix, that is to say that each

[14] Pinto (2016) studies the choice incentives in more general settings. He shows that Monotonic Incentives (*4.1*) imply Unordered Monotonicity (13) under WARP and Normal Choice (6) .

$\boldsymbol{B}_t; t \in \{t_h, t_m, t_h\}$ can be transformed into a lower triangular matrix by row and column permutations.[15] This *triangular equivalence* is easily verified. Consider a permutation of the rows and columns of matrix $\boldsymbol{B}_{t_h}$ towards increasing values of row-sums and decreasing values of column-sums. The resulting matrix $\widetilde{\boldsymbol{B}}_{t_h}$ is displayed in (15), which is indeed lower triangular. The same feature is shared by matrices $\boldsymbol{B}_{t_m}$ and $\boldsymbol{B}_{t_l}$.

$$\underbrace{\boldsymbol{B}_{t_h} = \begin{array}{c} \\ \left[\begin{array}{ccccccc} \boldsymbol{s}_1 & \boldsymbol{s}_2 & \boldsymbol{s}_3 & \boldsymbol{s}_4 & \boldsymbol{s}_5 & \boldsymbol{s}_6 & \boldsymbol{s}_7 \\ 1 & 0 & 0 & 1 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 1 \end{array}\right] \begin{array}{c} \\ z_c \\ z_8 \\ z_e \end{array} \end{array}}_{\text{Binary Matrix } \boldsymbol{B}_{t_h}} \Rightarrow \underbrace{\widetilde{\boldsymbol{B}}_{t_h} = \left[\begin{array}{ccccccc} \boldsymbol{s}_1 & \boldsymbol{s}_7 & \boldsymbol{s}_4 & \boldsymbol{s}_5 & \boldsymbol{s}_2 & \boldsymbol{s}_3 & \boldsymbol{s}_6 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 \end{array}\right] \begin{array}{c} \\ z_8 \\ z_e \\ z_c \end{array}}_{\text{Ordered Row and Column Sums} \rightarrow \text{Lower Triangular}} \quad (15)$$

It is useful to translate the triangular equivalence of binary matrices $\boldsymbol{B}_t$ into two useful properties of the response matrix $\boldsymbol{R}$. The first property is that the sets of response-types that take the choice value t for any two instrumental values z, z' are nested. Some notation is helpful to clarify this property. Let $\Sigma_t(z)$ be the set of response-types that take value t when in instrument is set to z, that is:

$$\Sigma_t(z) \equiv \{\boldsymbol{s} \in \text{supp}(\boldsymbol{S}); \boldsymbol{R}[z, \boldsymbol{s}] = t\}. \quad (16)$$

The triangular equivalence of $\boldsymbol{B}_t$ implies that for any two instrumental values z, z' it must be the case that $\Sigma_t(z)$ is a subset of $\Sigma_t(z')$ or vice versa. In other words, the sets $\Sigma_t(z); z \in \text{supp}(Z)$ are nested. To illustrate, consider the set $\Sigma_{t_h}(z_c)$ that consists of the response types in the response matrix $\boldsymbol{R}$ that take value t_h for voucher z_c. The first row of the MTO response matrix in **L-3** corresponds to z_c and shows that response-types $\boldsymbol{s}_1, \boldsymbol{s}_7, \boldsymbol{s}_4, \boldsymbol{s}_5$ take value t_h. Thus $\Sigma_{t_h}(z_c) = \{\boldsymbol{s}_1, \boldsymbol{s}_7, \boldsymbol{s}_4, \boldsymbol{s}_5\}$. For z_e, we have that $\Sigma_{t_h}(z_e) = \{\boldsymbol{s}_1, \boldsymbol{s}_7\}$, and, for z_8, we have that $\Sigma_{t_h}(z_8) = \{\boldsymbol{s}_1\}$. As expected, these sets are nested, that is, $\Sigma_{t_h}(z_8) \subset \Sigma_{t_h}(z_e) \subset \Sigma_{t_h}(z_c)$. This *Nested Property* is formally stated in **P-2**.

Property P-2. (**Nested Choices**) *If unordered monotonicity* (13) *holds, then for any values* $z, z' \in \text{supp}(Z)$, *and any* $t \in \text{supp}(T)$, *we have that* $\Sigma_t(z) \subset \Sigma_t(z')$ *or* $\Sigma_t(z') \subset \Sigma_t(z)$.

Proof. See Appendix A.6. ☐

Nested Property **P-2** is used to generate a novel application of the well-known method of two-stage least squares in Section 8.1.

The triangular equivalence of binary matrices $\boldsymbol{B}_t$ also implies a condition on every the 2×2 sub-matrices of the response matrix $\boldsymbol{R}$. Namely, if the diagonal elements of any 2×2 sub-matrix in $\boldsymbol{R}$ take value t, then at least one of the off-diagonal must also be t. This is termed *lonesum property* and is stated in **P-3**.

[15]This means that each binary matrix $\boldsymbol{B}_t; t \in \{t_h, t_m, t_h\}$ is a lonesum matrix. This type of matrix can be equivalently characterized in terms of its row and column sums. Namely, a binary matrix $\boldsymbol{A}$ is lonesum if and only if each of its elements $\boldsymbol{A}[i, j]$ can be expressed as a function of the matrix row-sums and column-sums (Brualdi, 1980). Heckman and Pinto (2018) exploit this feature to show that unordered monotonicity is equivalent to a separability condition.

Property P-3. (**Lonesum**) *Under unordered monotonicity* (13)*, the following condition holds for any* 2×2 *sub-matrices of the response matrix* $\boldsymbol{R}$:

$$\text{if} \begin{pmatrix} \boldsymbol{R}[z, \boldsymbol{s}] & \boldsymbol{R}[z, \boldsymbol{s}'] \\ \boldsymbol{R}[z', \boldsymbol{s}] & \boldsymbol{R}[z', \boldsymbol{s}'] \end{pmatrix} = \begin{pmatrix} t & t' \\ t'' & t \end{pmatrix} \text{ then } t' = t \text{ or } t'' = t. \tag{17}$$

for any $t, t', t'' \in \text{supp}(T)$, $z, z' \in \text{supp}(Z)$ and $\boldsymbol{s}, \boldsymbol{s}' \in \text{supp}(\boldsymbol{S})$.

Proof. See Heckman and Pinto (2018). □

The final property exploits the fact that unordered monotonicity (13) holds if and only if the indicator for each choice $t \in \text{supp}(T)$ can be expressed as a function that is weakly separable on the instrumental variable Z and unobserved variables $\boldsymbol{V}$ that affect choice T. This *Separability Property* is stated in **P-4**.

Property P-4. (**Separability**) *Let* $D_t \equiv \mathbf{1}[T = t]$ *be the binary indicator for each choice* $t \in \{t_h, t_m, t_l\}$*. Then there must exist real-valued functions* $\varphi_t : \text{supp}(Z) \to \mathbb{R}$ *and* $\tau_t : \text{supp}(\boldsymbol{V}) \to \mathbb{R}$ *such that* D_t *can be expressed by a weak separable equation below:*

$$D_t = \mathbf{1}\left[\varphi_t(Z) \geq \tau_t(\boldsymbol{V})\right] \; \forall t \in \text{supp}(T). \tag{18}$$

Proof. See Heckman and Pinto (2018). □

The lonesum property **P-3** and the separability property **P-4** are used to address the problem of partial identification of counterfactual outcomes in Section 9.

6 Assessing Model Assumptions

Lemma **L-4** implies that the choice restrictions generated by the revealed preference analysis (Panel C′ of Table 7) are equivalent to (and only to) the nine monotonicity relations in **P-1**.

A benefit of the monotonicity relations over choice restrictions is that each relation $\mathbf{1}[T_\omega(z) = t] \geq \mathbf{1}[T_\omega(z') = t] \, \forall \omega \in \Omega$ implies a propensity score inequality $P(T = t|Z = z) > P(T = t|Z = z')$ that can be evaluated using observed data. Thereby we can do inference on the seven choice restrictions induced by the revealed preference analysis by verifying if the nine propensity score inequalities evaluated from data corroborate the direction of the nine monotonicity relations in **P-1**.

Table 8 presents the nine unordered monotonicity relations that generate the MTO response matrix as well as the nine propensity score inequalities obtained from observed data. The table shows that the direction of each propensity score inequality matches the direction of each monotonicity inequality in **P-1**.

Table 8 provides strong evidence in favor of the revealed preference analysis of Section 4. Suppose that each of the nine propensity score inequality could be either greater than (>) or less than (<). The combination of all possible relations of these nine inequalities totals $2^9 = 512$. Some of these combinations are not feasible because they do not account for the fact that propensity

Table 8: MTO Unordered Monotonicity and Respective Propensity Scores Inequalities

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

The third column of this table displays the nine unordered monotonicity inequalities. Those inequalities are equivalent to the choice restrictions generated by revealed preference analysis displayed in Panel C′ of Table 7. Each unordered monotonicity inequality correspond to a propensity score inequality that can be evaluated by observed data. The last column of this table presets the estimates for the unconditional propensity scores of MTO. The relation of each propensity score inequality complies with its respective unordered monotonicity criteria.

scores must sum to one, i.e. $\sum_{t\in\{t_l,t_m t_h\}} P(T = t|Z = z) = 1$ for each $z \in \{z_c, z_8, z_e, \}$. Indeed, the total number of feasible combinations of propensity score inequalities is 336. The revealed preference analysis of Section 4 endorses a single pattern out of the 336 possible ones. This pattern of propensity score inequalities is exactly the one corroborated by data.

Remark 6.1. Response matrix $\boldsymbol{R}$ *in* ***L-3*** *stems from combining MTO incentives and revealed preference analysis. The same response matrix also arise under alternative approaches that do not evoke preference analysis nor the design of the MTO intervention. Appendix H presents an alternative approach that relies on the premise that families share similar choice behavior. The approach assumes that a change in the instrument that induces a family towards a choice cannot induce another family against the same choice. This assumption can be understood as a generalized version of the notions of compliers and definers for multiple choices models. In contrast with the revealed preference analysis, this assumption is not testable. Nevertheless, both approaches yield the same Response Matrix* $\boldsymbol{R}$ *in* ***L-3****.*

7 Merging the Economic Model into a Causal Model with IV

Section 4 elicits a simple economic model to examine counterfactual choices in MTO. The economic model however is not appropriate to investigate the causal effects of neighborhood choices on the outcomes. To advance the analysis, it is necessary to merge the economic model of Section 4 into a standard IV model that is suitable to investigate the causal effects of neighborhood choices on the outcomes.

In the economic model, the choice of family ω is driven by an unobserved utility function u_ω that represents family preferences. In the causal model, the choice of family ω is expressed as a function $f_T(Z_\omega, \boldsymbol{V}_\omega)$ where family preferences are characterized w.l.o.g. by an unobserved random vector $\boldsymbol{V}_\omega$ of arbitrary dimension. Equation (19) expresses the neighborhood choice of a family ω in both economic and causal models:

$$T_\omega = \underbrace{\underset{t \in \{t_l, t_m, t_h\}}{\arg\max} \left(\max_{g \in \mathcal{B}_\omega(Z_\omega, t)} u_\omega(t, g) \right)}_{\text{Economic Model}} \equiv \underbrace{f_T(Z_\omega, \boldsymbol{V}_\omega)}_{\text{Causal Model}} \tag{19}$$

The random vector $\boldsymbol{V}$ in (19) also plays the role of the unobserved confounding variable that generates selection bias. It causes both the choice T and outcomes Y. The standard IV model is then defined by equations (20)–(21) and the independent condition (22).[16]

$$\text{Choice Equation}: \quad T = f_T(Z, X, \boldsymbol{V}); \quad \text{supp}(T) = \{t_h, t_m, t_l\}. \tag{20}$$

$$\text{Outcome Equation}: \quad Y = f_Y(T, \boldsymbol{V}, X, \epsilon). \tag{21}$$

$$\text{Independence Condition}: \quad Z \perp\!\!\!\perp (\boldsymbol{V}, \epsilon) | X; \ \text{supp}(Z) = \{z_c, z_8, z_e\}. \tag{22}$$

Variable X stands for observed pre-intervention variables that we wish to control for; Y is an observed post-intervention outcome and ϵ is an unobserved error term; Condition (22) means that $(\boldsymbol{V}, \epsilon)$ and Z are mutually independent conditioned on X. This implies that the instrument Z affects Y only through its impact on choice T. I evoke two standard regularity conditions: the expectation of Y exists, $E(|Y|) < \infty$, and that each neighborhood option is chosen by at least some families for each voucher assignment, $P(T = t | Z = z, X) > 0 \; \forall \, (z, t) \in \text{supp}(Z) \times \text{supp}(T)$. All variables are defined on the common probability space $(\Omega, \mathcal{F}, P)$, where $(X_\omega, Z_\omega, \boldsymbol{V}_\omega, T_\omega, Y_\omega, \epsilon_\omega)$ denotes the realized values of random variables $(X, Z, \boldsymbol{V}, T, Y, \epsilon)$ for family $\omega \in \Omega$. No assumptions are made on the functional form of the choice equation $f_T(\cdot)$ or the outcome equation $f_Y(\cdot)$. For sake of notational simplicity, I suppress pre-treatment variables X henceforward. All the analysis can be understood as conditional on pre-treatment variables X.

Counterfactual outcome $Y(t) = f_Y(t, \boldsymbol{V}, \epsilon)$ is defined by *fixing* the treatment T of outcome equation (21) to the value $t \in \{t_h, t_m, t_l\}$.[17] The counterfactual choice $T(z) = f_T(z, \boldsymbol{V})$ stands for the neighborhood choice when the instrument Z is *fixed* to the value $z \in \{z_c, z_8, z_e\}$. In this notation, the observed outcome Y and neighborhood choice T can be expressed as

$$Y = \sum_{t \in \text{supp}(T)} Y(t) \cdot \mathbf{1}[T = t] \;\equiv\; Y(T), \quad \text{and} \quad T = \sum_{z \in \text{supp}(Z)} T(z) \cdot \mathbf{1}[Z = z] \;\equiv\; T(Z),$$

where $\mathbf{1}[\alpha]$ is the indicator function that takes value 1 if α is true and 0 otherwise.

The *Response Variable* $\boldsymbol{S}$ is the 3-dimensional random vector of counterfactual choices defined in (23). $\boldsymbol{S}$ is unobserved as it is a function of the unobserved confounding vector $\boldsymbol{V}$ and its support

[16] In the language of Frisch (1938); Haavelmo (1943, 1944), these functions are called are autonomous equations, that means deterministic functions that remain invariant under manipulation of their arguments.

[17] *Fixing* is causal operation that plays a central role in the study of causality. See Heckman and Pinto (2014) for a recent discussion on fixing and causality.

$\text{supp}(S) = \{\boldsymbol{s}_1, \ldots, \boldsymbol{s}_7\}$ consists of the seven response-types in **L-3**.[18]

$$\boldsymbol{S} = \big[T(z_c)\,,\, T(z_8)\,,\, T(z_e)\big]' = \big[f_T(z_c, \boldsymbol{V})\,,\, f_T(z_8, \boldsymbol{V})\,,\, f_T(z_e, \boldsymbol{V})\big]'. \tag{23}$$

Choice T can be expressed as a function of instrument Z and response variable $\boldsymbol{S}$:

$$\therefore T = \big[\mathbf{1}[Z = z_c]\,,\, \mathbf{1}[Z = z_8]\,,\, \mathbf{1}[Z = z_e]\big] \cdot \boldsymbol{S}. \tag{24}$$

Equation (24) implies that T is deterministic given Z and $\boldsymbol{S}$. Independence relations (25)–(27) hold in model described by equations (20)–(23).[19]

$$\text{IV Properties :} \quad Z \perp\!\!\!\perp Y(t) \text{ and } Z \not\perp\!\!\!\perp T \tag{25}$$

$$\text{Matching Properties :} \quad Y(t) \perp\!\!\!\perp T|\boldsymbol{V} \text{ and also } Y(t) \perp\!\!\!\perp T|\boldsymbol{S} \tag{26}$$

$$\text{Response Variable Properties :} \quad \boldsymbol{S} \perp\!\!\!\perp Z, \quad Y \perp\!\!\!\perp Z|(\boldsymbol{S}, T), \text{ and } Y \perp\!\!\!\perp T|(\boldsymbol{S}, Z) \tag{27}$$

IV Properties (25) are often called exclusion restriction and IV relevance. The Matching Properties (26) imply that if $\boldsymbol{V}$ (or $\boldsymbol{S}$) were observed, the counterfactual outcome $E(Y(t)|\boldsymbol{V})$ (or $E(Y(t)|\boldsymbol{S})$) could be evaluated by the conditional expectation $E(Y|\boldsymbol{V}, T = t)$ (or $E(Y|\boldsymbol{S}, T = t)$). Independence relations (27) are useful properties of the response variable $\boldsymbol{S}$.

In short, the identification problem consists of assessing unobserved quantities, i.e., response-type probabilities $P(\boldsymbol{S} = \boldsymbol{s})$ and counterfactual expectations $E(Y(t)|\boldsymbol{S} = \boldsymbol{s})$ for $(t, \boldsymbol{s}) \in \{t_h, t_m, t_l\} \times \{\boldsymbol{s}_1, \ldots, \boldsymbol{s}_7\}$, from observed data, i.e., propensity scores $P(T = t|Z = z)$ and outcome expectations $E(Y|T = t, Z = z)$ where $(t, z) \in \{t_h, t_m, t_l\} \times \{z_c, z_8, z_e\}$. Equation (28) connects these observed and unobserved quantities:[20]

$$\underbrace{E(\kappa(Y)|T = t, Z = z)\, P(T = t|Z = z)}_{\text{Observable}} = \sum_{s \in \text{supp}(\boldsymbol{S})} \underbrace{\mathbf{1}[T = t|\boldsymbol{S} = s, Z = z]}_{\text{Known from } \boldsymbol{R}} \underbrace{E(\kappa(Y(t))|\boldsymbol{S} = s)\, P(\boldsymbol{S} = s)}_{\text{Unobserved}}, \tag{28}$$

where $\kappa : \mathbb{R} \to \mathbb{R}$ is any real-valued function. A short list of properties of the equality (28) is given below:

1. The indicator $\mathbf{1}[T = t|\boldsymbol{S} = \boldsymbol{s}, Z = z]$ in (28) is deterministic (see (24)). Most important, this indicator is known as it is given by the element $\boldsymbol{B}_t[z, \boldsymbol{s}]$ of binary matrix $\boldsymbol{B}_t = \mathbf{1}[\boldsymbol{R} = t]$. Indeed, $\boldsymbol{R}[z, \boldsymbol{s}]$ is the value that choice T takes conditioned on $Z = z, \boldsymbol{S} = \boldsymbol{s}$, that is $\boldsymbol{R}[z, \boldsymbol{s}] = (T|Z = z, \boldsymbol{S} = \boldsymbol{s})$. Thereby, the indicator $\mathbf{1}[T = t|\boldsymbol{S} = \boldsymbol{s}, Z = z]$ is equal to the binary matrix element $\boldsymbol{B}_t[z, \boldsymbol{s}]$.

2. If we set $\kappa(Y) = Y$ in (28), then the expected value of the observed outcome conditioned on neighborhood choice T and the instrument Z is expressed as a mixture latent counterfactual means $E(Y(t)|\boldsymbol{S} = s)$ and response-type probabilities $P(\boldsymbol{S} = s)$. Setting $\kappa(Y) = \mathbf{1}[Y \leq y]$ enables to express the observed cumulative density function (CDF) of the outcome as a mixture the counterfactual outcome CDFs.

3. If $\kappa(Y)$ is replaced by the pre-program variable X is such that $X \perp\!\!\!\perp T|\boldsymbol{S}$, then equation (28)

[18] $\boldsymbol{S}$ does not add additional information to the model and can be understood as a discrete transformation of random vector $\boldsymbol{V}$. $\boldsymbol{S}$ is a balancing score for $\boldsymbol{V}$, that is, it is a coarse transformation of $\boldsymbol{V}$ that suffices to control for the confounding effects of unobserved variables $\boldsymbol{V}$ that generate bias.

[19] See Heckman and Pinto (2018); Pinto (2014) for a proof.

[20] See Heckman and Pinto (2018); Pinto (2014) for a proof.

reads:

$$E(\kappa(X)|T=t, Z=z)\, P(T=t|Z=z) = \sum_{s \in \text{supp}(\boldsymbol{S})} \mathbf{1}[T=t|\boldsymbol{S}=s, Z=z]\, E(\kappa(X)|\boldsymbol{S}=s)\, P(\boldsymbol{S}=s). \tag{29}$$

4. Setting $\kappa(Y)$ to a constant term enables to express observed propensity scores as a sum of unobserved response-type probabilities:

$$P(T=t|Z=z) = \sum_{s \in \text{supp}(\boldsymbol{S})} \mathbf{1}[T=t|\boldsymbol{S}=s, Z=z]\, P(\boldsymbol{S}=s). \tag{30}$$

For notational simplicity, I suppress the transformation $\kappa(\cdot)$ henceforward. Identification results that apply to Y also apply to $\kappa(Y)$.

7.1 Identifying Causal Parameters

It is useful to express equation (28) in matrix notation to investigate identification conditions. Let $\boldsymbol{Q}_Z(t)$ be the 3×1 vector of expected outcomes $E(Y \cdot \mathbf{1}[T=t]|Z=z); t \in \{t_h, t_m, t_l\}$ when z ranges in $\{z_c, z_8, z_e\}$. Let $\boldsymbol{Q}_Z$ be the 9×1 vector that stacks vectors $\boldsymbol{Q}_Z(t_h)$, $\boldsymbol{Q}_Z(t_m)$, $\boldsymbol{Q}_Z(t_l)$:

$$\boldsymbol{Q}_Z(t) = \begin{bmatrix} E(Y \cdot \mathbf{1}[T=t]|Z=z_c) \\ E(Y \cdot \mathbf{1}[T=t]|Z=z_8) \\ E(Y \cdot \mathbf{1}[T=t]|Z=z_e) \end{bmatrix} = \begin{bmatrix} E(Y|T=t, Z=z_c)\, P(T=t|Z=z_c) \\ E(Y|T=t, Z=z_8)\, P(T=t|Z=z_8) \\ E(Y|T=t, Z=z_e)\, P(T=t|Z=z_e) \end{bmatrix} \text{ and } \boldsymbol{Q}_Z = \begin{bmatrix} \boldsymbol{Q}_Z(t_h) \\ \boldsymbol{Q}_Z(t_m) \\ \boldsymbol{Q}_Z(t_l) \end{bmatrix}. \tag{31}$$

The vectors of propensity scores $\boldsymbol{P}_Z(t); t \in \{t_l, t_m, t_h\}$ and $\boldsymbol{P}_Z$ are defined by setting Y in equation (31) to 1:

$$\boldsymbol{P}_Z(t) = \begin{bmatrix} P(T=t|Z=z_c) \\ P(T=t|Z=z_8) \\ P(T=t|Z=z_e) \end{bmatrix} \text{ and } \boldsymbol{P}_Z = \begin{bmatrix} \boldsymbol{P}_Z(t_h) \\ \boldsymbol{P}_Z(t_m) \\ \boldsymbol{P}_Z(t_l) \end{bmatrix}. \tag{32}$$

Vectors $\boldsymbol{Q}_Z$ and $\boldsymbol{P}_Z$ correspond to the left-hand side of equations (28) and (30) respectively.

Let $\boldsymbol{P}_S$ be the 7×1 unobserved vector of response-type probabilities. Let $\boldsymbol{Q}_S(t)$ be the 7×1 vector of unobserved counterfactual outcome expectations for each $t \in \{t_h, t_m, t_l\}$. $\boldsymbol{Q}_S$ is the 21×1 vector that stacks $\boldsymbol{Q}_S(t_h)$, $\boldsymbol{Q}_S(t_m)$, $\boldsymbol{Q}_S(t_l)$:

$$\boldsymbol{P}_S = \begin{bmatrix} P(\boldsymbol{S}=\boldsymbol{s}_1) \\ \vdots \\ P(\boldsymbol{S}=\boldsymbol{s}_7) \end{bmatrix}, \boldsymbol{Q}_S(t) = \begin{bmatrix} E(Y(t)|\boldsymbol{S}=\boldsymbol{s}_1)\, P(\boldsymbol{S}=\boldsymbol{s}_1) \\ \vdots \\ E(Y(t)|\boldsymbol{S}=\boldsymbol{s}_7)\, P(\boldsymbol{S}=\boldsymbol{s}_7) \end{bmatrix}, \text{ and } \boldsymbol{Q}_S = \begin{bmatrix} \boldsymbol{Q}_S(t_h) \\ \boldsymbol{Q}_S(t_m) \\ \boldsymbol{Q}_S(t_l) \end{bmatrix}. \tag{33}$$

Vectors $\boldsymbol{Q}_S$ and $\boldsymbol{P}_S$ correspond to the unobserved term in the right-hand side of equations (28) and (30) respectively.

The binary matrix $\boldsymbol{B}_t$ organizes the indicators $\mathbf{1}[T=t|\boldsymbol{S}=\boldsymbol{s}, Z=z]$ of (29) for $(z, \boldsymbol{s}) \in \text{supp}(Z) \times \text{supp}(\boldsymbol{S})$ as a rectangular array. Let $\boldsymbol{B}_Q$ be the block diagonal matrix whose main diagonal consists of matrices $\boldsymbol{B}_{t_h}, \boldsymbol{B}_{t_m}$ and $\boldsymbol{B}_{t_l}$. Let $\boldsymbol{B}_P$ be the matrix that stacks matrices $\boldsymbol{B}_{t_h}, \boldsymbol{B}_{t_m}$ and $\boldsymbol{B}_{t_l}$. Notationally, we have that:

$$\boldsymbol{B}_Q = \begin{bmatrix} \boldsymbol{B}_{t_h} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{B}_{t_m} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{B}_{t_l} \end{bmatrix} \text{ and } \boldsymbol{B}_P = \begin{bmatrix} \boldsymbol{B}_{t_h} \\ \boldsymbol{B}_{t_m} \\ \boldsymbol{B}_{t_l} \end{bmatrix}, \tag{34}$$

where $\mathbf{0}$ stands for a null matrix that has three rows and seven columns. Under this notation, we can express Equations (28) and (30) by the matrix expressions (35) and (36) respectively:

$$\text{Outcome Equation: } \boldsymbol{Q}_Z(t) = \boldsymbol{B}_t \cdot \boldsymbol{Q}_S(t); t \in \{t_h, t_m, t_l\} \Rightarrow \boldsymbol{Q}_Z = \boldsymbol{B}_Q \cdot \boldsymbol{Q}_S; \tag{35}$$

$$\text{Propensity Score Equation: } \boldsymbol{P}_Z(t) = \boldsymbol{B}_t \cdot \boldsymbol{P}_S; t \in \{t_h, t_m, t_l\} \Rightarrow \boldsymbol{P}_Z = \boldsymbol{B}_P \cdot \boldsymbol{P}_S. \tag{36}$$

Remark 7.1. A consequence of Equations (35)–(36) is that the identification of the expected counterfactual outcomes $\boldsymbol{Q}_S$ and response-type probabilities $\boldsymbol{P}_S$ depends only on the properties of matrices binary matrices $\boldsymbol{B}_Q$ and $\boldsymbol{B}_P$. For instance, if $\boldsymbol{B}_P$ is inventible, then the response-types probabilities $\boldsymbol{P}_S$ are identified by $\boldsymbol{P}_S = \boldsymbol{B}_P^{-1} \cdot \boldsymbol{P}_Z$. Binary matrices $\boldsymbol{B}_Q, \boldsymbol{B}_P$ are generated upon the response matrix $\boldsymbol{R}$. Thus the identification of causal parameters stems only on the properties of response matrix $\boldsymbol{R}$.

Identification of Counterfactual Outcomes

Theorem **T-2** provides a general identification formula that applies to any response matrix $\boldsymbol{R}$ where the unordered monotonicity (condition **T-1**) holds. The formula is based on the binary matrix decomposition $\boldsymbol{B}_t = \boldsymbol{C}_t\boldsymbol{A}_t$ exemplified in (10)–(12).

Theorem T-2. If unordered monotonicity holds, all identified counterfactual outcomes can be expressed by $\Big(\boldsymbol{A}_t\boldsymbol{Q}_S(t)\Big) \div \Big(\boldsymbol{A}_t\boldsymbol{P}_S\Big)$ where $\boldsymbol{A}_t$ stems from decomposition $\boldsymbol{B}_t = \boldsymbol{C}_t\boldsymbol{A}_t$ and $\div$ denotes element-wise division. [21] Moreover, $\boldsymbol{A}_t\boldsymbol{Q}_S(t)$ and $\boldsymbol{A}_t\boldsymbol{P}_S$ are identified by $\boldsymbol{A}_t\boldsymbol{Q}_S(t) = \big((\boldsymbol{C}_t'\boldsymbol{C}_t)^{-1}\boldsymbol{C}_t'\boldsymbol{P}_Z(t)\big)$, and $\boldsymbol{A}_t\boldsymbol{P}_S = \big((\boldsymbol{C}_t'\boldsymbol{C}_t)^{-1}\boldsymbol{C}_t'\boldsymbol{Q}_Z(t)\big)$, that is:

$$\underbrace{\boldsymbol{A}_t\boldsymbol{Q}_S(t) \div \boldsymbol{A}_t\boldsymbol{P}_S(t)}_{\text{Identified Counterfactual Outcomes}} = \underbrace{(\boldsymbol{C}_t'\boldsymbol{C}_t)^{-1}\boldsymbol{C}_t' \cdot \boldsymbol{Q}_Z(t) \div (\boldsymbol{C}_t'\boldsymbol{C}_t)^{-1}\boldsymbol{C}_t' \cdot \boldsymbol{P}_Z(t)}_{\text{Identification Formulas}}; \ \forall \ t \in \text{supp}(T). \tag{37}$$

Proof. See Appendix A.7. □

Pinto (2016) shows that under monotonic incentives (3) the identification formula (37) simplifies to:

$$\underbrace{\Big(\boldsymbol{A}_t\boldsymbol{Q}_S(t)\Big) \div \Big(\boldsymbol{A}_t\boldsymbol{P}_S\Big)}_{\text{Identified Counterfactual Outcomes}} = \underbrace{\Big(\boldsymbol{C}_t^{-1}\boldsymbol{Q}_Z(t)\Big) \div \Big(\boldsymbol{C}_t^{-1}\boldsymbol{P}_Z(t)\Big)}_{\text{Identification Formulas}}; \ \boldsymbol{B}_t = \boldsymbol{C}_t\boldsymbol{A}_t, \ t \in \text{supp}(T). \tag{38}$$

The right-hand side of (38) summarizes all identified counterfactual outcomes. The left-hand side of (38) provides the identification equations that can be evaluated by observed data. Ap-

[21] 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 diag$(\cdot)$ is the operator that transform a vector into a diagonal matrix.

pendix J exemplifies the use of the identification formula (38) to the familiar LATE model with binary choices.

Equations (39)–(40) display matrices $\boldsymbol{A}_t\boldsymbol{Q}_S(t)$ and $\boldsymbol{A}_t\boldsymbol{P}_S$ in (38) for choice t_h.

$$\boldsymbol{A}_{t_h}\boldsymbol{Q}_S(t_h) = \begin{bmatrix} E(Y(t_h)\cdot\mathbf{1}[\boldsymbol{S}\in\{\boldsymbol{s}_4,\boldsymbol{s}_5\}]) \\ E(Y(t_h)\cdot\mathbf{1}[\boldsymbol{S}=\boldsymbol{s}_7]) \\ E(Y(t_h)\cdot\mathbf{1}[\boldsymbol{S}=\boldsymbol{s}_1]) \end{bmatrix},\ \boldsymbol{A}_{t_h}\boldsymbol{P}_S = \begin{bmatrix} P(\boldsymbol{S}\in\{\boldsymbol{s}_4,\boldsymbol{s}_5\}) \\ P(\boldsymbol{S}=\boldsymbol{s}_7) \\ P(\boldsymbol{S}=\boldsymbol{s}_1) \end{bmatrix}. \tag{39}$$

Therefore the identified parameters (left-hand side of (38)) for t_h are given by:

$$\therefore \left(\boldsymbol{A}_{t_h}\boldsymbol{Q}_S(t_h)\right) \div \left(\boldsymbol{A}_{t_h}\boldsymbol{P}_S\right) = \begin{bmatrix} E(Y(t_h)|\boldsymbol{S}\in\{\boldsymbol{s}_4,\boldsymbol{s}_5\}) \\ E(Y(t_h)|\boldsymbol{S}=\boldsymbol{s}_7) \\ E(Y(t_h)|\boldsymbol{S}=\boldsymbol{s}_1) \end{bmatrix}, \tag{40}$$

Equations (42)–(43) display matrices $\boldsymbol{C}_t^{-1}\boldsymbol{Q}_Z(t)$ and $\boldsymbol{C}_t^{-1}\boldsymbol{P}_Z(t)$ in (38) for t_h :

$$\boldsymbol{C}_{t_h} = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix} \Rightarrow \boldsymbol{C}_{t_h}^{-1} = \begin{bmatrix} 1 & 0 & -1 \\ 0 & -1 & 1 \\ 0 & 1 & 0 \end{bmatrix}, \tag{41}$$

$$\Rightarrow \boldsymbol{C}_{t_h}^{-1}\boldsymbol{Q}_Z(t_h) = \begin{bmatrix} E(Y\cdot\mathbf{1}[T=t_h]|Z=z_c) - E(Y\cdot\mathbf{1}[T=t_h]|Z=z_e) \\ E(Y\cdot\mathbf{1}[T=t_h]|Z=z_e) - E(Y\cdot\mathbf{1}[T=t_h]|Z=z_8) \\ E(Y\cdot\mathbf{1}[T=t_h]|Z=z_8) \end{bmatrix}, \tag{42}$$

$$\text{and } \boldsymbol{C}_{t_h}^{-1}\boldsymbol{P}_Z(t_h) = \begin{bmatrix} P(T=t_h|Z=z_c) - P(T=t_h|Z=z_e) \\ P(T=t_h|Z=z_e) - P(T=t_h|Z=z_8) \\ P(T=t_h|Z=z_8) \end{bmatrix}. \tag{43}$$

Equation (44) displays the final identification formula in (38) for t_h. The left-hand side lists all the identified counterfactual outcomes for t_h while the right-hand side can be evaluated from observed data.

$$\therefore \underbrace{\begin{bmatrix} E(Y(t_h)|\boldsymbol{S}\in\{bms_4,\boldsymbol{s}_5\}) \\ E(Y(t_h)|\boldsymbol{S}=\boldsymbol{s}_7) \\ E(Y(t_h)|\boldsymbol{S}=\boldsymbol{s}_1) \end{bmatrix}}_{\boldsymbol{A}_{t_h}\boldsymbol{Q}_S(t_h)\div\boldsymbol{A}_{t_h}\boldsymbol{P}_S} = \underbrace{\begin{bmatrix} \frac{E(Y\cdot\mathbf{1}[T=t_h]|Z=z_c)-E(Y\cdot\mathbf{1}[T=t_h]|Z=z_e)}{P(T=t_h|Z=z_c)-P(T=t_h|Z=z_e)} \\ \frac{E(Y\cdot\mathbf{1}[T=t_h]|Z=z_e)-E(Y\cdot\mathbf{1}[T=t_h]|Z=z_8)}{P(T=t_h|Z=z_e)-P(T=t_h|Z=z_8)} \\ \frac{E(Y\cdot\mathbf{1}[T=t_h]|Z=z_8)}{P(T=t_h|Z=z_8)} \end{bmatrix}}_{\boldsymbol{C}_{t_h}^{-1}\boldsymbol{Q}_Z(t_h)\div\boldsymbol{C}_{t_h}^{-1}\boldsymbol{P}_Z(t_h)}. \tag{44}$$

The identification formulas for t_m and t_l are displayed in (45)–(46).

$$\underbrace{\begin{bmatrix} E(Y(t_m)|\boldsymbol{S}\in\{\boldsymbol{s}_4,\boldsymbol{s}_7\}) \\ E(Y(t_m)|\boldsymbol{S}=\boldsymbol{s}_6) \\ E(Y(t_m)|\boldsymbol{S}=\boldsymbol{s}_2) \end{bmatrix}}_{\boldsymbol{A}_{t_m}\boldsymbol{Q}_S(t_m)\div\boldsymbol{A}_{t_m}\boldsymbol{P}_S} = \underbrace{\begin{bmatrix} \frac{E(Y\cdot\mathbf{1}[T=t_m]|Z=z_8)-E(Y\cdot\mathbf{1}[T=t_m]|Z=z_c)}{P(T=t_m|Z=z_8)-P(T=t_m|Z=z_c)} \\ \frac{E(Y\cdot\mathbf{1}[T=t_m]|Z=z_c)-E(Y\cdot\mathbf{1}[T=t_m]|Z=z_e)}{P(T=t_m|Z=z_c)-P(T=t_m|Z=z_e)} \\ \frac{E(Y\cdot\mathbf{1}[T=t_m]|Z=z_e)}{P(T=t_m|Z=z_e)} \end{bmatrix}}_{\boldsymbol{C}_{t_m}^{-1}\boldsymbol{Q}_Z(t_m)\div\boldsymbol{C}_{t_m}^{-1}\boldsymbol{P}_Z(t_m)}, \tag{45}$$

$$\underbrace{\begin{bmatrix} E(Y(t_l)|\boldsymbol{S}\in\{bms_4,\boldsymbol{s}_6\}) \\ E(Y(t_l)|\boldsymbol{S}=\boldsymbol{s}_5) \\ E(Y(t_l)|\boldsymbol{S}=\boldsymbol{s}_3) \end{bmatrix}}_{\boldsymbol{A}_{t_l}\boldsymbol{Q}_S(t_l)\div\boldsymbol{A}_{t_l}\boldsymbol{P}_S} = \underbrace{\begin{bmatrix} \frac{E(Y\cdot\mathbf{1}[T=t_l]|Z=z_e)-E(Y\cdot\mathbf{1}[T=t_l]|Z=z_8)}{P(T=t_l|Z=z_e)-P(T=t_l|Z=z_8)} \\ \frac{E(Y\cdot\mathbf{1}[T=t_l]|Z=z_8)-E(Y\cdot\mathbf{1}[T=t_l]|Z=z_c)}{P(T=t_l|Z=z_8)-P(T=t_l|Z=z_c)} \\ \frac{E(Y\cdot\mathbf{1}[T=t_l]|Z=z_c)}{P(T=t_l|Z=z_c)} \end{bmatrix}}_{\boldsymbol{C}_{t_l}^{-1}\boldsymbol{Q}_Z(t_l)\div\boldsymbol{C}_{t_l}^{-1}\boldsymbol{P}_Z(t_l)}. \tag{46}$$

Theorem T-3 lists all the causal parameters that are nonparametrically identified by the MTO response matrix:

Theorem T-3. Response matrix $\boldsymbol{R}$ in L-3 renders the identification of the following quantities:

(1) All response-type probabilities $P(\boldsymbol{S}=\boldsymbol{s})$; $\boldsymbol{s} \in \{\boldsymbol{s}_1,\ldots,\boldsymbol{s}_7\}$ are identified.

(2) The following counterfactual outcomes are identified:

	High Poverty (t_h)	Medium Poverty (t_m)	Low Poverty (t_l)
Always-takes	$E(Y(t_h)\|\boldsymbol{S}=\boldsymbol{s}_1)$	$E(Y(t_m)\|\boldsymbol{S}=\boldsymbol{s}_2)$	$E(Y(t_l)\|\boldsymbol{S}=\boldsymbol{s}_3)$
Single response-types	$E(Y(t_h)\|\boldsymbol{S}=\boldsymbol{s}_7)$	$E(Y(t_m)\|\boldsymbol{S}=\boldsymbol{s}_6)$	$E(Y(t_l)\|\boldsymbol{S}=\boldsymbol{s}_5)$
Double response-types	$E(Y(t_h)\|\boldsymbol{S}\in\{\boldsymbol{s}_4,\boldsymbol{s}_5\})$	$E(Y(t_m)\|\boldsymbol{S}\in\{\boldsymbol{s}_4,\boldsymbol{s}_7\})$	$E(Y(t_l)\|\boldsymbol{S}\in\{\boldsymbol{s}_4,\boldsymbol{s}_6\})$

(3) For any variable X such that $X \perp\!\!\!\perp T|\boldsymbol{S}$ holds, $P(X=x|\boldsymbol{S}=\boldsymbol{s})$ is identified for all $\boldsymbol{s} \in \text{supp}(S)$ and $x \in \text{supp}(X)$.

Proof. See Appendix A.8. □

Item (1) of **T-3** states that each response-type probability is identified. In particular, it is possible to evaluate the share of families whose neighborhood decision remains the same regardless of the voucher assignment – always-takers $\boldsymbol{s}_1,\boldsymbol{s}_2,\boldsymbol{s}_3$ – the share of families that most responsive to voucher assignment $\boldsymbol{s}_4$ and the share of those families who change their choice only for a specific voucher assignment – $\boldsymbol{s}_5,\boldsymbol{s}_6,\boldsymbol{s}_7$.

Item (2) of **T-3** states that nine outcome counterfactuals are identified. The first row stands for counterfactuals for always-takes (no choice variation), the second row lists the identified counterfactuals conditioned on a single response-type, and the last row presents the counterfactual outcomes conditioned on two response-types. These are termed *primitives* as any causal parameter that is point-identified must be a combination of these nine counterfactual outcomes.

Item (3) states that the distribution of pre-program variables X conditioned on each response-type $\boldsymbol{s} \in \text{supp}(\boldsymbol{S})$ is identified. This result enables to estimate the likelihood that a given family belongs to each unobserved response-type. Let the pre-program variables of a family ω, be $X_\omega = x \in \text{supp}(X)$. The probability that this family is of type $\boldsymbol{s} \in \text{supp}(\boldsymbol{S})$ is given by:

$$P(\boldsymbol{S}=\boldsymbol{s}|X=x) = \frac{P(X=x|\boldsymbol{S}=\boldsymbol{s})\,P(\boldsymbol{S}=\boldsymbol{s})}{P(X=x)}, \tag{47}$$

where $P(X=x)$ can be evaluated through observed data and $P(\boldsymbol{S}=\boldsymbol{s}), P(X=x|\boldsymbol{S}=\boldsymbol{s})$ are identified according to Items (1) and (3) of **T-3** respectively.

*Remark 7.2. The Matrix $\boldsymbol{A}_t$ of the decomposition $\boldsymbol{B}_t = \boldsymbol{C}_t\boldsymbol{A}_t; t \in \text{supp}(T)$ in (9) indicates the identified counterfactual outcomes. Consider the choice t_h. According to **T-3**, the identified counterfactual outcomes are: $E(Y(t_h)|\boldsymbol{S}=\boldsymbol{s_1})$, $E(Y(t_h)|\boldsymbol{S}=\boldsymbol{s_7})$, and $E(Y(t_h)|\boldsymbol{S}\in\{\boldsymbol{s}_4,\boldsymbol{s}_5\})$. The rows of $\boldsymbol{A}_{t_h}$ in (10) indicate this selection of response-types. The first row takes value 1 for response-types $\{\boldsymbol{s}_4,\boldsymbol{s}_5\}$, the second row indicates $\boldsymbol{s}_7$ and the third row indicates $\boldsymbol{s}_1$.*

7.2 Interpreting the TOT Parameter

The treatment-on-the-treated parameter (**TOT**) is defined by the causal effect of being offered a voucher divided by the voucher compliance rate.[22] Let the compliance rate for the experimental and Section 8 vouchers be respectively $P(T = t_l | Z = z_e)$ and $P(T \in \{t_m, t_l\} | Z = z_8)$. Parameter $\mathbf{TOT}(z_e, z_c)$ in (48) compares experimental families versus control families while $\mathbf{TOT}(z_8, z_c)$ in (49) compares Section 8 versus control:

$$\mathbf{TOT}(z_e, z_c) \equiv \frac{E(Y|Z = z_e) - E(Y|Z = z_c)}{P(T = t_l|Z = z_e)}, \tag{48}$$

$$\mathbf{TOT}(z_8, z_c) \equiv \frac{E(Y|Z = z_8) - E(Y|Z = z_c)}{P(T \in \{t_m, t_l\}|Z = z_8)}. \tag{49}$$

Lemma L-5 expresses the **TOT** parameters in terms of neighborhood effects.

Lemma L-5. Response matrix $\boldsymbol{R}$ in L-3 implies that:

$$\mathbf{TOT}(z_e, z_c) = \left(\frac{E(Y(t_l) - Y(t_h)|\boldsymbol{S} \in \{\boldsymbol{s}_4, \boldsymbol{s}_5\})\, P(\boldsymbol{S} \in \{\boldsymbol{s}_4, \boldsymbol{s}_5\}) + E(Y(t_l) - Y(t_m)|\boldsymbol{S} = \boldsymbol{s}_6)\, P(\boldsymbol{S} = \boldsymbol{s}_6)}{P(\boldsymbol{S} \in \{\boldsymbol{s}_4, \boldsymbol{s}_5, \boldsymbol{s}_6\})} \right) \cdot \big(1 - P(\boldsymbol{S} = \boldsymbol{s}_3|\boldsymbol{S} \in \{\boldsymbol{s}_3, \boldsymbol{s}_4, \boldsymbol{s}_5, \boldsymbol{s}_6\})\big) \tag{50}$$

$$\mathbf{TOT}(z_8, z_c) = \left(\frac{E(Y(t_m) - Y(t_h)|\boldsymbol{S} \in \{\boldsymbol{s}_4, \boldsymbol{s}_7\})\, P(\boldsymbol{S} \in \{\boldsymbol{s}_4, \boldsymbol{s}_7\}) + E(Y(t_l) - Y(t_h)|\boldsymbol{S} = \boldsymbol{s}_5)\, P(\boldsymbol{S} = \boldsymbol{s}_5)}{P(\boldsymbol{S} \in \{\boldsymbol{s}_4, \boldsymbol{s}_7, \boldsymbol{s}_5\})} \right) \cdot \big(1 - P(\boldsymbol{S} = \boldsymbol{s}_2|\boldsymbol{S} \in \{\boldsymbol{s}_2, \boldsymbol{s}_4, \boldsymbol{s}_7, \boldsymbol{s}_5\})\big). \tag{51}$$

Proof. See Appendix A.9. □

L-5 reveal that the **TOT** parameters evaluate a mixture neighborhood effects times a conditional response-type probability. The first term of $\mathbf{TOT}(z_e, z_c)$ in (50) is a weighted average of two neighborhood causal effects: (1) the outcome effect of low versus high poverty neighborhoods, i.e. $Y(t_l) - Y(t_h)$, for response-types $\boldsymbol{s}_4, \boldsymbol{s}_5$; and (2) low versus medium poverty neighborhoods, i.e. $Y(t_l) - Y(t_m)$, for the response-type $\boldsymbol{s}_6$. The second term is always between 0 and 1. The larger the share of low-poverty always-takers ($\boldsymbol{s}_3$), the smaller the $\mathbf{TOT}(z_e, z_c)$ parameter. The first term of $\mathbf{TOT}(z_8, z_c)$ in (51) is also a weighted average of two neighborhood effects – medium (t_m) versus high (t_h) for $\boldsymbol{s}_4, \boldsymbol{s}_7$ and low (t_l) versus high (t_h) for $\boldsymbol{s}_5$. The second term is a conditional probability: the larger the share of medium-poverty always-takers ($\boldsymbol{s}_2$), the smaller the $\mathbf{TOT}(z_8, z_c)$ parameter.

8 Estimation Methods

Monotonic incentives 3 enables to evaluate causal parameters using well-known econometric tools. This section shows that each identified counterfactual outcome can be estimated by a Two-stage Least Squares (2SLS) regression using a suitable data transformation. This result arises from the Nested Property P-2. Furthermore, response-type probabilities can be estimated by a Linear Probability Model (LPM) that uses binary matrices $\boldsymbol{B}_t; t \in \text{supp}(T)$ as covariates. This result

[22] See Appendix B for a discussion on the treatment-on-the-treated parameter.

stems from a rank property of matrix $\boldsymbol{B}_P$ in (34). A similar regression is employed to estimate the distribution of pre-program variables conditioned on response-types.

8.1 Revisiting the Two-stage Least Squares Method

Since the seminal papers of Theil (1953, 1958), economists have used the Two-stage Least Squares estimator (2SLS) to evaluate causal effects of endogenous variables. Unordered monotonicity offers a novel application of this well-known method.

Recall that $\Sigma_t(z)$ denotes the set of response-types that take value t when in instrument is set to z. Property **P-2** states that sets $\Sigma_t(z); z \in \text{supp}(Z)$ are nested, that is, for any $t \in \text{supp}(T)$ and any $z, z' \in \text{supp}(Z)$, we have that $\Sigma_t(z) \subset \Sigma_{t_h}(z')$ or $\Sigma_t(z') \subset \Sigma_{t_h}(z)$ holds. For example, we have that $\Sigma_{t_h}(z_c) = \{\boldsymbol{s_1}, \boldsymbol{s_4}, \boldsymbol{s_5}, \boldsymbol{s_7}\} \subset \{\boldsymbol{s_1}, \boldsymbol{s_7}\} = \Sigma_{t_h}(z_e)$ and thereby $\Sigma_{t_h}(z_c) \setminus \Sigma_{t_h}(z_e) = \{\boldsymbol{s_4}, \boldsymbol{s_5}\}$. This property, combined with equation (28), enables to identify the counterfactual outcome $Y(t_h)$ conditioned on the response-types in $\Sigma_{t_h}(z_c) \setminus \Sigma_{t_h}(z_e)$ as ratio of difference-in-means:

$$\begin{aligned}\frac{E(Y\mathbf{1}[T=t_h]|Z=z_c) - E(Y\mathbf{1}[T=t_h]|Z=z_e)}{E(\mathbf{1}[T=t_h]|Z=z_c) - E(\mathbf{1}[T=t_h]|Z=z_e)} &= \frac{E(Y(t_h)|\boldsymbol{S}=\boldsymbol{s_4})P(\boldsymbol{S}=\boldsymbol{s_4}) + E(Y(t_h)|\boldsymbol{S}=\boldsymbol{s_5})P(\boldsymbol{S}=\boldsymbol{s_5})}{P(\boldsymbol{S}=\boldsymbol{s_4}) + P(\boldsymbol{S}=\boldsymbol{s_5})} \\ &= E(Y(t_h)|\boldsymbol{S} \in \Sigma_{t_h}(z_c) \setminus \Sigma_{t_h}(z_e)).\end{aligned}$$

Lemma **L-6** exploit the fact that Property **P-2** applies to any choice and any pair of instrumental values.

Lemma L-6. If unordered monotonicity holds, then for any $t \in \text{supp}(T)$ and any values $z', z \in \text{supp}(Z)$ such that $0 < P(T=t|Z=z') < P(T=t|Z=z)$, we have that $E(Y(t)|\boldsymbol{S} \in \Sigma_t(z) \setminus \Sigma_t(z'))$ and $P(\boldsymbol{S} \in \Sigma_t(z) \setminus \Sigma_t(z'))$ are identified and can be evaluated by:

$$\text{and } E(Y(t)|\boldsymbol{S} \in \Sigma_t(z) \setminus \Sigma_t(z')) = \frac{E(Y\mathbf{1}[T=t]|Z=z) - E(Y\mathbf{1}[T=t_h]|Z=z')}{E(\mathbf{1}[T=t]|Z=z) - E(\mathbf{1}[T=t]|Z=z')}, \quad (52)$$

$$P(\boldsymbol{S} \in \Sigma_t(z) \setminus \Sigma_t(z')) = E(\mathbf{1}[T=t]|Z=z) - E(\mathbf{1}[T=t]|Z=z'). \quad (53)$$

Proof. See Appendix A.10. □

The ratio in the left-hand side of Equation (52) can be estimated by a 2SLS regression that sets the binary indicator $D_t \equiv \mathbf{1}[T=t]$ as endogenous variable, the interaction $Y \cdot D_t$ as the outcome and uses the indicators $\mathbf{1}[Z=z]$ and $\mathbf{1}[Z=z']$ as instrumental variables. Theorem **T-4** describes this procedure:

Theorem T-4. *Let the 2SLS regression in (54)–(55) where the instrumental variables consist of two binary indicators $\mathbf{1}[Z=z], \mathbf{1}[Z=z']$, the choice indicator D_t plays the role of the endogenous variable and the second stage that uses $D_t \cdot Y$ as dependent variable:*

$$\textit{First Stage } D_{t,\omega} = \gamma_z \cdot \mathbf{1}[Z_\omega = z] + \gamma_{z'} \cdot \mathbf{1}[Z_\omega = z'] + \epsilon_{\omega,D} \quad (54)$$

$$\textit{Second Stage } Y_\omega \cdot D_{t,\omega} = \kappa + \beta \cdot D_{t,\omega} + \epsilon_{\omega,Y}. \quad (55)$$

If unordered monotonicity (13) hold, then for any $t \in \text{supp}(T)$ and any two values $z, z' \in \text{supp}(Z)$, the estimator β evaluates $E(Y(t)|\boldsymbol{S} \in \Sigma_t(z) \oplus \Sigma_t(z'))$, where $\oplus$ denotes set symmetric difference.[23]

[23]The symmetric difference between two sets A, B is defined by $A \oplus B = (A \setminus B) \cup (B \setminus A)$. If $A \supset B$ then

Proof. See Appendix A.11. □

Example 8.1. Consider the estimation of the counterfactual outcome $E(Y(t_h)|\boldsymbol{S} \in \{\boldsymbol{s_4}, \boldsymbol{s_5}\})$. By checking the response matrix **L-3**, we observe that: (1) response-types $\boldsymbol{s_1}, \boldsymbol{s_4}, \boldsymbol{s_5}, \boldsymbol{s_7}$ take choice value t_h under z_c; and (2) response-types $\boldsymbol{s_1}, \boldsymbol{s_7}$ take choice value t_h under z_e. The set difference is given by $\Sigma_{t_h}(z_c) \setminus \Sigma_{t_h}(z_e) = \{\boldsymbol{s_4}, \boldsymbol{s_5}\}$. Thus, according to **T-4**, $E(Y(t_h)|\boldsymbol{S} \in \{\boldsymbol{s_4}, \boldsymbol{s_5}\})$ can be estimated by the following steps:

1. Generate two instrumental variable indicators: (1) $\mathbf{1}[Z_\omega = z_c]$ that takes value 1 if family ω is assigned to z_c; and (2) $\mathbf{1}[Z_\omega = z_e]$ that takes value 1 if family ω is assigned to z_e.

2. Generate the choice indicator $D_{t_h,\omega}$ that takes value 1 if family ω chooses $T_\omega = t_h$.

3. Generate the interaction variable $Y_\omega \cdot D_{t_h,\omega}$ that multiplies the choice indicator $D_{t_h,\omega}$ with the outcome Y_ω for each family ω.

4. Perform a 2SLS regression of the interaction variable $Y_\omega \cdot D_{t_h,\omega}$ on a constant term and on the choice indicator $D_{t_h,\omega}$ using the Z-indicators as instrumental variables.

Appendix O describes a generalized version of this 2SLS estimator that controls for baseline variables and allows for a weighting matrix.

Table 9 lists the counterfactual outcome means estimated by the 2SLS regression of **T-4**. The counterfactual outcome means for always-takes $\boldsymbol{s}_1, \boldsymbol{s}_2, \boldsymbol{s}_3$ can be evaluated by least-square regressions that exploit the following equalities: $E(Y(t_h)|\boldsymbol{S} = \boldsymbol{s}_1) = E(Y|T = t_h, Z = z_8)$, $E(Y(t_m)|\boldsymbol{S} = \boldsymbol{s}_2) = E(Y|T = t_m, Z = z_e)$, and $E(Y(t_l)|\boldsymbol{S} = \boldsymbol{s}_3) = E(Y|T = t_l, Z = z_c)$.[24]

Table 9: Two-stage Least Square Estimation for Identified Parameters

Data Transformations				Identified Parameters
Endogenous Variables Choice Indicator	Dependent Variable Outcome Interaction	Instrumental Variable IV Indicators		
$D_{t_h} \equiv \mathbf{1}[T = t_h]$	$D_{t_h} \cdot Y$	$\mathbf{1}[Z = z_c]$	$\mathbf{1}[Z = z_8]$	$E(Y(t_h)\|S \in \{s_4, s_5, s_7\})$
		$\mathbf{1}[Z = z_c]$	$\mathbf{1}[Z = z_e]$	$E(Y(t_h)\|S \in \{s_4, s_5\})$
		$\mathbf{1}[Z = z_8]$	$\mathbf{1}[Z = z_e]$	$E(Y(t_h)\|S = s_7)$
$D_{t_m} \equiv \mathbf{1}[T = t_m]$	$D_{t_m} \cdot Y$	$\mathbf{1}[Z = z_c]$	$\mathbf{1}[Z = z_8]$	$E(Y(t_m)\|S \in \{s_4, s_7\})$
		$\mathbf{1}[Z = z_c]$	$\mathbf{1}[Z = z_e]$	$E(Y(t_m)\|S = s_6)$
		$\mathbf{1}[Z = z_8]$	$\mathbf{1}[Z = z_e]$	$E(Y(t_m)\|S \in \{s_4, s_6, s_7\})$
$D_{t_l} \equiv \mathbf{1}[T = t_l]$	$D_{t_l} \cdot Y$	$\mathbf{1}[Z = z_c]$	$\mathbf{1}[Z = z_8]$	$E(Y(t_l)\|S = s_5)$
		$\mathbf{1}[Z = z_c]$	$\mathbf{1}[Z = z_e]$	$E(Y(t_l)\|S \in \{s_4, s_5, s_6\})$
		$\mathbf{1}[Z = z_8]$	$\mathbf{1}[Z = z_e]$	$E(Y(t_l)\|S \in \{s_4, s_6\})$

This table describes the counterfactual means estimated by 2SLS procedure in **T-4**. The first stage estimates uses two IV indicators (columns 3 and 4) for instrumental values and a choice indicator (column 1) as endogenous variable. The second stage uses the interaction of the outcome and the choice indicator (columns 2) as dependent variable. The 2SLS regression evaluates the counterfactual expectation $E(Y(t)|\boldsymbol{S} \in \Sigma_t(z)\Delta\Sigma_t(z'))$ as described in **T-4** (column 5).

$A \oplus B = A \setminus B$. Note that the first stage does not has an intercept while the second stage has.

[24]Counterfactual outcome means for the always-takes $\boldsymbol{s}_1, \boldsymbol{s}_2, \boldsymbol{s}_3$ can also be evaluated by the 2SLS in **T-4** that uses single instrumental value indicator in (54) and suppresses the constant in (55).

8.2 Estimating Response-type Probabilities and Pre-program Variables

The response-type probabilities can be nonparametric estimated using Equation (36), i.e. $\boldsymbol{P}_Z = \boldsymbol{B}_P \cdot \boldsymbol{P}_S$. If matrix $\boldsymbol{B}_P$ in (34) has full column-rank, then response-type probabilities are point identified and can be estimated by $\boldsymbol{P}_S = \boldsymbol{B}_P^+ \boldsymbol{P}_Z$, where $\boldsymbol{B}_P^+$ is the Moore-Penrose pseudo-inverse of $\boldsymbol{B}_P$ and the vector of propensity scores $\boldsymbol{P}_Z$ can be obtained by its respective sample means. Moreover, under full column-rank, the pseudo-inverse of $\boldsymbol{B}_P$ is given by $\boldsymbol{B}_P^+ = (\boldsymbol{B}_P'\boldsymbol{B}_P)^{-1}\boldsymbol{B}_P'$, therefore the response-type probabilities are identified by:

$$\boldsymbol{B}_P \text{ full column-rank } \Rightarrow \boldsymbol{P}_S = (\boldsymbol{B}_P'\boldsymbol{B}_P)^{-1}\boldsymbol{B}_P'\boldsymbol{P}_Z. \tag{56}$$

Equation (56) can be interpreted as an OLS estimator that uses propensity scores $\boldsymbol{P}_Z$ as a dependent variable and the columns of binary matrix $\boldsymbol{B}_P$ as covariates. Lemma **L-7** shows that Response-type probabilities can be estimated by a linear regression that mirrors equation (56).

Lemma L-7. Let $\boldsymbol{R}$ be a response matrix of dimension $N_T \times N_S$ whose elements take values in $\text{supp}(T) = \{t_1, \ldots, t_{N_T}\}$ and $\boldsymbol{B}_P$ be the matrix in (34) that stacks binary matrices $\boldsymbol{B}_t = \boldsymbol{1}[\boldsymbol{R} = t]; t \in \text{supp}(T)$. If $\boldsymbol{B}_P$ has full column-rank then the vector of response-type probabilities $\boldsymbol{P}_S$ can be nonparametrically estimated by the parameter $\boldsymbol{\beta}$ of the following Least Squares Regression:

$$D_{t,\omega} = \boldsymbol{B}_{t,\omega}\boldsymbol{\beta} + \epsilon_\omega \text{ across all } t \in \text{supp}(T), \tag{57}$$

where $D_{t,\omega} \equiv \boldsymbol{1}[T_\omega = t]; t \in \{t_h, t_m, t_l\}$ is the binary indicator whether family ω chooses neighborhood t, and $\boldsymbol{B}_{t,\omega} \equiv \boldsymbol{B}_t[Z_\omega, \cdot]$ denotes the row of matrix $\boldsymbol{B}_t$ associated with the instrumental value $Z_\omega \in \{z_c, z_8, z_e\}$ assigned to family ω.

Proof. See Appendix A.12. □

The regression in Lemma **L-7** stacks the data across possible choice values and resembles a Seemingly Unrelated Regression (SUR) on the indicator $D_{t,\omega}$ on $\boldsymbol{B}_{t,\omega}$ across $t \in \{t_l, t_m, t_h\}$. To clarify, let $\mathbb{D}_{t,\Omega} = [D_{t,\omega}; \omega \in \Omega]$ be the vector that stacks indicators $D_{t,\omega}$ across the families in the sample. Let $\mathbb{D}_\Omega = [\mathbb{D}'_{t_h,\Omega}, \mathbb{D}'_{t_m,\Omega}, \mathbb{D}'_{t_l,\Omega}]'$ be the vector that stacks $\mathbb{D}_{t,\Omega}$ across neighborhood choices t_h, t_m, t_l. In the same fashion, matrix $\mathbb{B}_{t,\Omega} = [\boldsymbol{B}_{t,\omega}; \omega \in \Omega]$ stacks $B_{t,\omega}$ across all families and $\mathbb{B}_\Omega = [\mathbb{B}'_{t_h,\Omega}, \mathbb{B}'_{t_m,\Omega}, \mathbb{B}'_{t_l,\Omega}]'$ stacks matrices $\mathbb{B}_{t,\Omega}$ across neighborhood choices. The estimator of response-type probabilities is given by $\widehat{\boldsymbol{P}}_S = (\mathbb{B}_\Omega'\mathbb{B}_\Omega)^{-1}\mathbb{B}_\Omega'\mathbb{D}_\Omega$, and the estimated probabilities are identical to the ones generated by replacing the propensity scores in (36) by its respective sample means.[25]

Item 3 of **T-3** states that the distribution of pre-program variable conditioned on response-types is identified. The identification also arises from the fact that matrix $\boldsymbol{B}_P$ has full column-rank and its estimation can be achieved by a linear regression closely related to Lemma **L-7**. The estimation of the expected value of pre-program variables conditioned on response-types $E(X|\boldsymbol{S} = \boldsymbol{s}_j); j = 1, .., 7$ is obtained by parameter $\boldsymbol{\beta}$ in regression (57) by replacing the dependent variable $D_{t,\omega}$ for $X_\omega \cdot D_{t,\omega}$ and the exploratory variable $\boldsymbol{B}_{t,\omega}$ for $\boldsymbol{B}_{t,\omega} \cdot \widehat{\boldsymbol{P}}_S'$. See Appendix A.13 for a proof.

[25]The estimated probabilities always sum to one. However, the method does not impose that the probabilities are positive.

9 Addressing the Problem of Partial Identification

Theorem **T-3** lists nine counterfactual outcomes that are nonparamterically identified. Six of those are conditioned on a single response-type, while remaining three are conditioned on two response-types. This entails a partial identification problem. For instance, $E(Y(t_l)|\boldsymbol{S} \in \{\boldsymbol{s}_4, \boldsymbol{s}_6\})$ is identified but it cannot be disentangled into $E(Y(t_l)|\boldsymbol{S} = \boldsymbol{s}_4)$ and $E(Y(t_l)|\boldsymbol{S} = \boldsymbol{s}_6)$ without additional assumptions. This partial identification problem is typical to discrete instruments, which, in general, cannot identify the average treatment effects (ATE) without additional assumptions.[26]

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. Their methods build on the Local Instrumental Variables (LIV) model of Heckman and Vytlacil (1999). The LIV model expresses ATE as an integral of the marginal treatment effect (MTE) as written below:

$$ATE = \int_0^1 MTE(v)dv \quad \text{such that} \quad MTE(v) \equiv E(Y(t_1) - Y(t_0)|V = v), \tag{58}$$

where t_1, t_0 denotes the values of the binary choice T and $V \sim Unif[0,1]$ stands for a confounding variable that causes the choice T and the outcome Y. Continuous instruments over the full support enables the identification of the $MTE(v)$ function for v in the unit interval. Thereby the integral $\int_a^b MTE(v)dv$ is also identified over any interval $[a,b] \subset [0,1]$, and, in particular, ATE is identified. Discrete instruments, on the other hand, identify the integral $\int_a^b MTE(v)dv$ only for discrete intervals and therefore ATE is not identified.

Current literature uses extrapolations of the MTE function to overcome the limitations posed by discrete instruments. I apply the main ideas of this literature to a slight distinct setting. Instead of a binary treatment, I investigate the case of multiple choices. I do not seek to assess ATE, but simply to disentangle the identified counterfactual outcomes that are conditioned in two response-types. This is obtained via interpolation instead of extrapolation.

Section 9.1 extends the Local Instrumental Variables (LIV) of Heckman and Vytlacil (1999) to the case of multiple choices. To do so, I exploit Lemma **L-6** and the Separability Property **P-4**. Section 9.2 explains the identification problems that arises when we migrate from the binary choice model to the case of multiple choices. Section 9.3 exploits the Lonesum Property **P-3** to solve these additional identification problems.

9.1 Marginal Effects Under Unordered Monotonicity

This section builds on Heckman and Vytlacil (1999) to the examine multiple choice models under unordered monotonicity.[27] The Separability Property **P-4** states that the choice indicator can be

[26] An exception occurs when full compliance holds. For instance, if there exist instrumental values $z_0, z_1 \in \text{supp}(Z)$ and treatment choices $t_1, t_0 \in \text{supp}(T)$ such that $P(T = t_1|Z = z_1) = 1$ and $P(T = t_0|Z = z_0) = 1$, then $E(Y(t_1) - Y(t_0))$ is identified by $E(Y|Z = z_1) - E(Y|Z = z_0)$.

[27] See Appendix K for a review of the LIV model for the case of binary choices.

expressed by a separable equation. This property is restated below:

$$D_t = \mathbf{1}\left[\varphi_t(Z) \geq \tau_t(\boldsymbol{V})\right] = \mathbf{1}\left[F_{\tau_t(\boldsymbol{V})}(\varphi_t(Z)) \geq F_{\tau_t(\boldsymbol{V})}(\tau_t(\boldsymbol{V}))\right] = \mathbf{1}\left[P_t \geq U_t\right], \tag{59}$$

where variables P_t, U_t are transformations of $Z, \tau_t(\boldsymbol{V})$ defined by:

$$P_t(Z) \equiv F_{\tau_t(\boldsymbol{V})}(\varphi_t(Z)) \text{ and } U_t(\boldsymbol{V}) \equiv F_{\tau_t(\boldsymbol{V})}(\tau_t(\boldsymbol{V})), \tag{60}$$

where $F_{\tau_t(\boldsymbol{V})}(\cdot)$ denotes the CDF of $\tau_t(\boldsymbol{V})$. Two properties of U_t, P_t are that U_t has a uniform distribution $U_t \sim Unif[0,1]$ for an absolutely continuous $\tau_t(\boldsymbol{V})$ and that $P_t \perp\!\!\!\perp (U_t, Y(t))$ due to $Z \perp\!\!\!\perp (\boldsymbol{V}, Y(t))$ of (25). These two properties imply that $P_t(Z)$ in (60) plays the role of the propensity score, as shown below:

$$P\big(T = t|Z = z\big) = E\big(D_t|Z = z\big) = E\big(\mathbf{1}\left[P_t(Z) \geq U_t\right]|Z = z\big) = P\big(P_t(z) \geq U_t\big) = P_t(z), \tag{61}$$

where the third equality is due to $P_t \perp\!\!\!\perp U_t$ and the last equality is due to uniform distribution of U_t. Consider the instrumental value $z \in \text{supp}(Z)$ which correspond to the propensity score $P(T = t|Z = z) = p$. Thus the expectation $E(Y \cdot D_t|Z = z)$ can be stated as $E(Y \cdot D_t|P_t = p)$ as expressed by the following equations:

$$E\Big(Y \cdot D_t|P_t = p\Big) = E\Big(Y(t) \cdot \mathbf{1}[P_t \geq U_t]|P_t = p\Big) = E\Big(Y(t) \cdot \mathbf{1}[p \geq U_t]\Big) = \int_0^p E\big(Y(t)|U_t = u\big)du. \tag{62}$$

The first equality in (62) uses separability (18), the second is due to $P_t \perp\!\!\!\perp (Y(t), U_t)$ and the last one is due to uniformity of U_t. A consequence of equation (62) is that the counterfactual outcome mean $E(Y(t))$ can be identified by the derivative of the observed outcome w.r.t. the propensity score P_t :

Theorem T-5. Let P_t be the propensity score of choice $t \in \text{supp}(T)$ and $E\big(Y \cdot D_t|P_t = p\big)$ be a.e. differentiable w.r.t. to $p \in [0,1]$. If unordered monotonicity holds, then $E(Y(t)|U_t = u)$ is identified for u at value p by $E(Y(t)|U_t = p) = \partial E\big(Y \cdot D_t|P_t = p\big)/\partial p$. Moreover, if $E\big(Y \cdot D_t|P_t = p\big)$ is differentiable for all $p \in [0,1]$ (full support) then the average counterfactual outcomes $E(Y(t))$ can be evaluated by:

$$E(Y(t)) = \int_0^1 E(Y(t)|U_t = u)du, \text{ where } E(Y(t)|U_t = u) = \left.\frac{\partial E\Big(Y \cdot D_t|P_t = p\Big)}{\partial p}\right|_{p=u} \tag{63}$$

Proof. See Appendix A.14. □

T-5 states that if the instrument Z ensures enough variation around $P_t(Z) = u$, then $E(Y(t)|U_t = u)$ is identified. If Z ensures enough variation over the full support of $P_t(Z)$, then $E(Y(t)|U_t = u)$ is identified for all $u \in [0,1]$ and $E(Y(t))$ is obtained by integrating $E(Y(t)|U_t = u)$ over the unity interval.[28]

The primary feature of **T-5** is that the identification of $E(Y(t))$ depends only on the propensity score of choice t and *not* on the propensity scores of remaining choices.

Remark 9.1. The recent work of Lee and Salanié (2018) offers substantial contributions to the literature on the identification of counterfactual outcomes in multivalued treatment models. They investigate a general IV setting with N_T treatment choices that are determined by an arbitrary set

[28] In particular, if $E(Y \cdot D_t|P_t = p)$ is expressed by a polynomial $E(Y \cdot D_t|P_t = p) = \sum_{k=1}^{K} \beta_{k,t} \cdot p^k$, then the counterfactual outcome expectation $E(Y(t))$ is given by $E(Y(t)) = \sum_{k=1}^{K} \beta_{k,t}$.

separable equations. Unordered monotonicity consists of a subfamily of this class of models. Lee and Salanié (2018) show that the identification of a counterfactual outcome mean $E(Y(t))$ generally requires a N_T-th order derivative across the propensity scores of all N_T choices. ***T-5*** *states that in the particular case of unordered monotonicity, the identification of a counterfactual outcome mean $E(Y(t))$ requires only a first-order derivative w.r.t the propensity score of choice t. This facilities the estimation of causal effects as it avoids the evaluation of higher order derivatives that can be empirically cumbersome.*

Discrete instruments do not render the identification of $E(Y(t)|U_t = u)$ as the support of Z is finite. Consider two instrumental values $z, z' \in \text{supp}(Z)$ corresponding to propensity scores $P_t(z) = p, P_t(z') = p'$ for choice t where $p > p'$. According to equation (62), we have that:

$$\frac{E\big(Y \cdot D_t|P_t = p\big) - E\big(Y \cdot D_t|P_t = p'\big)}{p - p'} = \frac{\int_{p'}^{p} E(Y(t)|U_t = u)du}{p - p'} = E(Y(t)|p \leq U_t \leq p'). \tag{64}$$

Lemma **L-6** employs the Nested Property **P-4** to identify the counterfactual outcome $Y(t)$ conditioned on the response-types in $\Sigma_t(z) \setminus \Sigma_t(z')$. Lemma **L-6** can be restated using (64) as:

$$\begin{aligned}
E(Y(t)|\boldsymbol{S} \in \Sigma_t(z) \setminus \Sigma_t(z')) &= \frac{E(Y\mathbf{1}[T = t]|Z = z) - E(Y\mathbf{1}[T = t_h]|Z = z')}{E(\mathbf{1}[T = t]|Z = z) - E(\mathbf{1}[T = t]|Z = z')} && (65)\\
&\equiv \frac{E\big(Y \cdot D_t|P_t = P_t(z)\big) - E\big(Y \cdot D_t|P_t = P_t(z')\big)}{P_t(z) - P_t(z')} && (66)\\
&= \frac{\int_0^{P_t(z)} E(Y(t)|U_t = u)du - \int_0^{P_t(z')} E(Y(t)|U_t = u)du}{P_t(z) - P_t(z')} && (67)\\
&= \frac{\int_{P_t(z')}^{P_t(z)} E(Y(t)|U_t = u)du}{P_t(z) - P_t(z')} && (68)\\
&\equiv g_t(P_t(z'), P_t(z)), \quad \text{such that } P_t(z') < P_t(z). && (69)
\end{aligned}$$

Equation (68) states that each identified counterfactual outcome $E(Y(t)|\boldsymbol{S} \in \Sigma_t(z) \setminus \Sigma_t(z'))$ can be expressed as an integral of $E(Y(t)|U_t = u)$ over the interval $[P_t(z'), P_t(z)]$ which depends only on the propensity scores of choice t. Equation (69) simply emphasizes that the integral in (68) can be expressed as a function of two probabilities: $P_t(z')$ and $P_t(z)$.

9.2 Understanding the Problem of Partial Identification

Equations (64)–(69) are useful to characterize the identification problem of disentangling the counterfactual outcomes. Consider the case of $E(Y(t_l)|\boldsymbol{S} \in \{\boldsymbol{s}_4, \boldsymbol{s}_6\})$. This counterfactual outcome can also be expressed as $E(Y(t_l)|\boldsymbol{S} \in \Sigma_{t_l}(z_e) \setminus \Sigma_{t_l}(z_8))$ and, according to (68), it can be expressed as an integral of $E(Y(t)|U_t = u)du$ over the interval $[P_{t_l}(z_8), P_{t_l}(z_e)]$. This integral can be represented as a function $g_{t_l}(\cdot, \cdot)$ evaluated at the propensity score boundaries, that is, $g_{t_l}\big(P_{t_l}(z_8), P_{t_l}(z_e)\big)$. Equation (70) summarizes this information.

$$E(Y(t_l)|\boldsymbol{S} \in \{\boldsymbol{s}_4, \boldsymbol{s}_6\}) = \underbrace{\overbrace{E(Y(t_l)|\boldsymbol{S} \in \Sigma_{t_l}(z_e) \setminus \Sigma_{t_l}(z_8))}^{\text{Identified Counterfactual T-3}} = \overbrace{\frac{\int_{P_{t_l}(z_8)}^{P_{t_l}(z_e)} E(Y(t_l)|U_{t_l}=u)du}{P_{t_l}(z_e)-P_{t_l}(z_8)}}^{\text{As in Equation (69)}}}_{\text{By Lemma L-6 \& Equation (64)}} \equiv g_{t_l}\Big(P_{t_l}(z_8), P_{t_l}(z_e)\Big) \tag{70}$$

Our goal is to disentangle $E(Y(t_l)|\boldsymbol{S} \in \{\boldsymbol{s}_4, \boldsymbol{s}_6\})$ into $E(Y(t_l)|\boldsymbol{S} = \boldsymbol{s}_4)$ and $E(Y(t_l)|\boldsymbol{S} = \boldsymbol{s}_6)$. Under unordered monotonicity, the extended LIV model of Section 9.1 holds. This implies that $E(Y(t_l)|\boldsymbol{S} = \boldsymbol{s}_4)$ and $E(Y(t_l)|\boldsymbol{S} = \boldsymbol{s}_6)$ can also be expressed as integrals of $E(Y(t)|U_t = u)du$ over two contiguous and non-overlapping intervals. Moreover, the union of these two intervals must be equal to $[P_{t_l}(z_8), P_{t_l}(z_e)]$. Otherwise stated, the problem of identifying $E(Y(t_l)|\boldsymbol{S} = \boldsymbol{s}_4)$ and $E(Y(t_l)|\boldsymbol{S} = \boldsymbol{s}_6)$ consists of determining the propensity score $p^*_{t_l} \in (P_{t_l}(z_8), P_{t_l}(z_e))$ that splits the interval $[P_{t_l}(z_8), P_{t_l}(z_e)]$ into two integration intervals $[P_{t_l}(z_8), p^*_{t_l}]$ and $[p^*_{t_l}, P_{t_l}(z_e)]$ such that $g_{t_l}\big(P_{t_l}(z_8), p^*_{t_l}\big)$ and $g_{t_l}\big(p^*_{t_l}, P_{t_l}(z_e)\big)$ identify the targeted counterfactual outcomes. Notationally we have that:

$$E(Y(t_l)|\boldsymbol{S} = \boldsymbol{s}') = \frac{\int_{P_{t_l}(z_8)}^{P^*_{t_l}} E(Y(t_l)|U_{t_l} = u)du}{P^*_{t_l} - P_{t_l}(z_8)} \equiv g_{t_l}\Big(P_{t_l}(z_8), p^*_{t_l}\Big), \tag{71}$$

$$E(Y(t_l)|\boldsymbol{S} = \boldsymbol{s}'') = \frac{\int_{p^*_{t_l}}^{P_{t_l}(z_e)} E(Y(t_l)|U_{t_l} = u)du}{P_{t_l}(z_e) - p^*_{t_l}} \equiv g_{t_l}\Big(p^*_{t_l}, P_{t_l}(z_e)\Big), \tag{72}$$

$$\text{where } (\boldsymbol{s}', \boldsymbol{s}'') = (\boldsymbol{s}_4, \boldsymbol{s}_6) \text{ or } (\boldsymbol{s}', \boldsymbol{s}'') = (\boldsymbol{s}_6, \boldsymbol{s}_4). \tag{73}$$

There are two identification challenges in equations (71)–(73) that must be overcome. The first one is that the probability $p^*_{t_l} \in (P_{t_l}(z_8), P_{t_l}(z_e))$ is unknown. The second one is that even if the probability $p^*_{t_l}$ were known, an uncertainty exists as we cannot determine which of the two intervals $[P_{t_l}(z_8), p^*_{t_l}], [p^*_{t_l}, P_{t_l}(z_e)]$ identify each of the counterfactual outcomes $E(Y(t_l)|\boldsymbol{S} = \boldsymbol{s}_4), E(Y(t_l)|\boldsymbol{S} = \boldsymbol{s}_6)$. Next section explains how the Lonesum Property **P-3** helps in solving both identification challenges.

9.3 Solving the Problem of Partial Identification

The partial identification problem of discrete instruments stems from the lack of variability of the instrumental variable. The matrices in (74) help to illustrate this fact. The first matrix lists the response-types in the response-matrix $\boldsymbol{R}$ in which t_l appears. Those are $\boldsymbol{s}_3, \boldsymbol{s}_4, \boldsymbol{s}_5, \boldsymbol{s}_6$. The second matrix indicates the t_l-choices. The matrix consists of the columns $\boldsymbol{s}_3, \boldsymbol{s}_4, \boldsymbol{s}_5, \boldsymbol{s}_6$ of $\boldsymbol{B}_{t_l}$. The third matrix reorders these columns into a triangular matrix.

$$
\underbrace{\begin{array}{c} \\ z_c \\ z_8 \\ z_e \end{array}\begin{array}{c}\begin{array}{cccc} \boldsymbol{s}_3 & \boldsymbol{s}_4 & \boldsymbol{s}_5 & \boldsymbol{s}_6 \end{array}\\ \begin{bmatrix} t_l & t_h & t_h & t_m \\ t_l & t_m & t_l & t_m \\ t_l & t_l & t_l & t_l \end{bmatrix}\end{array}}_{\text{Selection of } \boldsymbol{R}} \Rightarrow \underbrace{\begin{array}{c}\begin{array}{cccc} \boldsymbol{s}_3 & \boldsymbol{s}_4 & \boldsymbol{s}_5 & \boldsymbol{s}_6 \end{array}\\ \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 1 \end{bmatrix}\end{array}}_{\text{Columns of } \boldsymbol{B}_{t_l}} \Rightarrow \underbrace{\begin{array}{c}\begin{array}{cccc} \boldsymbol{s}_3 & \boldsymbol{s}_5 & \boldsymbol{s}_4 & \boldsymbol{s}_6 \end{array}\\ \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 \end{bmatrix}\end{array}}_{\text{Reordered } \boldsymbol{B}_{t_l} \text{ Columns}} \therefore \begin{array}{l} \Sigma_{t_l}(z_c) = \{\boldsymbol{s}_3\} \\ \Sigma_{t_l}(z_8) = \{\boldsymbol{s}_3, \boldsymbol{s}_5\} \\ \Sigma_{t_l}(z_e) = \{\boldsymbol{s}_3, \boldsymbol{s}_5, \boldsymbol{s}_4, \boldsymbol{s}_6\} \end{array} \tag{74}
$$

The first row of the matrices in (74) correspond to the instrumental value z_c. We observed that under z_c, the response-type $\boldsymbol{s}_3$ is the only one that takes value t_l. Under z_8 (second row) response-types $\boldsymbol{s}_3$ and $\boldsymbol{s}_5$ take value t_l. The difference of response-type sets between z_8 and z_c is $\Sigma_{t_l}(z_8) \setminus \Sigma_{t_l}(z_c) = \{\boldsymbol{s}_5\}$ and thereby $E(Y(t_l)|\boldsymbol{S} = \boldsymbol{s}_5)$ is identified (Lemma **L-6**). Partial identification arises because the difference between response-type sets between z_8 and z_e is $\Sigma_{t_l}(z_e) \setminus \Sigma_{t_l}(z_8) = \{\boldsymbol{s}_4, \boldsymbol{s}_6\}$. Thereby $E(Y(t_l)|\boldsymbol{S} \in \{\boldsymbol{s}_4, \boldsymbol{s}_6\})$ is identified but not its counterparts $E(Y(t_l)|\boldsymbol{S} = \boldsymbol{s}_4)$ and $E(Y(t_l)|\boldsymbol{S} = \boldsymbol{s}_6)$.

This identification problem would be solved if there were an instrumental value z^* that bridges the response-types between $\Sigma_{t_l}(z_8)$ and $\Sigma_{t_l}(z_e)$. Notationally it means that $\Sigma_{t_l}(z_8) \subset \Sigma_{t_l}(z^*) \subset \Sigma_{t_l}(z_8)$ such that the set differences $\Sigma_{t_l}(z_e) \setminus \Sigma_{t_l}(z^*)$ and $\Sigma_{t_l}(z^*) \setminus \Sigma_{t_l}(z_8)$ would render the response-types $\boldsymbol{s}_4$ and $\boldsymbol{s}_6$. For these conditions to be satisfied, it must be the case that $\Sigma_{t_l}(z^*) = \{\boldsymbol{s}_3, \boldsymbol{s}_5, \boldsymbol{s}_4\}$ or $\Sigma_{t_l}(z^*) = \{\boldsymbol{s}_3, \boldsymbol{s}_5, \boldsymbol{s}_6\}$. In summary, the additional variation z^* of the instrument Z that enables the identification of $E(Y(t_l)|\boldsymbol{S} = \boldsymbol{s}_4)$ and $E(Y(t_l)|\boldsymbol{S} = \boldsymbol{s}_6)$ is such that response-types $\boldsymbol{s}_3, \boldsymbol{s}_5$ that take value t_l for z^* (same for z_8 and z_e) but either $\boldsymbol{s}_6$ or $\boldsymbol{s}_4$ may take value t_l. Matrix (75) illustrates this condition by inserting a z^*-row between the rows z_8, z_e of the response-types displayed in (74).

$$
\underbrace{\begin{array}{c} \\ z_c \\ z_8 \\ z^* \\ z_e \end{array}\begin{array}{c}\begin{array}{cccc} \boldsymbol{s}_3 & \boldsymbol{s}_4 & \boldsymbol{s}_5 & \boldsymbol{s}_6 \end{array}\\ \begin{bmatrix} t_l & t_h & t_h & t_m \\ t_l & t_m & t_l & t_m \\ t_l & ? & t_l & ? \\ t_l & t_l & t_l & t_l \end{bmatrix}\end{array}}_{\text{New Instrument Value } z^*} \Rightarrow \underbrace{\overbrace{\begin{array}{c}\begin{array}{cccc} \boldsymbol{s}_3 & \boldsymbol{s}_4 & \boldsymbol{s}_5 & \boldsymbol{s}_6 \end{array}\\ \begin{bmatrix} t_l & t_h & t_h & t_m \\ t_l & t_m & t_l & t_m \\ t_l & t_l & t_l & t_m \\ t_l & t_l & t_l & t_l \end{bmatrix}\end{array}}^{\boldsymbol{s}_4\text{-shift}}}_{\substack{\Sigma_{t_l}(z^*) = \{\boldsymbol{s}_3, \boldsymbol{s}_5, \boldsymbol{s}_4\} \\ \text{Unordered monotonicity holds}}} \text{ or } \underbrace{\overbrace{\begin{array}{c}\begin{array}{cccc} \boldsymbol{s}_3 & \boldsymbol{s}_4 & \boldsymbol{s}_5 & \boldsymbol{s}_6 \end{array}\\ \begin{bmatrix} t_l & t_h & t_h & t_m \\ t_l & t_m & t_l & t_m \\ t_l & t_m & t_l & t_l \\ t_l & t_l & t_l & t_l \end{bmatrix}\end{array}}^{\boldsymbol{s}_6\text{-shift}}}_{\substack{\Sigma_{t_l}(z^*) = \{\boldsymbol{s}_3, \boldsymbol{s}_5, \boldsymbol{s}_6\} \\ \text{Unordered monotonicity does \textbf{not} hold}}} \tag{75}
$$

Matrix (75) shows that the instrument change from z_8 to z_e induces *both* the response-types $\boldsymbol{s}_4, \boldsymbol{s}_6$ to shift from t_m to t_l. The pseudo instrumental value z^* bridges z_8–z_e by allowing for just one shift. Some uncertainty exists as there are two possible cases: shifting the neighborhood choice in $\boldsymbol{s}_4$ implies in $\Sigma_{t_l}(z^*) = \{\boldsymbol{s}_3, \boldsymbol{s}_5, \boldsymbol{s}_4\}$ while shifting $\boldsymbol{s}_6$ implies $\Sigma_{t_l}(z^*) = \{\boldsymbol{s}_3, \boldsymbol{s}_5, \boldsymbol{s}_6\}$. The Lonesum Property **P-3** solves this uncertainty by ruling out one of these cases.

The Lonesum Property **P-3** prohibits the shift of the neighborhood choice in response-type $\boldsymbol{s}_6$. The 2×2 sub-matrix generates by rows (z_c, z^*) and columns $(\boldsymbol{s}_4, \boldsymbol{s}_6)$ has t_m in its anti-diagonal but there is no t_m in its diagonal, which violates **P-3**. On the other hand, Property **P-3** holds for the $\boldsymbol{s}_4$-shift. We then conclude that $\Sigma_{t_l}(z^*) = \{\boldsymbol{s}_3, \boldsymbol{s}_5, \boldsymbol{s}_4\}$. This analysis solves the two identification challenges posted in the previous section.

The first challenge is to determine the probability $p^*_{t_l}$ that split the interval $[P_{t_l}(z_8), P_{t_l}(z_e)]$. and sets the boundaries of the integration for $E(Y(t_l)|\boldsymbol{S} = \boldsymbol{s}_4)$, $E(Y(t_l)|\boldsymbol{S} = \boldsymbol{s}_6)$. This probability is given by the propensity score for z^*, which is equal to the following response-type probabilities:

$$P_{t_l}(z^*) \equiv P(T = t_l|Z = z^*) = P(\boldsymbol{S} \in \Sigma_{t_l}(z^*)) = P(\boldsymbol{S} \in \{\boldsymbol{s}_3, \boldsymbol{s}_5, \boldsymbol{s}_4\}).$$

In the same token, we have that the propensity scores $P_{t_l}(z_e), P_{t_l}(z_8)$ identify the following response-types probabilities:

$$P_{t_l}(z_8) = P(\boldsymbol{S} \in \Sigma_{t_l}(z_8)) = P(\boldsymbol{S} \in \{\boldsymbol{s}_3, \boldsymbol{s}_5\}) \text{ and } P_{t_l}(z_e) = P(\boldsymbol{S} \in \Sigma_{t_l}(z_e)) = P(\boldsymbol{S} \in \{\boldsymbol{s}_3, \boldsymbol{s}_5, \boldsymbol{s}_4, \boldsymbol{s}_6\}). \tag{76}$$

Therefore we have that $P_{t_l}(z^*) \in (P_{t_l}(z_8), P_{t_l}(z_e))$, as desired.

The second challenge refers to the uncertainty in assigning integration intervals to the counterfactual outcomes. This uncertainty is resolved as $\Sigma_{t_l}(z^*)\backslash P_{t_l}(z_8) = \{\boldsymbol{s}_4\}$ and $\Sigma_{t_l}(z_e)\backslash P_{t_l}(z^*) = \{\boldsymbol{s}_6\}$. Thereby $E(Y(t_l)|\boldsymbol{S} = \boldsymbol{s}_4)$ is identified by integration over $[P_{t_l}(z_8), P_{t_l}(z^*)]$, and $E(Y(t_l)|\boldsymbol{S} = \boldsymbol{s}_4)$ by integration over $[P_{t_l}(z^*), P_{t_l}(z_e)]$, that is:

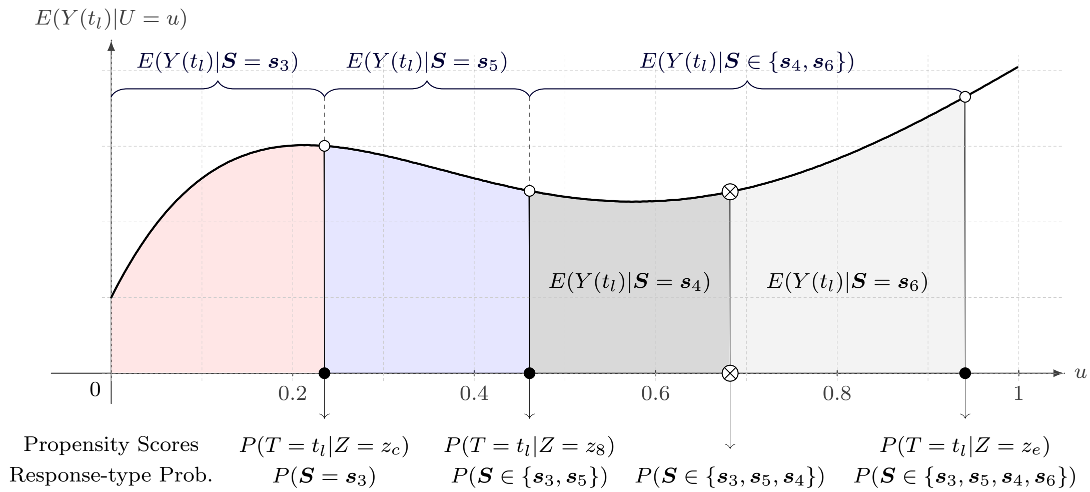
$$E(Y(t_l)|\boldsymbol{S} = \boldsymbol{s}_4) = \frac{\int_{P_{t_l}(z_8)}^{P_{t_l}(z^*)} E(Y(t_l)|U_{t_l} = u)du}{P_{t_l}(z^*) - P_{t_l}(z_8)} \equiv g_{t_l}\Big(P_{t_l}(z_8), P_{t_l}(z^*)\Big), \tag{77}$$

$$E(Y(t_l)|\boldsymbol{S} = \boldsymbol{s}_6) = \frac{\int_{P_{t_l}(z^*)}^{P_{t_l}(z_e)} E(Y(t_l)|U_{t_l} = u)du}{P_{t_l}(z_e) - P_{t_l}(z^*)} \equiv g_{t_l}\Big(P_{t_l}(z^*), P_{t_l}(z_e)\Big), \tag{78}$$

$$\text{where } P_{t_l}(z^*) = P(\boldsymbol{S} \in \{\boldsymbol{s}_3, \boldsymbol{s}_5, \boldsymbol{s}_4\}). \tag{79}$$

Figure 3 represents the identification of $E(Y(t_l)|\boldsymbol{S} = \boldsymbol{s}_4)$ and $E(Y(t_l)|\boldsymbol{S} = \boldsymbol{s}_6)$ as a graph. While $P_{t_l}(z_e), P_{t_l}(z_8)$ are obtained directly from observed data, $P_{t_l}(z^*)$ is not. However, $P_{t_l}(z^*)$ is identified as all response-type probabilities are identified (**T-3**) and it can be evaluated by **L-7**. The values $g_{t_l}(p^*_{t_l}, P_{t_l}(z_8))$ and $g_{t_l}(P_{t_l}(z_e), p^*_{t_l})$ are obtained by the interpolation the function $\int_0^p E(Y(t_l)|U_{t_l} = u)du$ as discussed in the next section.

Figure 3: Identification of Counterfactual Outcomes $E(Y(t_l)|\boldsymbol{S}=\boldsymbol{s}_4)$ and $E(Y(t_l)|\boldsymbol{S}=\boldsymbol{s}_6)$



This figure summarises the identification strategy to disentangle $E(Y(t_l)|\boldsymbol{S}\in\{\boldsymbol{s}_4,\boldsymbol{s}_6\})$ into $E(Y(t_l)|\boldsymbol{S}=\boldsymbol{s}_4)$ and $E(Y(t_l)|\boldsymbol{S}=\boldsymbol{s}_6)$. The Extended LIV model (Section 9.1) shows that $E(Y(t_l)|\boldsymbol{S}\in\{\boldsymbol{s}_4,\boldsymbol{s}_6\})$ can be expressed as an integral of $E(Y(t_l)|U_{t_l}=u)$ over and interval which depends only on the propensity scores of choice t_l. This interval is given by $[P_{t_l}(z_8), P_{t_l}(z_e)]$ where $P_{t_l}(z_8) \equiv P(T=t_l|Z=z_8)$ and $P_{t_l}(z_e) \equiv P(T=t_l|Z=z_e)$. The Lonesum Property **P-3** implies $E(Y(t_l)|\boldsymbol{S}=\boldsymbol{s}_4)$ is identified by integration of $E(Y(t_l)|U_{t_l}=u)$ over $[P_{t_l}(z_8), P_{t_l}(z^*)]$, and $E(Y(t_l)|\boldsymbol{S}=\boldsymbol{s}_4)$ by integration of $E(Y(t)|U_t=u)$ over $[P_{t_l}(z^*), P_{t_l}(z_e)]$, where $P_{t_l}(z_8), P_{t_l}(z_e)$ identify $P(\boldsymbol{S}=\{\boldsymbol{s}_3,\boldsymbol{s}_5\}), P(\boldsymbol{S}=\{\boldsymbol{s}_3,\boldsymbol{s}_5,\boldsymbol{s}_4,\boldsymbol{s}_8\})$ and $P_{t_l}(z^*)$ is given by $P(\boldsymbol{S}=\{\boldsymbol{s}_3,\boldsymbol{s}_5,\boldsymbol{s}_4\})$ which is also identified (**T-3**).

The identification analysis that disentangles the counterfactual outcome $E(Y(t_l)|\boldsymbol{S}\in\{\boldsymbol{s}_4,\boldsymbol{s}_6\})$ also applies to $E(Y(t_h)|\boldsymbol{S}\in\{\boldsymbol{s}_4,\boldsymbol{s}_5\})$ and $E(Y(t_m)|\boldsymbol{S}\in\{\boldsymbol{s}_4,\boldsymbol{s}_7\})$. Table 10 focus on choice t_h and presents the identification formulas that disentangle $E(Y(t_h)|\boldsymbol{S}\in\{\boldsymbol{s}_4,\boldsymbol{s}_5\})$ into $E(Y(t_h)|\boldsymbol{S}=\boldsymbol{s}_4)$ and $E(Y(t_h)|\boldsymbol{S}=\boldsymbol{s}_5)$. Table 11 focus on choice t_m and presents the respective identification equations.

Table 10: Identification Formulas for Counterfactual Means $E(Y(t_h)|\boldsymbol{S}=\boldsymbol{s}_4)$ and $E(Y(t_h)|\boldsymbol{S}=\boldsymbol{s}_5)$

Counterfactual Outcome	Integral Representation	Function of Propensity Scores
$E(Y(t_h)|\boldsymbol{S}\in\{\boldsymbol{s}_4,\boldsymbol{s}_5\})=$	$\frac{\int_{P_{t_h}(z_e)}^{P_{t_h}(z_c)} E(Y(t_h)|U_{t_h}=u)du}{P_{t_h}(z_c)-P_{t_h}(z_e)}$	$\equiv g_{t_h}\Big(P_{t_h}(z_c), P_{t_h}(z_e)\Big)$
$E(Y(t_h)|\boldsymbol{S}=\boldsymbol{s}_4)=$	$\frac{\int_{P_{t_h}(z_e)}^{p^*_{t_h}} E(Y(t_h)|U_{t_h}=u)du}{p^*_{t_h}-P_{t_h}(z_e)}$	$\equiv g_{t_h}\Big(p^*_{t_h}, P_{t_h}(z_e)\Big)$
$E(Y(t_h)|\boldsymbol{S}=\boldsymbol{s}_5)=$	$\frac{\int_{p^*_{t_h}}^{P_{t_h}(z_c)} E(Y(t_h)|U_{t_h}=u)du}{P_{t_h}(z_c)-p^*_{t_h}}$	$\equiv g_{t_h}\Big(P_{t_h}(z_c), p^*_{t_h}\Big)$
where $p^*_{t_h}$	$=P(\boldsymbol{S}\in\{\boldsymbol{s}_1,\boldsymbol{s}_7,\boldsymbol{s}_4\})\in\Big(P_{t_h}(z_e), P_{t_h}(z_c)\Big)$	
because $P_{t_h}(z_e)$	$=P(\boldsymbol{S}\in\{\boldsymbol{s}_1,\boldsymbol{s}_7\})$ and $P_{t_h}(z_c)=P(\boldsymbol{S}\in\{\boldsymbol{s}_1,\boldsymbol{s}_7,\boldsymbol{s}_4,\boldsymbol{s}_5\})$	

Table 11: Identification Formulas for Counterfactual Means $E(Y(t_m)|\boldsymbol{S}=\boldsymbol{s}_4)$ and $E(Y(t_m)|\boldsymbol{S}=\boldsymbol{s}_7)$

Counterfactual Outcome	Integral Representation	Function of Propensity Scores
$E(Y(t_m)\|\boldsymbol{S}\in\{\boldsymbol{s}_4,\boldsymbol{s}_7\})=$	$\frac{\int_{P_{t_m}(z_8)}^{P_{t_m}(z_c)} E(Y(t_m)\|U_{t_m}=u)du}{P_{t_m}(z_8)-P_{t_m}(z_c)}$	$\equiv g_{t_m}\Big(P_{t_m}(z_8),P_{t_m}(z_c)\Big)$
$E(Y(t_m)\|\boldsymbol{S}=\boldsymbol{s}_4)=$	$\frac{\int_{P_{t_m}(z_8)}^{p^*_{t_m}} E(Y(t_m)\|U_{t_m}=u)du}{p^*_{t_m}-P_{t_m}(z_c)}$	$\equiv g_{t_m}\Big(p^*_{t_m},P_{t_m}(z_8)\Big)$
$E(Y(t_m)\|\boldsymbol{S}=\boldsymbol{s}_7)=$	$\frac{\int_{p^*_{t_m}}^{P_{t_m}(z_8)} E(Y(t_m)\|U_{t_m}=u)du}{P_{t_m}(z_8)-p^*_{t_m}}$	$\equiv g_{t_m}\Big(P_{t_m}(z_8),p^*_{t_m}\Big)$
where $p^*_{t_m}$	$=P(\boldsymbol{S}\in\{\boldsymbol{s}_2,\boldsymbol{s}_6,\boldsymbol{s}_4\})\in\Big(P_{t_m}(z_c),P_{t_m}(z_8)\Big)$	
because $P_{t_m}(z_e)$	$=P(\boldsymbol{S}\in\{\boldsymbol{s}_2,\boldsymbol{s}_6\})$ and $P_{t_m}(z_c)=P(\boldsymbol{S}\in\{\boldsymbol{s}_2,\boldsymbol{s}_6,\boldsymbol{s}_4,\boldsymbol{s}_7\})$	

9.4 Estimating the Counterfactual Outcomes

Similar to Section 8.1, I show that the evaluation of the counterfactual outcome discussed in the previous section can be achieved through simple econometric techniques. Indeed, all counterfactual outcomes and causal effects of this paper are estimated by 2SLS regressions applied to particular transformations of the observed data.

Theorem **T-6** state that counterfactual outcome means can be estimated by a ratio of least square estimates. The result is motivated by a control function approach that uses local polynomials of the propensity score to correct for endogeneity.[29] I adopt a slightly more general setting where $H_t(Z)$ is a function of the instrumental variable that would represent propensity scores in the control function approach while $\boldsymbol{\lambda}(H_t)$ would represent the polynomial functions. Specifically, let the instrument Z takes N_Z values in $\{z_1,...,Z_{N_Z}\}$ and choice T takes values N_T values in $\{t_1,...,t_{N_T}\}$. For each $t\in\text{supp}(T)$, let $H_t(z)$ be any injection function $H_t:\text{supp}(Z)\to\mathbb{R}$ whose image is denoted by $h_{t,i}\equiv H_t(z_i); i=1,...,N_Z$. Let $\boldsymbol{\lambda}(h)=\big[\lambda_1(h),...,\lambda_{N_Z}(h)\big]'$ be any vector of N_Z linearly independent[30] real-valued functions. [31] In this notation, the expectations $E(Y\cdot D_t|Z=z_i)$ and $E(D_t|Z=z_i)$ for $i=1,...,N_Z$ can always be expressed as:

$$E(Y\cdot D_t|Z=z_i) \;=\; E(Y\cdot D_t|H_t(Z)=h_{t,i}) \;=\; \boldsymbol{\lambda}(h_{t,i})'\boldsymbol{\beta}_t, \tag{80}$$
$$E(D_t|Z=z_i) \;=\; E(\,D_t\,|H_t(Z)=h_{t,i}) \;=\; \boldsymbol{\lambda}(h_{t,i})'\boldsymbol{\theta}_t, \tag{81}$$

where $\boldsymbol{\beta}_t,\boldsymbol{\theta}_t$ are N_Z-dimensional vectors determined by the unique solution of the linear system generated by equations (80)–(81) across N_Z instrumental values. Let $D_{t,\omega}$ be the choice indicator for family ω and $\boldsymbol{\lambda}_{t,\omega}\equiv\boldsymbol{\lambda}(H_t(Z_\omega))$ be the N_Z-dimensional vector $\boldsymbol{\lambda}$ associated with family ω.

[29] See Appendix N for the connection with the control function approach.

[30] By linearly independent I mean that the square matrix generated by concatenating vectors $\boldsymbol{\lambda}(H_t(z_i))$ for the $i\in\{1,...,N_Z\}$ has full rank, that is to say that $\det([\boldsymbol{\lambda}(H_t(z_1)),...,\boldsymbol{\lambda}(H_t(z_{N_Z}))])\neq 0$.

[31] An example of function $H_t(Z)$ is the propensity score $P_t(z); z\in\text{supp}(Z)$ defined by $P_t(z)\equiv P(D_t|Z=z)$ where $D_t\equiv\mathbf{1}[\mathbf{T}=\mathbf{t}]$ is a treatment indicator. An example of function $\boldsymbol{\lambda}(h)$ for $N_Z=3$ is the third degree polynomial $\boldsymbol{\lambda}(h)=\big[h,h^2,h^3\big]'$.

Theorem T-6. *Let $t \in \text{supp}(T)$ be any treatment choice and $z, z' \in \text{supp}(Z)$ be two instrumental values. If unordered monotonicity (13) holds, then the counterfactual outcome $\Lambda_t(z,z') = E(Y(t)|\boldsymbol{S} \in \Sigma_t(z) \oplus \Sigma_t(z'))$ is identified (**L-6**) and can be estimated by:*[32]

$$\widehat{\Lambda}_t(z,z') = \frac{\Big(\boldsymbol{\lambda}(h_t) - \boldsymbol{\lambda}(h'_t)\Big)'\widehat{\boldsymbol{\beta}}_t}{\Big(\boldsymbol{\lambda}(h_t) - \boldsymbol{\lambda}(h'_t)\Big)'\widehat{\boldsymbol{\theta}}_t}, \tag{82}$$

where $h_t = H_t(z)$, $h'_t = H_t(z')$ and $\widehat{\boldsymbol{\beta}}_t, \widehat{\boldsymbol{\theta}}_t$ are the estimates of the following least squares regressions:

$$D_{t,\omega} = \boldsymbol{\lambda}_{t,\omega}\boldsymbol{\theta}_t + \epsilon_{\omega,D} \quad \text{ and } \quad Y_\omega \cdot D_{t,\omega} = \boldsymbol{\lambda}_{t,\omega}\boldsymbol{\beta}_t + \epsilon_{\omega,Y}, \tag{83}$$

where $\epsilon_{\omega,Y}, \epsilon_{\omega,D}$ denote error terms.

Proof. See Appendix A.15. □

The estimator (82) can be understood as a generalized LATE estimator for unordered monotonicity. The numerator of (82) estimates the expected difference $E(Y \cdot D_t|Z = z) - E(Y \cdot D_t|Z = z')$ while the numerator estimates the propensity score difference $P(T = t|Z = z) - P(T = t|Z = z')$.

Remark 9.2. ***T-6*** *characterizes a broad class of estimators for multiple choice models in which unordered monotonicity holds. The estimator applies to all identified counterfactual outcomes. It also applies to any number of choices and for an arbitrary number of instrumental values. In particular, it applies to binary choice models with categorical instrumental variables examined in Brinch et al. (2017) and Kline and Walters (2017).*

Remark 9.3. ***T-6*** *applies to any choice of injection function $H_t(z)$ and any choice of linearly independent functions in $\boldsymbol{\lambda}(h)$. Moreover, estimates $\widehat{\Delta}_t(z,z')$ are numerically the same regardless of the choice of $H_t(z)$ or $\boldsymbol{\lambda}(h)$ (see See Appendix A.15 for proof). Estimates $\widehat{\Delta}_t(z,z')$ are also numerically identical to the 2SLS estimates in* ***T-4****.*

Remark 9.4. *Estimator (82) can be expressed in terms of known matrices $\boldsymbol{B}_t, \boldsymbol{A}_t, \boldsymbol{C}_t$ and it can be understood as the empirical counterpart of the identifying equations (38) of Section 7.1 (see Appendix M). The estimator can also be interpreted as an example of the control function approach (see Appendix N).*

Remark 9.5. *Theorem* ***T-7*** *in Appendix P.2 shows that the counterfactual outcomes in (85)–(86) can be evaluated by a Three Stage Least Square Regression (3SLS) that fist performs a prediction of instrumental variable indicators which are in turn used as the instrumental variable of an standard 2SLS regression.*

Estimator **T-6** is useful to evaluate the counterfactual outcomes of previous section. Let $[h_{t_l,z_c}, h_{t_l,z_8}, h_{t_l,z_e}]'$ be the vector of propensity scores estimates for $[P_{t_l}(z_c), P_{t_l}(z_8), P_{t_l}(z_e)]'$ and let $\boldsymbol{\lambda}(h) = [\lambda_1(h), \lambda_2(h), \lambda_3(h)]$ be any selection of linearly independent functions. The counterfactual $E(Y(t_l)|\boldsymbol{S} \in \{\boldsymbol{s}_4, \boldsymbol{s}_6\})$ can be written as $E(Y(t_l)|\boldsymbol{S} \in \Sigma_{t_l}(z_e) \setminus \Sigma_{t_l}(z_8))$ and, according to

[32] As mention in **T-4**, $\Sigma_t(z) \oplus \Sigma_t(z')$ stands for the symmetric difference between sets $\Sigma_t(z)$ and $\Sigma_t(z')$. If $P_t(z) > P_t(z')$ then $\Sigma_t(z) \oplus \Sigma_t(z')$ yields $\Sigma_t(z) \setminus \Sigma_t(z')$. If $P_t(z) < P_t(z')$ then $\Sigma_t(z) \oplus \Sigma_t(z') = \Sigma_t(z') \setminus \Sigma_t(z)$.

T-6,

$$E(Y(t_l)|\boldsymbol{S} \in \{\boldsymbol{s}_4, \boldsymbol{s}_6\}) \text{ is estimated by } \frac{\Big(\boldsymbol{\lambda}(h_{t_l,z_e}) - \boldsymbol{\lambda}(h_{t_l,z_8})\Big)'\widehat{\boldsymbol{\beta}}_{t_l}}{\Big(\boldsymbol{\lambda}(h_{t_l,z_e}) - \boldsymbol{\lambda}(h_{t_l,z_8})\Big)'\widehat{\boldsymbol{\theta}}_{t_l}}, \tag{84}$$

where $h_{t_l,z_8} = P(T = t_l|Z = z_8) = P(\boldsymbol{S} \in \{\boldsymbol{s}_3, \boldsymbol{s}_5\}) < h_{t_l,z_e} = P(T = t_l|Z = z_e) = P(\boldsymbol{S} \in \{\boldsymbol{s}_3, \boldsymbol{s}_5, \boldsymbol{s}_4, \boldsymbol{s}_6\})$.

Equation (76) states that $P_{t_l}(z_8) = P(\boldsymbol{S} \in \{\boldsymbol{s}_3, \boldsymbol{s}_5\})$, $P_{t_l}(z_e) = P(\boldsymbol{S} \in \{\boldsymbol{s}_3, \boldsymbol{s}_5, \boldsymbol{s}_4, \boldsymbol{s}_6\})$ and the identification of $E(Y(t_l)|\boldsymbol{S} = \boldsymbol{s}_4)$ and $E(Y(t_l)|\boldsymbol{S} = \boldsymbol{s}_6)$ hinges on using propensity score $P_{t_l}(z^*) = P(\boldsymbol{S} \in \{\boldsymbol{s}_3, \boldsymbol{s}_5, \boldsymbol{s}_4\})$ to evaluate equations (77)–(78). Let $h^*_{t_l}$ be the estimate of response-type probability $P(\boldsymbol{S} \in \{\boldsymbol{s}_3, \boldsymbol{s}_5, \boldsymbol{s}_4\})$ obtained from $\widehat{\boldsymbol{P}}_S$ of **L-7**. Thus we can use interpolation to evaluate equations (77)–(78) as:

$$E(Y(t_l)|\boldsymbol{S} = \boldsymbol{s}_4) \text{ is estimated by } \frac{\Big(\boldsymbol{\lambda}(h^*_{t_l}) - \boldsymbol{\lambda}(h_{t_l,z_8})\Big)'\widehat{\boldsymbol{\beta}}_{t_l}}{\Big(\boldsymbol{\lambda}(h^*_{t_l}) - \boldsymbol{\lambda}(h_{t_l,z_8})\Big)'\widehat{\boldsymbol{\theta}}_{t_l}}, \tag{85}$$

where $h_{t_l,z_8} = P(T = t_l|Z = z_8) = P(\boldsymbol{S} \in \{\boldsymbol{s}_3, \boldsymbol{s}_5\}) < h^*_{t_l} = P(\boldsymbol{S} \in \{\boldsymbol{s}_3, \boldsymbol{s}_5, \boldsymbol{s}_4\})$

$$E(Y(t_l)|\boldsymbol{S} = \boldsymbol{s}_6) \text{ is estimated by } \frac{\Big(\boldsymbol{\lambda}(h_{t_l,z_e}) - \boldsymbol{\lambda}(h^*_{t_l})\Big)'\widehat{\boldsymbol{\beta}}_{t_l}}{\Big(\boldsymbol{\lambda}(h_{t_l,z_e}) - \boldsymbol{\lambda}(h^*_{t_l})\Big)'\widehat{\boldsymbol{\theta}}_{t_l}}, \tag{86}$$

where $h^*_{t_l} = P(\boldsymbol{S} \in \{\boldsymbol{s}_3, \boldsymbol{s}_5, \boldsymbol{s}_4\}) < h_{t_l,z_e} = P(T = t_l|Z = z_e) = P(\boldsymbol{S} \in \{\boldsymbol{s}_3, \boldsymbol{s}_5, \boldsymbol{s}_4, \boldsymbol{s}_6\})$.

The estimation procedure that evaluates the counterfactual outcomes (77)–(78) for t_l in also applies to t_h and t_m whose identification equations are displayed in Table 10 and Table 11 respectively. Appendix P.3 provides a detailed description of estimation procedures that explore the variation of pre-program variables and accounts for weighting schemes of the observed data.

10 A Summary of Theoretical Contributions

Figure 4 organizes the theoretical contributions of this paper in a comprehensible diagram. Section 4 uses an economic model to examine the counterfactual choices that MTO families can take. The design of the intervention determines the incentive matrix $\boldsymbol{L}$, that is combined with economic assumptions to generate choice restrictions. Those, in turn, produce the response matrix $\boldsymbol{R}$. A key property of the MTO incentive matrix is that it presents monotonic incentives *4.1*, in which a change in the instrument induce incentives towards the same direction for all choices.

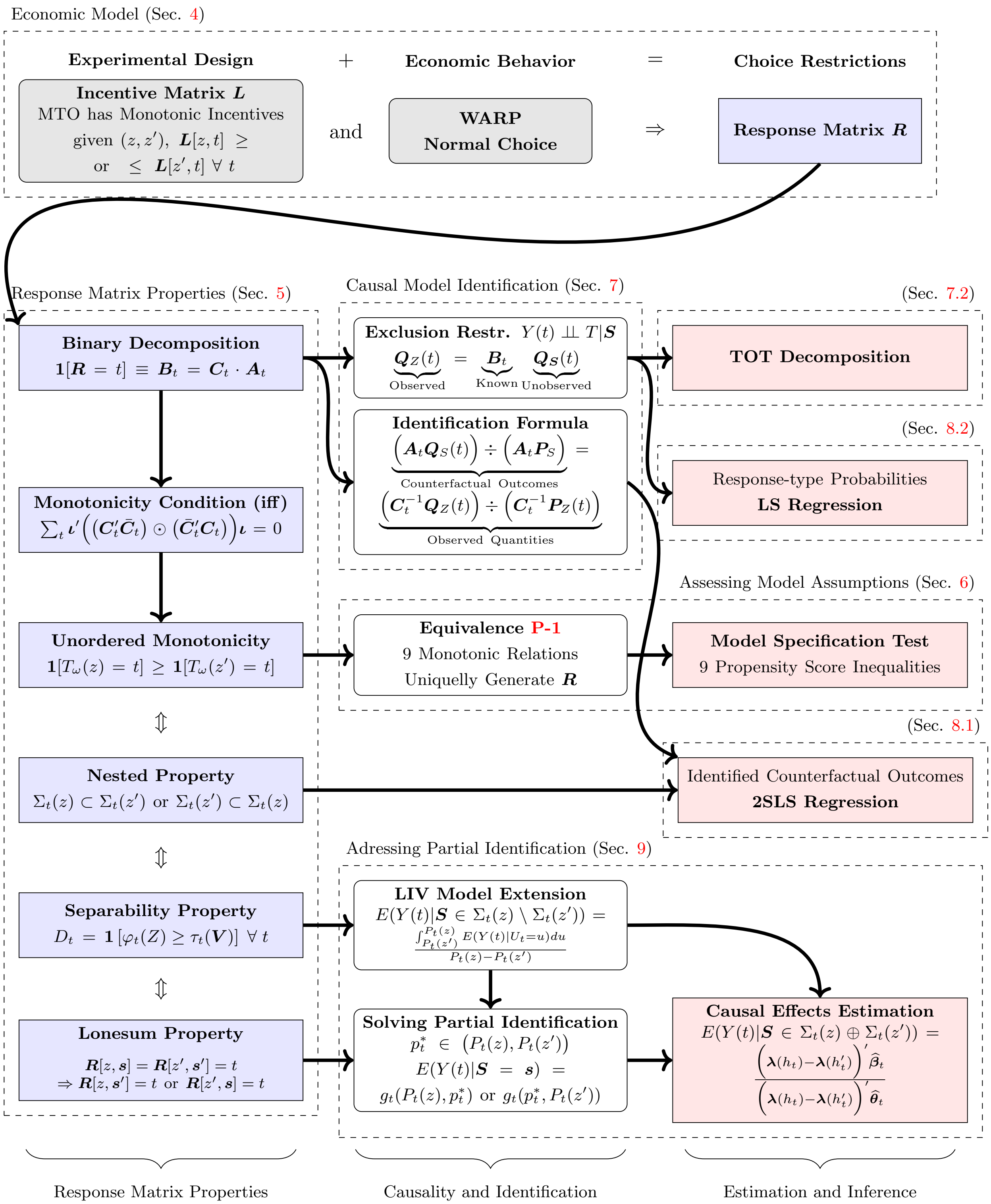
All identification and estimation results stem from the properties of the response matrix. Those are revealed in Section 5 and are termed as: Binary Decomposition (9), Monotonicity (**P-1**), Nested Choices (**P-2**), Lonesum (**P-3**), and Separability (**P-4**). These properties are shared by social experiments characterized by monotonic incentives.

Section 7 transcribes the economic model into a causal model. The response matrix $\boldsymbol{R}$ is expressed in terms of binary matrices $\boldsymbol{B}_t$ (Section 5.1) which are used to equate observed quantities and unobserved causal parameters (equations (35)–(36) in Section 7.1). These matrices are also used to examine the TOT parameter (Section 7.2) and to estimate response-type probabilities (**L-7** in Section 8.2). Binary matrices $\boldsymbol{B}_t$ are decomposed into matrices $\boldsymbol{C}_t$ and $\boldsymbol{A}_t$. These matrices are used to characterize a necessary and sufficient condition for unordered monotonicity to hold (**T-1**) and to yield a general identification formula (**T-2** in Section 7.1).

Unordered Monotonicity implies an equivalence between nine monotonicity inequalities and the seven choice restrictions generated by the revealed preference analysis (Property **P-1**). This equivalence is used to assess assumptions on choice behavior (Section 6). Identification formula **T-2** and the nested property **P-2** are used to show that each identified counterfactual outcome can be estimated via 2SLS (**L-6** and **T-4** in Section 8.1).

Some counterfactual outcomes are only partially identified. This problem is addressed in Section 9. The separability property **P-4** is used to extend the LIV model to the case of multiple choices (**T-5** in Section 9.1) while the lonesum Property **P-3** enables to point-identify counterfactual outcome means via interpolation (**T-6**). Causal effects can be evaluated by 2SLS under a suitable transformation of observed data (Appendix P).

Figure 4: **Summary of Identification and Estimation Results**



11 Empirical Results

This section applies the method described in previous sections to the MTO data. Section 8 shows that all causal parameters can be evaluated by standard econometric tools when proper data transformations are applied. Figure 5 displays the estimates of response-type probabilities which are evaluated by a linear probability model. A similar model is used to estimate the means of pre-program variables conditioned on response-types of Table 12. Figure 6 presents descriptive statistics of three main labor market outcomes: income, employment and the likelihood of breaking out of poverty. Figures 7–8 display the counterfactual outcome estimates for all response-types and for each outcome. Figure 10 presents the estimation of neighborhood causal effects for the outcomes. Counterfactual outcomes and treatment effects are estimated by 2SLS regressions discussed in Section 8.1 and 9.

All estimations are conditioned on site and on baseline pre-program variables regarding family characteristics, mobility, neighborhood safety and neighborhood satisfaction.[33] All estimations account for the adult survey weights of the MTO Interim Impacts Evaluation (2003). Appendix O provides detailed information on the estimation of response-type probabilities and the expected value of pre-program variables conditioned on response-types. Appendix P describe the procedures regarding the estimation of causal effects.

Response-type Probabilities

Figure 5 presents the response-type estimates.[34] Almost half of the families are always-takes ($\hat{P}(\boldsymbol{S} \in \{\boldsymbol{s}_1, \boldsymbol{s}_2, \boldsymbol{s}_3\}) = 0.46$) that do not change their neighborhood choice regardless of the voucher assignment. In particular, one in each three families remain on high-poverty neighborhood ($\hat{P}(\boldsymbol{S} = \boldsymbol{s}_1) = 0.35$) regardless of the voucher assignment. These consists of families for whom MTO incentives are not sufficient to incite neighborhood relocation. The second most frequent family-type is $\hat{P}(\boldsymbol{S} = \boldsymbol{s}_4) = 0.31$ which account to about a third of families. Response-type $\boldsymbol{s}_4$ comprises the most responsive families, those who choose high, medium and low poverty neighborhoods if the family is assigned to control, Section 8 and Experimental vouchers respectively. The remaining compliers $\boldsymbol{s}_5, \boldsymbol{s}_6, \boldsymbol{s}_7$ account to 23.5% of families. Some variation exists across cities. In particular, about half of families in Los Angeles are classified as $\boldsymbol{s}_4$.

Baseline Variables by Response-types

MTO vouchers have no impact on the neighborhood choice of always-taker families ($\boldsymbol{s}_1, \boldsymbol{s}_2, \boldsymbol{s}_3$), which account for almost half of all eligible families. If family response-types were observed, a

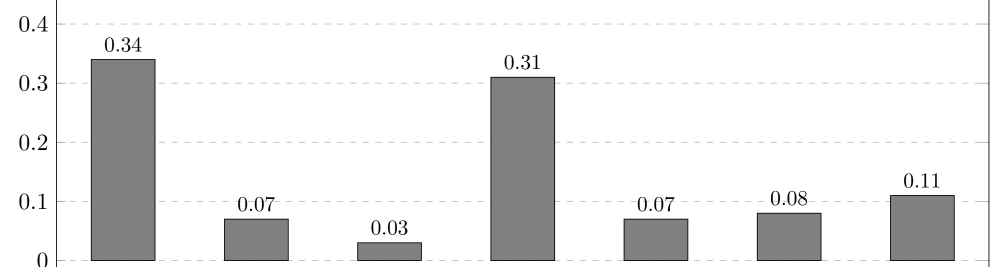
[33] Family characteristics: if resident ever married, if has no teenagers and if has a disabled family member. Mobility: if participant had applied for a Section 8, if has moved at least 3 times within 5 years previous to the intervention; Neighborhood safety: if being beaten/assaulted in the past 6 months (prior to intervention), if has moved in the past due to gangs, and if feels unsafe at night. Neighborhood satisfaction: reported no friends, has watched for neighbor's children, if has no family in the neighborhood, if chats with neighbor, and neighborhood dissatisfaction index.

[34] Appendix O.3 describes the estimation of response-type probabilities in detail.

Figure 5: Response-type Probabilities

Response-types	s_1	s_2	s_3	s_4	s_5	s_6	s_7
$T(z_c)$	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

All Data	$P(\boldsymbol{S} = \boldsymbol{s}_1)$	$P(\boldsymbol{S} = \boldsymbol{s}_2)$	$P(\boldsymbol{S} = \boldsymbol{s}_3)$	$P(\boldsymbol{S} = \boldsymbol{s}_4)$	$P(\boldsymbol{S} = \boldsymbol{s}_5)$	$P(\boldsymbol{S} = \boldsymbol{s}_6)$	$P(\boldsymbol{S} = \boldsymbol{s}_7)$



	s_1	s_2	s_3	s_4	s_5	s_6	s_7
s.d.	(0.01)	(0.01)	(0.01)	(0.02)	(0.01)	(0.01)	(0.01)
Baltimore	0.21	0.09	0.04	0.27	0.12	0.13	0.16
(s.d)	(0.03)	(0.02)	(0.03)	(0.04)	(0.03)	(0.03)	(0.03)
Boston	0.43	0.07	0.04	0.29	0.05	0.11	0.01
(s.d)	(0.02)	(0.02)	(0.02)	(0.03)	(0.03)	(0.03)	(0.03)
Chicago	0.27	0.09	0.02	0.20	0.05	0.07	0.28
(s.d)	(0.02)	(0.02)	(0.02)	(0.03)	(0.03)	(0.03)	(0.03)
Los Angeles	0.13	0.04	0.02	0.45	0.09	0.10	0.17
(s.d)	(0.02)	(0.02)	(0.02)	(0.03)	(0.03)	(0.03)	(0.03)
New York	0.49	0.06	0.04	0.34	0.06	0.01	0.00
(s.d)	(0.02)	(0.02)	(0.02)	(0.03)	(0.03)	(0.02)	(0.02)
	s_1	s_2	s_3	s_4	s_5	s_6	s_7

The first panel of the figure lists the counterfactual choices of each response-types. The second panel displays the estimated probabilities of the response-types according to Section 8.2. The last panes provides the response-type probabilities by city. Sample sizes are 633, 957, 889, 676, 1072 respectively. All estimations account for the person-level weight for adult survey of the interim analyses as described in the MTO Interim Impacts Evaluation manual, 2003, Appendix B. Estimates are conditional on site and baseline variables regarding family characteristics, mobility, neighborhood safety and satisfaction. [1] Family Characteristics: if resident ever married, if has no teenagers and if has a disable family member; [2] Mobility: applied for a Section 8, if has moved at least 3 times within 5 years; [3] Neighborhood safety: if being beaten/assaulted in the past 6 months (prior to intervention), if has moved in the past due to gangs, and if feels unsafe at night. [4] Neighborhood satisfaction: reported no friends, has watched for neighbor's children, if has no family in the neighborhood, if chats with neighbor, and neighborhood dissatisfaction index. See Appendix O.3 for detailed description of the estimation procedure.

policy maker could improve the efficiency of the MTO intervention by targeting only the families who respond to the MTO incentives. Those are the compliers $\boldsymbol{s}_4, \boldsymbol{s}_5, \boldsymbol{s}_6$ and $\boldsymbol{s}_7$.

Unfortunately response-types are not observed. Nevertheless the distribution of pre-program variables conditioned on response-types is identified (**T-3**). Thus it is possible to estimate the likelihood that a family is of a particular response-type type conditioned on its baseline characteristics at the onset of the intervention (Section 7.1). This information in useful in designing an eligibility criteria toward more efficient interventions.[35]

It is also possible to investigate how family baseline characteristics vary by response-type. Table 12 presents the means of selected pre-program variables conditioned on response-types.[36] The first variable measures education, the second assesses if the family has a car, the third indicates if the family had applied to housing welfare previously and the last question inquires if the family believe that it is able to move from housing projects regardless of public assistance. These variables are proxies for family poverty. A clear pattern arises among always-takes: $\boldsymbol{s}_1$-families are most disadvantaged, followed by $\boldsymbol{s}_2$-families and then by $\boldsymbol{s}_3$-families. These are the families that choose high, medium and low-poverty neighborhood respectively. Among compliers, the most disadvantaged families are $\boldsymbol{s}_5$ which choose to relocate to a low-poverty neighborhood whenever subsidy is available.

The next four questions assess family behavior and composition. Families that always remain in high-poverty neighborhoods ($\boldsymbol{s}_1$) are the ones most likely to have a disable house member (27% above baseline), most likely to have lived in the neighborhood for an extended period (14% above baseline), and most likely to have teenagers (16% above baseline). In summary, those are families who have less mobility and more ties to the community they live in. In constrast, families that are most responsive to the voucher incentives ($\boldsymbol{s}_4$) are the ones who are less likely to have teenagers as family members (20% below baseline) and families that always move to low-poverty neighborhood ($\boldsymbol{s}_3$) are the ones less likely to chat with the neighbors during their stay at high-poverty neighborhoods (43% below baseline).

The last four questions assess the reasons for moving from high poverty neighborhood. Always takes $\boldsymbol{s}_3$ are most likely to move to low-poverty neighborhoods seeking better schools (58% above baseline). The main reason for $\boldsymbol{s}_6$-families is gang activity (16% above baseline). These families are also most likely to live in unsafe areas (17% above baseline) and are often victims of crime (22% above baseline).

Outcome Means by Voucher and Neighborhood Choice

Figure 6 displays the conditional means of labor market outcomes surveyed at the interim evaluation of MTO. The first outcome is the income of the head of the family in thousand dollars per year. The second outcome is called *breaking through poverty* and indicates if the total household

[35]In particular, interventions could target compliers $\boldsymbol{s}_4, \boldsymbol{s}_5, \boldsymbol{s}_6$ and $\boldsymbol{s}_7$ or that increase the incentives for families that are likely to be of type $\boldsymbol{s}_1$.

[36]Appendix O.4 describes the estimation procedure for pre-program variables mean conditional on response-types.

Table 12: Pre-program Variables Means by Response-types

	Baseline	Always-takes			Compliers			
		s_1	s_2	s_3	s_4	s_5	s_6	s_7
Completed high school	0.38	0.35	0.41	0.71	0.36	0.26	0.42	0.44
(s.d.)	(0.01)	(0.02)	(0.10)	(0.24)	(0.04)	(0.16)	(0.12)	(0.09)
% diff. from baseline	–	−7.5%	6.5%	87.1%	−4.9%	−27.8%	11.2%	15.2%
Car Owner	0.17	0.14	0.23	0.53	0.20	0.10	0.05	0.13
(s.d.)	(0.01)	(0.02)	(0.07)	(0.17)	(0.03)	(0.11)	(0.08)	(0.06)
% diff. from baseline	–	−18.0%	38.1%	216.3%	21.6%	−40.8%	−68.4%	−22.6%
Applied for Section 8?	0.42	0.43	0.35	0.18	0.41	0.73	0.47	0.31
(s.d.)	(0.01)	(0.02)	(0.10)	(0.25)	(0.04)	(0.17)	(0.12)	(0.09)
% diff. from baseline	–	2.4%	−14.7%	−57.1%	−2.6%	75.9%	12.1%	−26.1%
Prospective mover	0.46	0.36	0.55	0.78	0.53	0.31	0.50	0.51
(s.d.)	(0.01)	(0.03)	(0.11)	(0.26)	(0.04)	(0.17)	(0.13)	(0.10)
% diff. from baseline	–	−22.6%	20.1%	70.0%	14.4%	−33.4%	7.8%	9.7%
Disable Household Member	0.16	0.21	0.16	0.15	0.16	0.19	0.12	0.08
(s.d.)	(0.01)	(0.02)	(0.07)	(0.17)	(0.03)	(0.11)	(0.08)	(0.06)
% diff. from baseline	–	26.6%	−1.8%	−7.4%	−4.8%	13.5%	−28.2%	−50.8%
Resident for 5 + yrs.	0.61	0.69	0.65	0.54	0.61	0.56	0.50	0.47
(s.d.)	(0.01)	(0.03)	(0.11)	(0.28)	(0.05)	(0.19)	(0.14)	(0.10)
% diff. from baseline	–	13.7%	7.1%	−11.7%	−0.6%	−8.3%	−18.1%	−23.4%
No teens (ages 13-17)	0.61	0.51	0.63	0.58	0.73	0.58	0.59	0.61
(s.d.)	(0.01)	(0.03)	(0.11)	(0.28)	(0.05)	(0.19)	(0.14)	(0.10)
% diff. from baseline	–	−15.9%	2.3%	−4.9%	19.7%	−5.6%	−3.4%	0.0%
Chat with neighbor	0.51	0.51	0.48	0.29	0.45	0.65	0.59	0.61
(s.d.)	(0.01)	(0.03)	(0.11)	(0.27)	(0.04)	(0.18)	(0.13)	(0.10)
% diff. from baseline	–	−0.1%	−5.3%	−42.5%	−11.3%	27.2%	15.9%	20.5%
Moved to seek schools	0.48	0.48	0.50	0.76	0.57	0.16	0.38	0.41
(s.d.)	(0.01)	(0.03)	(0.11)	(0.26)	(0.04)	(0.17)	(0.13)	(0.10)
% diff. from baseline	–	−0.7%	2.9%	57.5%	18.5%	−66.4%	−20.6%	−14.8%
Moved due to gangs	0.77	0.73	0.71	0.63	0.78	0.80	0.89	0.81
(s.d.)	(0.01)	(0.03)	(0.12)	(0.29)	(0.05)	(0.20)	(0.15)	(0.11)
% diff. from baseline	–	−4.9%	−7.9%	−18.1%	1.9%	4.3%	15.8%	5.9%
Unsafe at night	0.49	0.43	0.47	0.31	0.53	0.57	0.57	0.52
(s.d.)	(0.01)	(0.03)	(0.11)	(0.26)	(0.04)	(0.18)	(0.13)	(0.10)
% diff. from baseline	–	−12.7%	−3.1%	−36.4%	8.3%	17.2%	17.2%	5.9%
Victim last 6 months	0.42	0.41	0.30	0.46	0.43	0.51	0.44	0.36
(s.d.)	(0.01)	(0.02)	(0.10)	(0.25)	(0.04)	(0.17)	(0.12)	(0.09)
% diff. from baseline	–	−0.4%	−26.7%	11.4%	4.2%	22.2%	4.8%	−13.9%

The first column lists pre-program variables surveyed at the intervention onset. The second column (baseline) presents the variable mean across all response-types. The remaining seven columns present the variable mean conditioned on response-types. The first line provides the estimated variable means, the second line gives the standard deviation of the estimate and the third line gives the difference between the variable mean conditioned on the response-types and the baseline mean in percentage points. 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). The sample size is 4227. See Appendix O.4 for detailed description of the estimation procedure.

income is above poverty line. The third outcome indicates if the head of the family is employed and not on welfare. These outcome means must be interpreted as statistical description of the data as they do not account for the selection bias induced by noncompliance.

Graph A in Figure 6 shows that the income for control families that decide for high, median and low-poverty neighborhoods are \$11.36, \$13.01 and \$16.17 respectively. If there were no selection bias, the causal effect of low versus high-poverty neighborhoods on income would be $\$16.17 - \$11.36 = \$4.81$ thousand dollars per year.

The income difference decreases conditioned on Section 8 and experimental vouchers. The income gap between low and high poverty neighborhoods for Section 8 is \$2.51 and for experimental voucher is \$0.67 thousand dollars per year respectively. This pattern suggests strong selection bias. The lack of subsidy of the control group prevents the lower income families to move. Section 8 allows for a share of these lower income families to move to either high or medium poverty neighborhoods. As a consequence, we observe a decrease in the average income when comparing control with section 8 families for either medium-poverty neighborhoods ($-\$1.02$) and high-poverty neighborhoods ($-\$2.34$). The experimental voucher incentivizes lower income families to move to low-poverty neighborhoods. When comparing control with experimental families, we observe that the income for low-poverty neighborhoods decreases from \$16.17 to \$12.56. Similar pattern is observed for the remaining outcomes.

Counterfactual Outcomes Conditioned on Response-types

Figures 7–9 display the estimated counterfactual outcome means for income, breaking though poverty and employment respectively.[37] Each counterfactual estimate is evaluated by a 2SLS regression discussed in Section 8.1 and 9. As mentioned, estimates account for the adult survey weights of the MTO Interin Evaluation and are conditioned on site and baseline control variables regarding family characteristics, mobility, neighborhood safety and satisfaction.

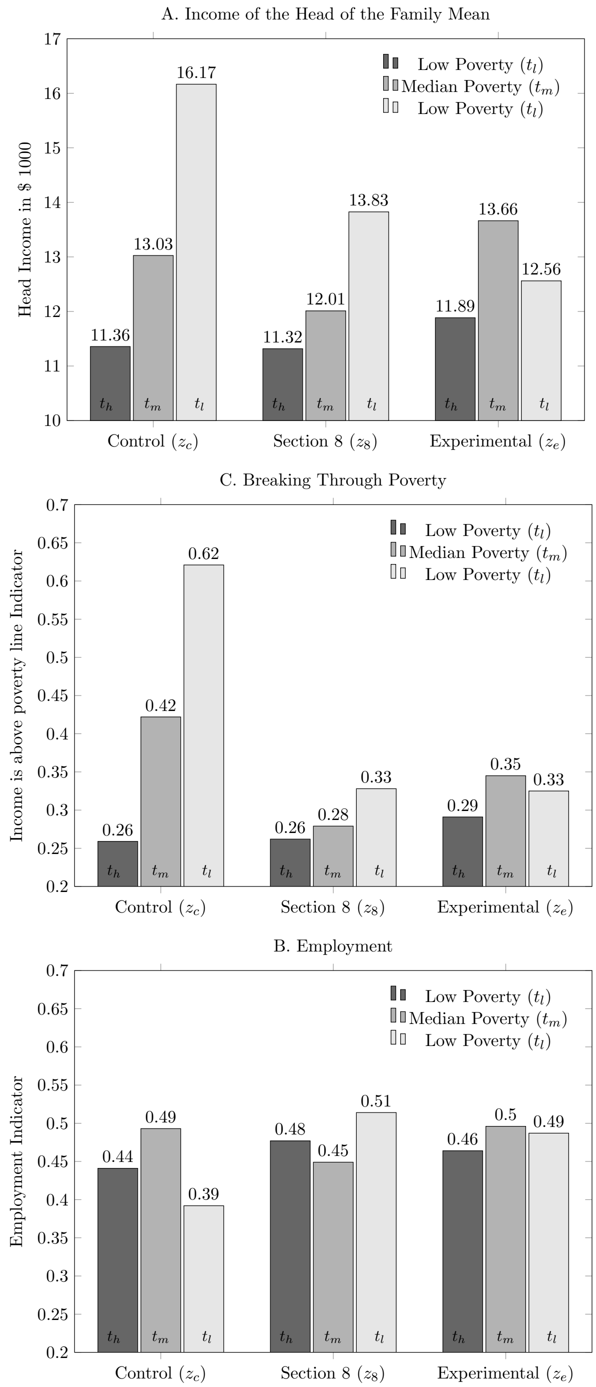
Graph A in Figure 7 shows the counterfactual income estimates for the always-takers. We observe a sharp increase in income of the always-taker families as the neighborhood choices range across high, median and low-poverty. These counterfactual outcomes are associated with families with distinct unobserved variables. Therefore the income difference across neighborhood types cannot be understood as neighborhood causal effects. As expected, the poorer families among always-takes are the ones who choose high-poverty neighborhoods ($\boldsymbol{s}_1$). These are also the most disadvantaged families in terms of baseline characteristics and they comprise a third of the MTO sample.

Graph B investigates the $\boldsymbol{s}_4$-families who are the focus of this empirical analysis. Families of response-type $\boldsymbol{s}_4$ also comprise a third of the sample. These families are the most responsive to MTO incentives as they chose among all neighborhood poverty levels as the voucher assignment varies. Graph B shows a steep increase in income as families move to better neighborhoods. The same pattern is observe in Graph B of Figures 8 and 9.

[37] Appendices O.5 and P describe the estimation procedures in detail.

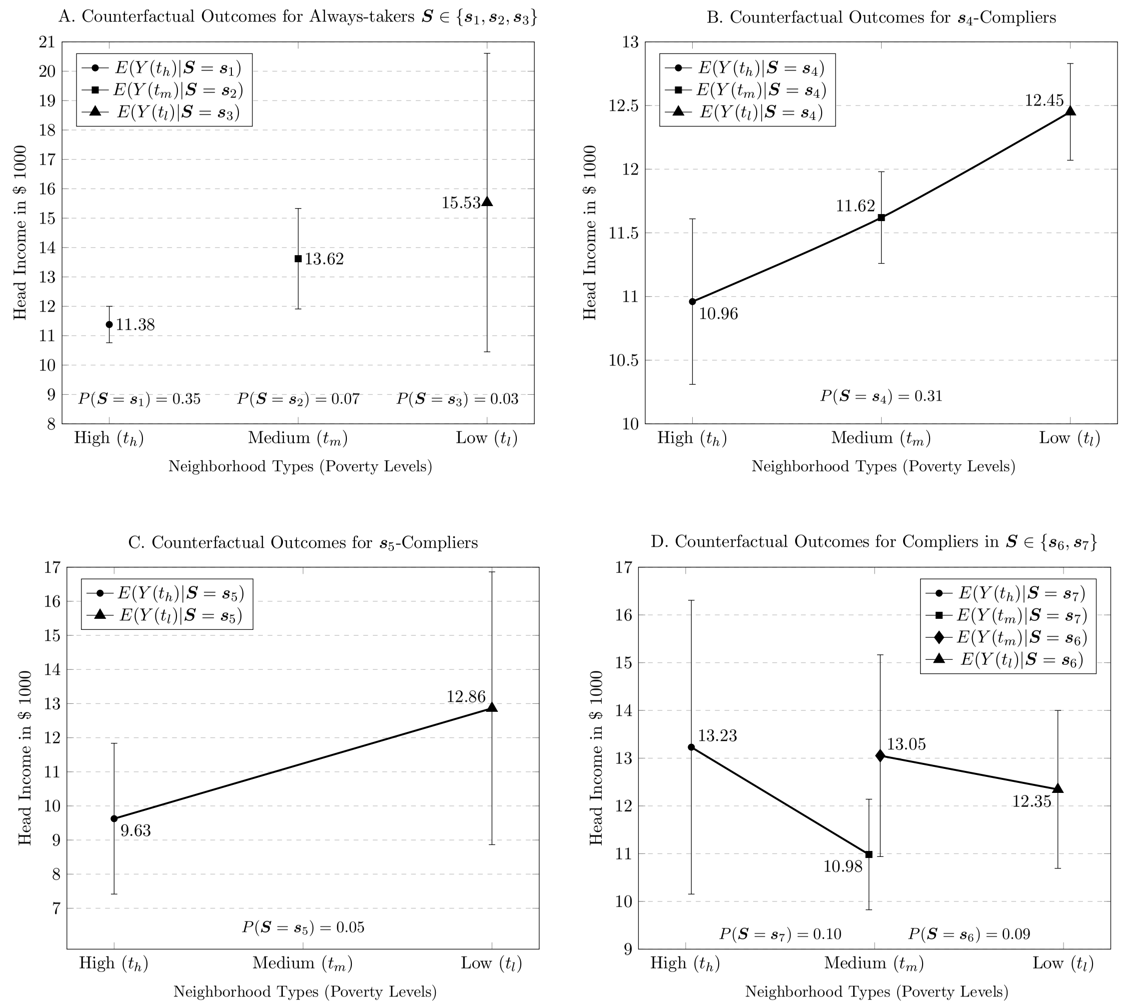
Graph C in Figure 7 examines $\boldsymbol{s}_5$-families. Estimates show a large income difference between low and high-poverty neighborhoods. However these families account for only 5% of the total sample and these values lack statistical precision. Graph D examines families of type $\boldsymbol{s}_6$ and $\boldsymbol{s}_7$. Families of type $\boldsymbol{s}_7$ only chose between high and medium-poverty neighborhoods while $\boldsymbol{s}_6$ families chose between medium and low-poverty neighborhoods. Income estimates for these families have a large standard deviation as these families comprise only about 10% of the sample. Similar patterns are observed in Graphs C,D of Figures 8–9.

Figure 6: Descriptive Statistics of Outcome Means by Neighborhood Choice and Voucher Assignment



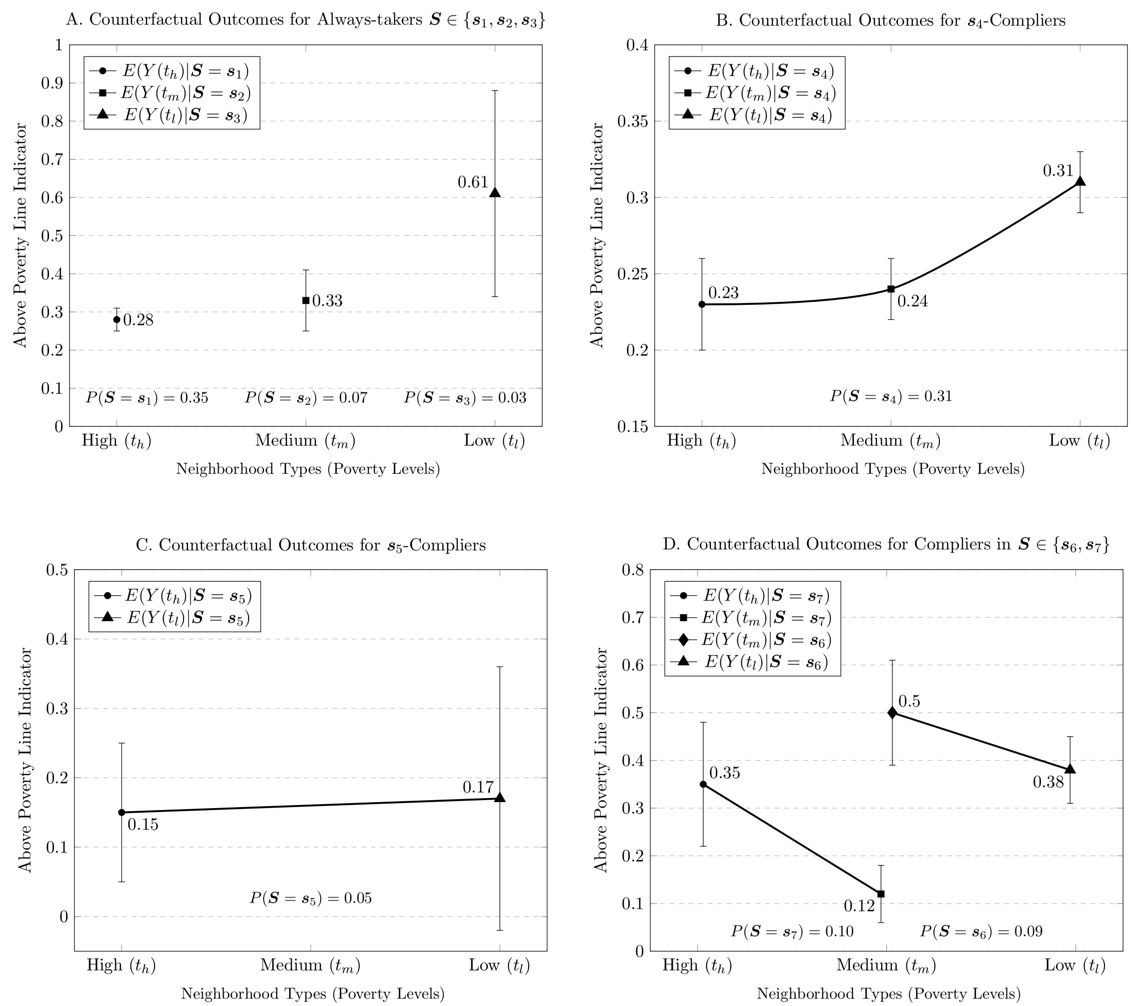
This figure presents display the statistical description of outcome means conditioned on voucher assignments and neighborhood choices. Graph A displays Income of the Head of the Family (in $1000). Graph B displays the Breaking Through Poverty, namely the indicator whether household income is above poverty line. Graph C displays Employment (household head employed and not on welfare); All estimates are conditioned on site and on baseline control variables regarding family characteristics, mobility, neighborhood safety and satisfaction. [1] Family Characteristics: if resident ever married, if has no teenagers and if has a disable family member; [2] Mobility: applied for a Section 8, if has moved at least 3 times within 5 years; [3] Neighborhood safety: if being beaten/assaulted in the past 6 months (prior to intervention), if has moved in the past due to gangs, and if feels unsafe at night. [4] Neighborhood satisfaction: reported no friends, has watched for neighbor's children, if has no family in the neighborhood, if chats with neighbor, and neighborhood dissatisfaction index. All estimations account for the person-level weight for adult survey of the interim analyses as described in the MTO Interim Impacts Evaluation manual, 2003, Appendix B.

Figure 7: Counterfactual Outcome Estimates for Income of the Head of the Family



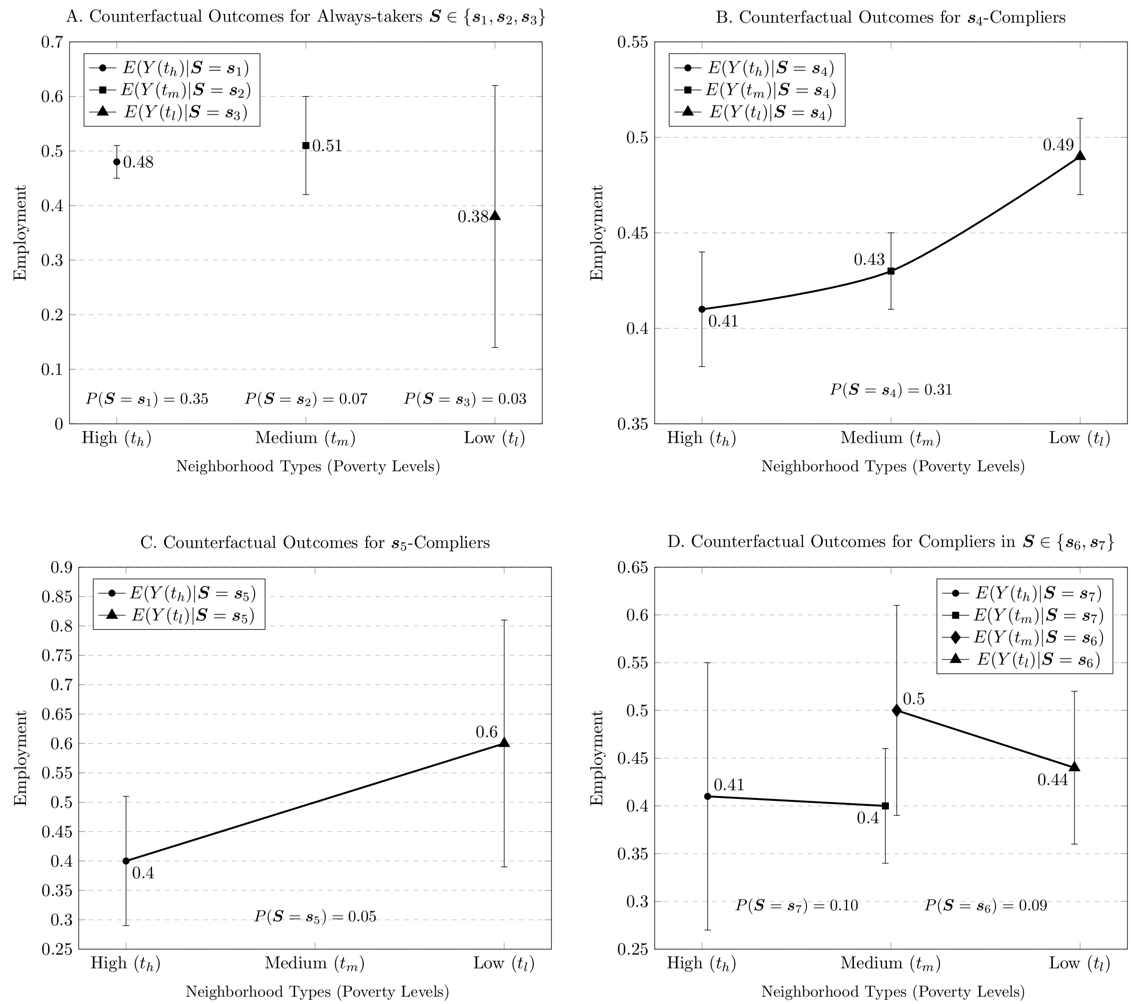
This figure presents four graphs that display the estimates of the counterfactual outcomes associated with the MTO response matrix in **L-3**. The outcome is the income of the head of the family. Graph A displays the counterfactual outcome estimates for always-takes (s_1, s_2, s_3); Graph B examines response-type s_4; Graph C investigates type s_5 and Graph D presents results for response-types s_6 and s_7. Each counterfactual estimate is evaluated by the 2SLS regressions discussed in Section 8.1 and 9. All 2SLS estimates are conditioned on site and on baseline control variables regarding family characteristics, mobility, neighborhood safety and satisfaction. [1] Family Characteristics: if resident ever married, if has no teenagers and if has a disable family member; [2] Mobility: applied for a Section 8, if has moved at least 3 times within 5 years; [3] Neighborhood safety: if being beaten/assaulted in the past 6 months (prior to intervention), if has moved in the past due to gangs, and if feels unsafe at night. [4] Neighborhood satisfaction: reported no friends, has watched for neighbor's children, if has no family in the neighborhood, if chats with neighbor, and neighborhood dissatisfaction index. All estimations account for the person-level weight for adult survey of the interim analyses as described in the MTO Interim Impacts Evaluation manual, 2003, Appendix B. See Appendices O and P for the description of the estimation procedures.

Figure 8: Counterfactual Outcome Estimates on Breaking Through Poverty



This figure presents four graphs that display the estimates of the counterfactual outcomes associated with the MTO response matrix in **L-3**. The outcome is the indicator whether the household income is above the poverty line. Graph A displays the counterfactual outcome estimates for always-takes $(\boldsymbol{s}_1, \boldsymbol{s}_2, \boldsymbol{s}_3)$; Graph B examines response-type $\boldsymbol{s}_4$; Graph C investigates type $\boldsymbol{s}_5$ and Graph D presents results for response-types $\boldsymbol{s}_6$ and $\boldsymbol{s}_7$. Each counterfactual estimate is evaluated by the 2SLS regressions discussed in Section 8.1 and 9. All 2SLS estimates are conditioned on site and on baseline control variables regarding family characteristics, mobility, neighborhood safety and satisfaction. [1] Family Characteristics: if resident ever married, if has no teenagers and if has a disable family member; [2] Mobility: applied for a Section 8, if has moved at least 3 times within 5 years; [3] Neighborhood safety: if being beaten/assaulted in the past 6 months (prior to intervention), if has moved in the past due to gangs, and if feels unsafe at night. [4] Neighborhood satisfaction: reported no friends, has watched for neighbor's children, if has no family in the neighborhood, if chats with neighbor, and neighborhood dissatisfaction index. All estimations account for the person-level weight for adult survey of the interim analyses as described in the MTO Interim Impacts Evaluation manual, 2003, Appendix B. See Appendices O and P for the description of the estimation procedures.

Figure 9: Counterfactual Outcome Estimates for Employment

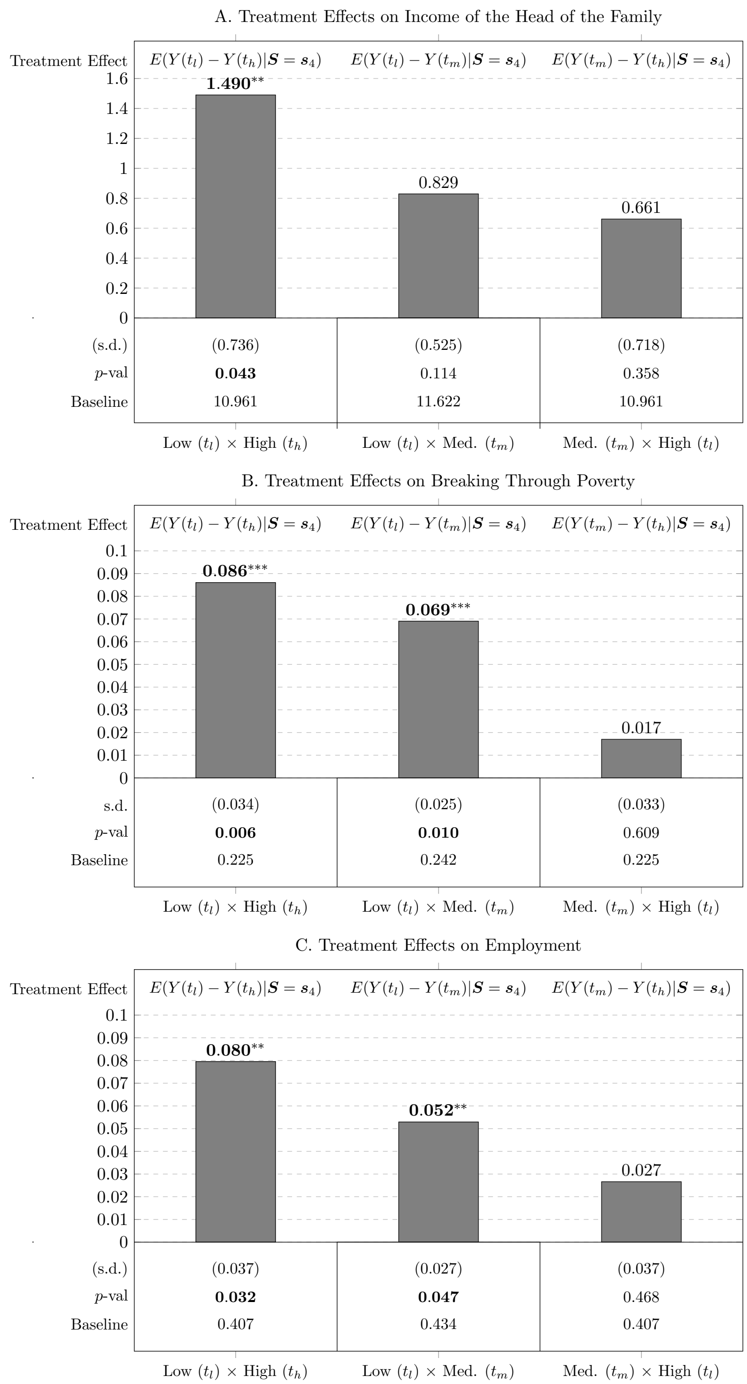


This figure presents four graphs that display the estimates of the counterfactual outcomes associated with the MTO response matrix in **L-3**. The outcome is the indicator whether the adult respondent is working and not receiving welfare. Graph A displays the counterfactual outcome estimates for always-takes $(\boldsymbol{s}_1, \boldsymbol{s}_2, \boldsymbol{s}_3)$; Graph B examines response-type $\boldsymbol{s}_4$; Graph C investigates type $\boldsymbol{s}_5$ and Graph D presents results for response-types $\boldsymbol{s}_6$ and $\boldsymbol{s}_7$. Each counterfactual estimate is evaluated by the 2SLS regressions discussed in Section 8.1 and 9. All 2SLS estimates are conditioned on site and on baseline control variables regarding family characteristics, mobility, neighborhood safety and satisfaction. [1] Family Characteristics: if resident ever married, if has no teenagers and if has a disable family member; [2] Mobility: applied for a Section 8, if has moved at least 3 times within 5 years; [3] Neighborhood safety: if being beaten/assaulted in the past 6 months (prior to intervention), if has moved in the past due to gangs, and if feels unsafe at night. [4] Neighborhood satisfaction: reported no friends, has watched for neighbor's children, if has no family in the neighborhood, if chats with neighbor, and neighborhood dissatisfaction index. All estimations account for the person-level weight for adult survey of the interim analyses as described in the MTO Interim Impacts Evaluation manual, 2003, Appendix B. See Appendices O and P for the description of the estimation procedures.

Figure 10 displays the treatment effects of compliers $\boldsymbol{s}_4$ for income (Graph A), breaking though poverty (Graph B) and employment (Graph C). The first column of each graph displays the treatment effects for low versus high poverty neighborhoods, the second column examines low versus median poverty and the last column shows causal effects for medium versus low poverty neighborhoods. Treatment effects for high versus low-poverty neighborhoods are statistically significant for all labor market outcomes. Families who move from high-poverty neighborhoods to low-poverty neighborhoods experience, on average, a 14% increase in income, a 20% increase in employment and an increase of 38% in their likelihood of breaking out of poverty. The treatment effects for low versus median poverty neighborhoods (second columns) are smaller than those for low versus high. This result is coherent with the decrease in the quality gap between neighborhood types. Treatment effects are still significant for employment and breaking trough poverty outcomes. The last columns display the treatment effect for medium versus low poverty neighborhoods. None of these estimates are statistically significant.

The treatment effects corroborate a large literature on social sciences claiming that neighborhood characteristics have substantial influence the economic well-being of its residents (Wilson, 2009). All estimates are positive supporting the claim that better neighborhoods impact labor market outcomes positively. Estimates are larger the bigger the gap in neighborhood quality. The largest treatment effects occur when comparing low versus high poverty neighborhoods, followed by low versus median and then by median versus high.

Figure 10: Causal Effects on Employment, Poverty Reduction and Income for Compliers ($\boldsymbol{S} = \boldsymbol{s}_4$)



This figure displays three graphs showing the causal effects for response-type $\boldsymbol{s}_4$. Graph A examines the income of the head of the family in \$1000 per year. Graph B focus on Breaking Through Poverty, that is an indicator whether household income is above poverty line. Graph C displays the Employment Indicator (household head employed and not on welfare); The first column of each graph displays the causal effect for low versus high poverty neighborhood ($E(Y(t_l) - Y(t_h)|\boldsymbol{S} = \boldsymbol{s}_4)$), the second column for low versus medium poverty neighborhood ($E(Y(t_l) - Y(t_m)|\boldsymbol{S} = \boldsymbol{s}_4)$) and the last column for medium versus high poverty neighborhood ($E(Y(t_m) - Y(t_h)|\boldsymbol{S} = \boldsymbol{s}_4)$). The following information is displayed for each causal effect: (1) the estimated standard deviation (s.d); (2) the p-valued for two-tailed test on the null hypothesis that equates the causal effect to zero (p-val); and (3) the counterfactual outcome mean of the comparison group (Baseline). Each estimate is obtained by performing 2SLS regressions on transformed data as discussed in Section 8.1 and 9. All estimates are conditioned on site and on baseline control variables regarding family characteristics, mobility, neighborhood safety and satisfaction in the same fashion as the estimations in Figures 7–8. Estimations use adult person-level weight as described in the MTO Interim Impacts Evaluation manual, 2003, Appendix B. See Appendix P for detailed description of the estimation procedure.

12 Conclusions

This paper offers a framework that enables researchers to exploit information on the incentives induced by social experiments and classical economic behavior to identify treatment effects. Two concepts play a primary role in the method: the *incentive matrix*, which characterizes the incentives induced by the design of the intervention, and the *response matrix*, which describes the possible counterfactual choices that agents can choose. The incentive matrix is determined by the experimental design. The method uses reveled preference analysis translates incentives into choice restrictions. Those restrictions generate the response matrix, which contains all the necessary information to examine the nonparametric identification of causal parameters.

This framework is used to evaluate neighborhood effects of the Moving to Opportunity, a housing experiment that designed to investigate the economic consequences of relocating poor families living in high poverty neighborhoods to low-poverty communities. The intervention randomly assigned families into three groups: control (z_c), Section 8 (z_8) and experimental (z_e). Experimental families received incentives to move to a low-poverty neighborhood. Section 8 families received incentives to move to either low or medium-poverty neighborhoods and control families received no incentives. MTO noncompliance was substantial. Half of the families that received incentives to relocate to low-poverty neighborhoods did not move while 20% of the that did not receive incentives moved to either medium or low-poverty neighborhoods. Noncompliance still allows for the evaluation the causal effect of being offered a voucher, that is, voucher effects. examples of voucher effects are the intention-to-treat and the treatment-on-the-treated parameters. An influential literature on MTO shows that voucher effects on labor market outcomes are not statistically significant.

This paper differs from previous literature by exploiting the information on the incentives induced by the MTO intervention. This enables to evaluate *neighborhood effects*, that is, the causal effect of residing in different neighborhood types. A useful characteristic of the MTO incentive matrix is that it presents *monotonic incentives*: changes in the instrument affect incentives in the same direction *for all choices*. Otherwise stated, a change in the instrumental values from z to z' that increases incentives for a choice t cannot decrease incentives for another choice t'. Monotonic incentives attributes non-trivial properties to the response matrix. The response matrix properties are then used in the identification of causal parameters, estimation of treatment effects and testing of model assumptions. The method also permits to decompose voucher effects in terms of neighborhood effects.

Although voucher effects are not significant, neighborhood effects are. On average, moving from a high-poverty neighborhoods to low-poverty neighborhoods yields a 14% increase in income, a 20% increase in employment and an increase of 38% on the likelihood of breaking out of poverty for the families that respond to MTO incentives. Moreover, causal effects are greater the greater the gap between neighborhood poverty levels. This result reconciles MTO with a large literature which claims that neighborhood characteristics has a significant influence the lives of its residents (Wilson, 2009).

A major benefit of this framework is that noncompliance is not perceived as an econometric problem, but rather an essential tool for policy evaluation. The estimation of causal parameters can be achieved using well-known econometric methods such as OLS and 2SLS regressions under a suitable transformation of the data. The method can be broadly applied to exploit economic incentives in multiple choice models with heterogeneous agents and categorical instrumental variables.

References

Ahn, H. and J. Powell (1993, July). Semiparametric estimation of censored selection models with a nonparametric selection mechanism. *Journal of Econometrics 58*(1–2), 3–29.

Aliprantis, D. and F. G. Richter (2014). Evidence of neighborhood effects from mto: Lates of neighborhood quality. *Umpublished Manuscript*.

Angrist, J. D. and G. W. Imbens (1995, June). Two-stage least squares estimation of average causal effects in models with variable treatment intensity. *Journal of the American Statistical Association 90*(430), 431–442.

Angrist, J. D., G. W. Imbens, and D. Rubin (1996). Identification of causal effects using instrumental variables. *Journal of the American Statistical Association 91*(434), 444–455.

Barnett, S. (1990). *Oxford applied mathematics and computing science series*. New York, NY: Oxford Applied Mathematics Clarendon Press.

Bloom, H. S. (1984, April). Accounting for no-shows in experimental evaluation designs. *Evaluation Review 8*(2), 225–246.

Blundell, R., M. Browning, and I. Crawford (2003). Nonparametric engel curves and revealed preference. *Econometrica 71*, 205–240.

Blundell, R., M. Browning, and I. Crawford (2008). Best nonparametric bounds on demand responses. *Econometrica 76*, 1227–1262.

Blundell, R., D. Kristensen, and R. Matzkin (2014). Bounding quantile demand functions using revealed preference inequalities. *Journal of Econometrics 179*, 112–127.

Brinch, C. N., M. Mogstad, and M. Wiswall (2017). Beyond late with a discrete instrument. *Journal of Political Economy 125*(4), 985–1039.

Brualdi, R. (1980). Matrices of zeros and ones with fixed row and column sum vectors. *Lin. Algebra Appl. 33*, 159231.

Chetty, R., J. Friedman, N. Hendren, M. Jones, and S. Porter (2017). The opportunity atlas. *Unpublished Working Paper* (23568).

Chetty, R., N. Hendren, and L. Katz (2016). The effects of exposure to better neighborhoods on children: New evidence from the moving to opportunity experiment. *American Economic Review 106*, 855–902.

Chyn, E. (2016). Moved to opportunity: The long-run effect of public housing demolition on labor market outcomes of children. *Umpublished Manuscript*.

Clampet-Lundquist, S. and D. Massey (2008). Neighborhood effects on economic self-sufficiency: A reconsideration of the moving to opportunity experiment. *American Journal of Sociology 114*, 107 – 143.

Conti, G., J. J. Heckman, and R. Pinto (2016). The effects of two influential early childhood interventions on health and healthy behaviours. Economic Journal *126*(596), F28–F65.

Fan, J. and I. Gijbels (1996). *Local Polynomial Modelling and its Applications.* New York: Chapman and Hall.

Feller, F., T. Grindal, M. Luke, and L. Page (2014). Compared to what? variation in the impacts of early childhood education by alternative care-type settings. *Umpublished Manuscript*.

Fisher, R. A. (1935). *The Design of Experiments.* London: Oliver and Boyd.

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

Gennetian, L. A., M. Sciandra, L. Sanbonmatsu, J. Ludwig, L. F. Katz, G. J. Duncan, J. R. Kling, and R. C. Kessler (2012). The long-term effects of moving to opportunity on youth outcomes. *Cityscape [Internet] 14*, 137–68.

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

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

Hanratty, M. H., S. A. McLanahan, and B. Pettit (2003). Los angeles site findings. In J. Goering and J. Feins (Eds.), *Choosing a Better Life*, pp. 245–274. Washington, DC: The Urban Institute Press.

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

Heckman, J. J. (1980). Sample selection bias as a specification error with an application to the estimation of labor supply functions. In J. P. Smith and J. F. Cogan (Eds.), *Female Labor Supply: Theory and Estimation*, pp. 206–248. Princeton University Press.

Heckman, J. J., V. J. Hotz, and J. R. Walker (1985, May). New evidence on the timing and spacing of births. *American Economic Review 75*(2), 179–184. Papers and Proceedings of the Ninety-Seventh Annual Meeting of the American Economic Association.

Heckman, J. J., S. H. Moon, R. Pinto, P. A. Savelyev, and A. Q. Yavitz (2010, July). Analyzing social experiments as implemented: A reexamination of the evidence from the HighScope Perry Preschool Program. *Quantitative Economics 1*(1), 1–46.

Heckman, J. J. and R. Pinto (2014). Causal analysis after haavelmo. *Econometric Theory*, 1–37.

Heckman, J. J., R. Pinto, and P. A. Savelyev (2013, October). Understanding the mechanisms through which an influential early childhood program boosted adult outcomes. *American Economic Review 103*(6), 2052–2086.

Heckman, J. J. and R. Robb (1985). Alternative methods for evaluating the impact of interventions. In J. J. Heckman and B. S. Singer (Eds.), *Longitudinal Analysis of Labor Market Data*, Volume 10, pp. 156–245. New York: Cambridge University Press.

Heckman, J. J. and S. Urzúa (2010). Comparing IV with Structural Models: What Simple IV Can and Cannot Identify. *Journal of Econometrics 156*(1), 27–37.

Heckman, J. J. and E. J. Vytlacil (1999, April). Local instrumental variables and latent variable models for identifying and bounding treatment effects. *Proceedings of the National Academy of Sciences 96*(8), 4730–4734.

Heckman, J. J. and E. J. Vytlacil (2005, May). Structural equations, treatment effects and econometric policy evaluation. *Econometrica 73*(3), 669–738.

Imbens, G. W. and J. D. Angrist (1994, March). Identification and estimation of local average treatment effects. *Econometrica 62*(2), 467–475.

Katz, L. F., J. Kling, and J. B. Liebman (2001). Moving to opportunity in boston: Early results of a randomized mobility experiment. *Quarterly Journal of Economics 116*, 607–654.

Katz, L. F., J. Kling, and J. B. Liebman (2003). Boston site findings. In J. Goering and J. Feins (Eds.), *Choosing a Better Life*, pp. 177–212. Washington, DC: The Urban Institute Press.

Kitamura, Y., , and J. Stoye (2014). Nonparametric analysis of random utility models. *Umpublished Maniscript*.

Kline, P. and C. Walters (2016). Evaluating public programs with close substitutes: The case of head start. *Umpublished Manuscript*.

Kline, P. and C. Walters (2017). Through the looking glass: Heckits, late, and numerical equivalence. *Umpublished working paper*.

Kling, J. R., J. B. Liebman, and L. F. Katz (2007). Experimental analysis of neighborhood effects. *Econometrica 75*, 83–119.

Kling, J. R., J. Ludwig, and L. F. Katz (2005). Neighborhood effects on crime for female and male youth: Evidence from a randomized housing voucher experiment. *The Quarterly Journal of Economics 120*, 87–130.

Ladd, H. F. and J. Ludwig (2003). The effects of mto on educational op-portunities in baltimore. In J. Goering and J. Feins (Eds.), *Choosing a Better Life*, pp. 117–151. Washington, DC: The Urban Institute Press.

Lee, S. and B. Salanié (2018). Identifying effects of multivalued treatments. *Econometrica 86*, 1939–1963.

Leventhal, T. and J. Brooks-Gunn (2003). New york site findings. In J. Goering and J. Feins (Eds.), *Choosing a Better Life*, pp. 213–244. Washington, DC: The Urban Institute Press.

Ludwig, J., G. J. Duncan, and J. Pinkston (2005). Housing mobility programs and economic self-sufficiency: Evidence from a randomized experiment. *Journal ofPublic Economics 89*, 131–156.

Ludwig, J., P. Hirschfield, and G. J. Duncan (2001). Urban poverty and juvenile crime: Evidence from a randomized housing mobility experiment. *Quarterly Journal of Economics 116*, 665–679.

Ludwig, J., J. R. Kling, L. F. Katz, J. B. Liebman, G. J. Duncan, R. C. Kessler, and L. Sanbonmatsu (2008). What can we learn about neighborhood effects from the moving to opportunity experiment. *American Journal of Sociology 114*(1), 144–88.

Magnus, J. and H. Neudecker (1999). *Matrix Differential Calculus with Applications in Statistics and Econometrics* (2 ed.). Wiley.

McFadden, D. (2005). Revealed stochastic preference: A synthesis. *Economic Theory 26*, 245–264.

McFadden, D. and K. Richter (1991). Stochastic rationality and revealed stochastic preference. *Preferences, Uncertainty and Rationality,ed. by J. Chipman, D. McFadden, and M.K. Richter. Boulder: Westview Press*, 161–186.

Mogstad, M., S. Andres, and A. Torgovitsky (2017). Using instrumental variables for inference about policy relevant treatment effects. *NBER Working Paper* (23568).

Mogstad, M. and A. Torgovitsky (2018). Identification and extrapolation of causal effects with instrumental variables. *Annual Review of Economics 2*, 577–613.

Orr, L., J. D. Feins, R. Jacob, and E. Beecroft (2003). *Moving to Opportunity Interim Impacts Evaluation.* Washington, DC: U.S. Department of Housing and Urban Development Office of Policy Development & Research.

Pinto, R. (2014). Selection bias in a controlled experiment: The case of moving to opportunity. *Umpublished Manuscript*.

Pinto, R. (2016). Monotonict incentives. *Working Paper*.

Powell, J. L. (1994). Estimation of semiparametric models. In R. Engle and D. McFadden (Eds.), *Handbook of Econometrics, Volume 4*, pp. 2443–2521. Amsterdam: Elsevier.

Richter, M. K. (1971). Rational choice. In *Preferences, Utility, and Demand: A Minnesota Symposium* (1 ed.), Volume 1, Chapter 2, pp. 29–58. New York: Harcourt, Brace, Jovanovich.

Sampson, R. J. (2008). Moving to inequality: Neighborhood effects and experiments meet social structure. *American Journal of Sociology 114*, 189 – 231.

Samuelson, P. A. (1938). A note on the pure theory of consumer's behaviour. *Economica 5*(17), 61–71.

Sanbonmatsu, L., J. Ludwig, L. F. Katz, L. A. Gennetian, G. J. Duncan, R. C. Kessler, E. Adam, T. McDade, and S. T. Lindau (2011). Moving to opportunity for fair housing demonstration program: Final impacts evaluation. *Department of Housing and Urban Development (HUD), Office of Policy Development and Research*.

Theil, H. (1953). *Esimation and Simultaneous Correlation in Complete Equation Systems.* The Hague: Central Planning Bureau. Mimeographed memorandum.

Theil, H. (1958). *Economic Forecasts and Policy.* Number 15 in Contributions to Economic Analysis. Amsterdam: North-Holland Publishing Company.

Vytlacil, E. J. (2004). Ordered discrete choice selection models: Equivalence, nonequivalence, and representation results. Unpublished manuscript, Stanford University, Department of Economics.

Wilson, W. J. (2009). *More than just race: Being black and poor in the inner city.* New York, NY: WW Norton and Company, Inc.

Wooldridge, J. M. (2015). Control function methods in applied econometrics. *Journal of Human Resources*, 420–445.

Yazejian, N. and D. M. Bryant (2012). Educare implementation study findings. Technical report, Frank Porter Graham Child Development Institute, Chapel Hill, NC.