working paper

12 08R

Marginal Neighborhood Effects from Moving to Opportunity

Dionissi Aliprantis and Francisca G.-C. Richter

FEDERAL RESERVE BANK OF CLEVELAND

Working Paper 12-08R October 2012*

Marginal Neighborhood Effects from Moving to Opportunity

Dionissi Aliprantis and Francisca G.-C. Richter

This paper estimates Marginal Treatment Effects (MTEs) of neighborhood quality from the Moving to Opportunity (MTO) housing mobility experiment in a model with multiple treatment levels. We propose and implement a new identification strategy that exploits the identification of the idiosyncratic component of an ordered choice model. Due to the limited changes in neighborhood quality induced by MTO, we only estimate MTEs of moving from the first to second decile of the national distribution of neighborhood quality. These MTEs are heterogeneous over observable characteristics: Labor market outcomes were affected most positively for individuals at the sites in which larger changes in neighborhood quality were induced by MTO. Estimated MTEs are also heterogeneous over unobservables, which we consider evidence in favor of selection occurring in a model with essential heterogeneity. Although there is not enough structure in our model to clearly interpret MTE heterogeneity, we discuss possible reasons for the surprising result that effects are best for those with characteristics that make them less likely to move without the program.

Keywords: Marginal Treatment Effect, Essential Heterogeneity, Strong Ignorability, Local Average Treatment Effect, Average Causal Response, Moving to Opportunity.

JEL Classification Numbers: C31, C36, C50, D04, I38, R23.

*First version posted March 2012 under the title “Local Average Neighborhood Effects from Moving to Opportunity.”

Dionissi Aliprantis is at the Federal Reserve Bank of Cleveland (dionissi.aliprantis@clev.frb.org). Francisca G.-C. Richter is also at the Federal Reserve Bank of Cleveland (francisca.g.richter@clev.frb.org). The authors thank Jeffrey Kling, Joel Elvery, Fabian Lange, Jon James, Ben Craig, Carlos Lamarche, and Alvin Murphy for helpful comments. They also thank Emily Burgen and Nelson Oliver for valuable research assistance, and Paul Joice at HUD for his assistance with the data set.

1 Introduction

Since neighborhood externalities are widely believed to be a major determinant of outcomes in the United States, empirical evidence on the existence and magnitude of neighborhood effects has been of great interest to social scientists (Wilson (1987)). Despite this interest, conclusively documenting the nature of neighborhood effects has proven difficult because it is infeasible to randomly allocate neighborhood characteristics to individuals. Researchers have yet to agree on the most appropriate methodology for investigating neighborhood effects under endogenous neighborhood selection.

The results from the Moving to Opportunity (MTO) housing mobility program are a prominent example of the difficulties confronted when studying neighborhood effects. In MTO rental vouchers were randomly assigned to public housing residents of poor neighborhoods that could be used to relocate to low-poverty neighborhoods. The lack of large effects from the program were disappointing (Kling et al. (2007), Sanbonmatsu et al. (2006)), especially given the strong positive effects from the program serving as its motivation, Gautreaux (Polikoff (2006), Rubinowitz and Rosenbaum (2000)). Furthermore, there has been disagreement regarding the most appropriate interpretation of results from MTO. One view interprets Intent to Treat (ITT) and Treatment on the Treated (TOT) effects from moving with an MTO voucher in terms of neighborhood effects (Ludwig et al. (2008)), while another view interprets these effects as program effects and is agnostic about their relationship to neighborhood effects (Clampet-Lundquist and Massey (2008), Aliprantis (2012)). Despite this disagreement about the interpretation of effects from moving with an MTO voucher, there have been analyses estimating effects from neighborhood characteristics using the MTO data (Kling et al. (2007), Ludwig and Kling (2007), Ludwig et al. (2008)).

This paper builds on previous analyses of effects from neighborhood quality using MTO voucher assignment as an instrumental variable. We take the analysis in Kling et al. (2007) as our starting point and relax two key identifying assumptions as suggested in Aliprantis (2012). We first relax the assumption of homogeneous Marginal Treatment Effects (MTEs) from neighborhood quality over unobservables. We also create and use in our analysis a linear index of neighborhood quality that uses several neighborhood characteristics in addition to the neighborhood poverty rate. Neighborhood externalities are captured in the resulting model through this index of neighborhood quality, with neighborhood effects defined as effects of moving between neighborhoods of varying quality.

In order to identify neighborhood effects under these relaxed assumptions we specify and estimate a joint model of potential outcomes and selection into a multi-valued treatment. Our model adapts the assumptions from Heckman et al. (2006) to the specific application of MTO. We identify the unobservable, idiosyncratic component of each individual's latent index in an ordered choice model, something not possible in the case of a binary treatment. After imposing parametric assumptions on expected outcomes, we are able to estimate MTEs over a region of the observed and unobserved components of the choice model. The region over which we can identify MTEs is determined by the mobility induced by the experiment, and our parametric assumptions allow us to extend this set beyond the subset of compliers used in the estimation of Local Average Treatment Effects (LATEs).

The estimation results from the ordered choice model indicate that vouchers had strongly heterogeneous effects across sites and voucher type. Residents in Los Angeles and New York City experienced much greater changes in neighborhood quality from the offer of a voucher than did participants in Baltimore or Chicago. The ordered choice model estimates also show that Section 8 vouchers were more effective at improving neighborhood quality at very low levels, but that MTO experimental vouchers became equally effective well before the fifth percentile of the national distribution of neighborhood quality. The experimental voucher was much more effective than the Section 8 voucher by the 10th percentile of quality and remained so at all higher levels of quality.

Perhaps more important than the heterogeneity in mobility induced by MTO is the fact that MTO only induced a very small share of participants to move to high quality neighborhoods. Categorizing neighborhood quality into the deciles of the national distribution, the most common single move induced by MTO was from the first to second decile. Nevertheless, this move was only induced for about 11 percent of our restricted sample of MTO volunteers. Since the subgroups of compliers induced to move across higher margins of quality represent even smaller shares of our sample, we only estimate MTE and LATE parameters pertaining to moves from the first to the second decile of the national distribution of neighborhood quality. Even these parameter are imprecisely estimated due to the fact that voucher assignment induced few individuals into higher quality neighborhoods.

Effects of neighborhood quality on employment and household income are heterogeneous and have large magnitudes. For example, even conditioning on having the median unobservable, the MTE on household income at one extreme of observables is estimated to be over $5,000, while the MTE at the other extreme is actually negative. Effects were similarly heterogeneous by unobservables, even accounting for observables.

The heterogeneity we find in MTEs indicates that improving neighborhood quality would have the most positive effects on individuals with either observed and/or unobserved characteristics that would make them the least likely to select into higher quality neighborhoods without a voucher. There is not enough structure in our model to clearly interpret this MTE heterogeneity, but we hypothesize that other social programs, violence, and information could all play a role in driving this heterogeneity. Furthermore, one preliminary conclusion from heterogeneity over observables is that labor market outcomes were affected most positively for individuals in Los Angeles and New York City, the two sites with the most mobility induced by MTO. We interpret MTE heterogeneity over unobservables as evidence in favor of selection occurring in a model under essential heterogeneity rather than strong ignorability. We also broadly interpret MTE heterogeneity over unobservables to suggest the need for richer models of how neighborhood characteristics impact outcomes.

The remainder of the paper is organized as follows: Section 2 develops the ordered choice model, defines the treatment effect parameters of interest, and discusses assumptions necessary to identify these parameters. Section 3 describes the MTO experiment and Section 4 discusses the data used in estimation. Sections 5 and 6 present the estimation results, and Section 7 concludes.

2 Model

The model we use is based on the Marginal Treatment Effect (MTE) framework developed in Heckman et al. (2006) and Heckman and Vytlacil (2005), henceforth referred to as HUV and HV, respectively, and originally introduced by Björklund and Moffitt (1987). In a multi-valued treatment setting, identification of level-specific MTEs over the whole range of unobserved heterogeneity would typically be achieved using continuous instruments for each transition. Experimental voucher assignment from MTO is not such a set of instruments. Thus while we are able to relax some of the identifying assumptions in Kling et al. (2007), the three-level instrument from MTO only allows us to identify level-specific parameters related to those transitions induced by the experiment.

2.1 Ordered Choice Model

We are interested in estimating the effects that neighborhood quality has on low-income individuals' outcomes such as employment, income, and health. We begin by modeling the way agents select into different levels of a multi-valued treatment- in this case, neighborhood quality. Consider a model in which there is a treatment D with J levels (ie, $D \in \{1, \ldots, J\} \equiv \mathcal{D}$), a binary instrument Z (ie, $Z \in \{0, 1\}$), and individuals have observable characteristics X. We assume that for each treatment level $j \leq J-1$ there is a latent index D_j^* representing the net marginal benefit of moving to treatment level $j+1$ from level j. Agents select level k^* that maximizes total utility $u(D)$:

$$k^* = \underset{k \in \{1, \ldots, J\}}{\operatorname{argmax}} \; u(k) \tag{1}$$

where

$$u(k) = \begin{cases} u(1) & \text{if} \quad k = 1; \\ u(1) + \sum_{j=1}^{k-1} D_j^* & \text{if} \quad k \geq 2. \end{cases}$$

We specify the $J-1$ latent indeces to be functions of observable characteristics (Z, X) and unobservable characteristics V with the following structure:

A1: $D_{ij}^* = \mu(X_i) + \gamma_j Z_i - V_i - C_j$ for $j = 1, \ldots, J-1$,

A2: $C_j < C_{j+1}$ for $j = 1, \ldots, J-1$ with $C_0 = -\infty$ and $C_J = \infty$,

A3: $C_j - \gamma_j < C_{j+1} - \gamma_{j+1}$ for $j = 1, \ldots, J-1$ with $C_0 - \gamma_0 = -\infty$ and $C_J + \gamma_J = \infty$.

Here $\mu(X)$ is the gain from moving up one level and is independent of the level. C_j is a transition-specific cost that increases at each level. V is an absolutely continuously distributed random variable representing the unobserved cost for individual i from moving up one level. And γ_j can be interpreted as a reduction in the transition specific cost brought about by varying the instrument that preserves the cost ranking among levels (A3). Note that the initial level $u(1)$ is important only as a way of indexing the model.

Let $D_0^* > \max\{0, D_1^*\}$ and define $D_J^* \equiv -\infty$. Then given Z, the net marginal benefit D_j^* is decreasing in j, so for $k^* \in \{1, \ldots, J\}$

$$D = k^* \quad \iff \quad D_{k^*}^* \leq 0 < D_{k^*-1}^*. \tag{2}$$

The model described in Equation 1, together with assumptions A1-A3, allow us to write condition 2 as an ordered choice model in terms of the unobservable characteristics that determine selection into treatment, V:

$$D = j \quad \iff \quad \mu(X) + \gamma_j Z - C_j < V \leq \mu(X) + \gamma_{j-1} Z - C_{j-1}. \tag{3}$$

Denoting the CDF of V conditional on X by $F_{V|X}(\cdot)$, we can define $\pi_j^Z(X) \equiv Pr(D > j|Z, X) = F_{V|X}(\mu(X) + \gamma_j Z - C_j)$ and $U_D \equiv F_{V|X}(V|X)$. Defining $\pi_0^Z \equiv 1$ for the sake of exposition, we can alternatively write the model in terms of the conditional quantiles of V, U_D:

$$D(U_D, X, Z) = \sum_{j=1}^{J} j \times \mathbf{1}\left\{\pi_j^Z(X) < U_D \leq \pi_{j-1}^Z(X)\right\}.$$

Figure 1 shows the ordered choice model in A1-A3 graphically. The top panel reflects marginal utilities D_j^* and the bottom panel illustrates the level of utility. $D = k$ is selected by the agent with the given realization of (Z, X, V) since $D_{k-1}^* > 0$ and $D_k^* \leq 0$.

2.2 Outcomes and Treatment Effect Parameters

We are interested in learning about features of the joint distribution of some outcome Y given treatment levels $D \in \{1, \ldots, J\}$, observable characteristics X, and the unobservable component of selection into treatment, U_D. In this analysis we will focus on the conditional mean $E[Y(D, X)|U_D = u_D]$. We assume potential outcomes are a function of treatment level, observable characteristics, and level-specific unobservable characteristics as follows:

$$Y_j(X, U_j) = Y(D = j, X, U_j) = \mu_j(X) + U_j \quad \text{for} \quad j = 1, \ldots, J. \tag{4}$$

As in HUV we add the following assumptions to A1-A3:

A4 $(U_{ij}, U_{Di}) \perp\!\!\!\perp Z_i | X_i$ for all $j = 1, \ldots, J$

A5 $\gamma_j \geq 0$ for all $j = 1, \ldots, J-1$ and $\gamma_j > 0$ for at least one $j \in \{1, \ldots, J-1\}$

A6 $U_{Di} \sim U[0, 1]$

A7 $E[\,|Y_j|\,] < \infty$ for all $j = 1, \ldots, J$

A8 $0 < Pr(D_i = j|X_i) < 1$ for all $j = 1, \ldots, J$ for all $X \subset \text{supp}X$

Given a realization of $(X, U_D) = (x, u_D)$, the Marginal Treatment Effect (MTE) for the transition j to $j+1$ is defined as the average effect of moving individuals with (x, u_D) from level j to $j+1$:

$$\triangle_{j,j+1}^{MTE}(x, u_D) \equiv E[Y_{j+1} - Y_j | X = x, U_D = u_D] \text{ for } j = 1, \ldots, J-1. \tag{5}$$

Note that assuming $U_j \equiv \tilde{U}$ for all j, as in the model in Kling et al. (2007), restricts the MTE to be homogeneous for all U_D.[1]

Given a realization of $X = x$, the corresponding Local Average Treatment Effect (LATE) for individuals with $U_D \in (a, b)$ is defined as the average MTE over that interval:

$$\triangle_{j,j+1}^{LATE}(x, a, b) \equiv \frac{\int_a^b \triangle_{j,j+1}^{MTE}(x, u_D) du_D}{b - a} \text{ for } j = 1, \ldots, J-1. \tag{6}$$

2.3 Identification of MTE and LATE Parameters

One way of identifying MTE parameters in this model is to augment assumptions A1-A8, as done by HUV, with an assumption about the ordered choice model. HUV assumes there exist instrumental variables W_j for each $j = 1, \ldots, J-1$ such that the distribution of $C_j(W_j)$, conditional on Z, X, and $\{C_h : h \neq j\}$, is nondegenerate and continuous. We label this assumption, which does not hold in our model, A9. Under A9 each margin of choice can be varied exogenously and independently of all others.

When evaluated at $U_D = \pi_j^Z(x)$, $\triangle_{j,j+1}^{MTE}(x, \pi_j^Z(x))$ represents the *gross* gain of moving from j to $j+1$ for individuals that are indifferent between levels j and $j+1$. HUV show that index sufficiency holds in this model so that $E[Y|Z, X]$ is equivalent to $E[Y|\boldsymbol{\pi}^Z(X)]$, where $\boldsymbol{\pi} \equiv [\pi_1^Z(X), \pi_2^Z(X), \ldots, \pi_{J-1}^Z(X)]$. HUV further show that $E[Y|\boldsymbol{\pi}^Z]$ is differentiable due to Assumptions A4, A6, and A7. The j to $j+1$ MTE can be interpreted as the change in mean outcome due to externally increasing π_j^Z while leaving all other π_k^Z's fixed for $k \neq j$:

$$\triangle_{j,j+1}^{MTE}(x, \pi_j) = \frac{\partial E[Y|X = x, \boldsymbol{\pi}^Z(x) = \boldsymbol{\pi}]}{\partial \pi_j}. \tag{7}$$

Identification of MTE parameters is achieved in HUV using the exogenous variation in $\pi_j^Z(x)$ induced by the W_j to estimate the right hand side of Equation 7.

In the context of residential choice, it is difficult to imagine a set of instruments $\{W_j\}$ each of which exogenously varies one margin of choice while leaving all other margins of choice unaffected. In large part, this problem arises because, unlike schooling, neighborhood quality levels are not clearly defined. Even if they existed, it is doubtful these instruments could be manipulated to identify the MTE function over the entire support of the distribution of (X, U_D). It would be more likely that each level-specific instrument W_j would vary the choice margin over some interval, but

[1] This is true even if $COV(\tilde{U}, U_D) \neq 0|X$.

not over the entire unit interval. A discussion of related issues can be found in Carneiro et al. (2011) for a binary context.

The discrete instrument of voucher assignment in MTO does not continuously vary $\pi_j^Z(X)$ for any j nor does it provide an independent source of variation for each of the transitions. Violation of assumption A9 leads us to adopt an alternative approach to identifying parameters that is based on the identification of U_D in the ordered choice model and is therefore not possible in the binary model.

Consider first the case of fixing observable characteristics $X = x$. Two such examples with $D \in \{1, 2, 3\}$ are illustrated in Figures 2 and 3. Remembering that $D = j$ iff $\pi_j^Z < U_D \leq \pi_{j-1}^Z$ and assuming $\pi_j^0(x) < \pi_j^1(x) < \pi_{j-1}^0(x)$, an implication from the ordered choice model is that when $Z = 0$ an individual with $X = x$ and $U_D \in [\pi_j^0(x), \pi_j^1(x)]$ selects into $D = j$, but exogenously changing Z to 1 induces such individuals into $D = j + 1$. Thus for $u_D \in [\pi_j^0(x), \pi_j^1(x)]$,

$$\begin{aligned} E[Y|x, u_D, Z = 1] - E[Y|x, u_D, Z = 0] &= E[Y(j + 1)|x, u_D] - E[Y(j)|x, u_D] \\ &= \triangle_{j,j+1}^{MTE}(x, u_D). \end{aligned} \tag{8}$$

We can identify LATEs non-parametrically by averaging MTEs over $u_D \in [\pi_j^0(x), \pi_j^1(x)]$.

The case when $X = x$ is fixed motivates the more general case in which X is not fixed. The basic idea of the identification strategy can be seen in Figure 4. Define the intervals $\mathcal{M}$ and $\mathcal{Y}$ such that $\mu(X) \subset \mathcal{M} \subset \mathbb{R}$ and $E[Y(D = j)|\mu(X), u_D] \subset \mathcal{Y} \subset \mathbb{R}$ for all $j \in \{1, \ldots, J\}$. Then, as illustrated in Figure 4, we might assume $E[Y(D = j)|\mu(X), u_D]$ is a surface in the cartesian product space $\mathcal{Y} \times \mathcal{M} \times [0, 1]$. The joint distribution of individuals' $(\mu(X), U_D, Z)$ will determine the supports over which we can identify $E[Y(D = j)|\mu(X), u_D]$ for each j, and hence which parameters we can identify from the experiment.

Given a parametrically specified ordered choice model, we estimate parameters based on both non-parametric and parametric specifications of $E[Y(D = j)|\mu(X), u_D]$. The specific parameterizations of potential outcomes we adopt are discussed in Section 6.2 and Appendix B, and we provide an intuitive discussion of identification for fixed observables in greater depth in Appendix A. A related discussion can also be found in Section VI A of HUV.

2.4 Discussion of Identifying Assumptions

We believe Assumptions A1-A8 represent the weakest assumptions under which we can identify interesting parameters in a model with a multi-valued treatment. We now consider alternative sets of identifying assumptions and discuss why we believe these assumptions are either too strong or too weak relative to A1-A8.

2.4.1 Strengthening A1-A8

One alternative approach would be to strengthen assumptions A1-A8. One particular assumption would allow us to estimate average treatment effects over the support of the distribution of a continuous treatment D using the generalized propensity score as developed in Imai and van Dyk (2004) or Hirano and Imbens (2005). However, while relatively standard, the Strong Ignorability (SI) assumption necessary for identification in Imai and van Dyk (2004) is restrictive relative to the framework we have adopted from HV and HUV.[2] When f denotes the distribution of potential outcomes, SI can be written as:

$$f\{y_j|X\} = f\{y_j|X, U_D \in [a,b]\} \text{ for all } j \in \{1,\dots,J\}. \tag{9}$$

From our specification of potential outcomes in Equation 4, this is the same as:

$$f\{u_j|X\} = f\{u_j|X, U_D \in [a,b]\} \text{ for all } j \in \{1,\dots,J\}, \tag{10}$$

or

$$U_j \perp\!\!\!\perp U_D|X \text{ for all } j \in \{1,\dots,J\}. \tag{11}$$

When calculating $E[Y(D=j)|U_D]$ as in Figures 2 and 3, the expectation is taken over the distribution of U_j conditional on X and U_D. Thus SI requires that conditional on X, the distribution of the Y_j, and therefore their expected values as well, would have to be the same for all values of U_D as shown in Figure 5. This assumption is most likely to hold when observable characteristics in X are able to explain most of the variability in choice, so deciding whether to adopt the SI assumption will depend on the particular application and data available.

Figure 6 shows a binary example to help illustrate the differences between the heterogeneity in treatment effects allowed under our assumptions A1-A8 and under SI. Let $\theta = \mu(X)$ be an index of observable characteristics. The average effect of treatment varies across θ as shown in the top panel of the Figure. In the center panel we can see a cross section of potential outcomes conditional on $\theta = \theta^*$. Since $E[\beta|\theta] = E[Y(1) - Y(0)|\theta]$ is the same for all u_D conditional on $\theta = \theta^*$, any variation in treatment identifies $E[\beta|\theta]$. We could use variation in treatment induced by an instrument, but we could also simply compare those individuals in the population or control group with $u_D < \pi_1^0$ and those with $u_D > \pi_1^0$ to estimate the treatment effect. That is, although a valid instrument is likely to make the assumption more plausible, when matching under SI there is no theoretical need for an instrument.

In contrast to the center panel, the bottom panel shows a possible example of MTEs that depend on u_D even conditional on θ. This is defined in HUV as a model with Essential Heterogeneity (EH).[3] In this case we need an instrument to generate variation in treatment status, and the variation generated by the instrument determines what part of the distribution of $Y(2) - Y(1)|\theta^*$ we can identify. Since $E[Y(2) - Y(1)|\theta^*] = \int_0^1 \triangle_{1,2}^{MTE}(\theta^*, u_D)du_D$, in the example in the bottom panel we cannot identify $E[Y(2) - Y(1)|\theta^*]$, but rather only $\int_{\pi_1^0}^{\pi_1^1} \triangle_{1,2}^{MTE}(\theta^*, u_D)du_D$. That the treatment effects we

[2] See Imbens (2004) for a discussion of the SI assumption in models with a binary treatment.
[3] Essential heterogeneity between levels j and $j+1$ is defined as

EH $COV(U_{j+1} - U_j, V) \mid X \neq 0.$

can identify are determined by the response of individuals to the instrument re-emphasizes that the neighborhood effects identified by MTO, or any other housing mobility experiment, depend on how people endogenously respond to the experiment. That is, under EH it is not possible to clearly interpret the neighborhood effects we observe through MTO without first understanding how the experiment impacted selection into treatment (Aliprantis (2012), Clampet-Lundquist and Massey (2008), Sampson (2008)).

2.4.2 Weakening A1-A8

Another alternative would be to weaken A1-A8, which would allow us to estimate the Average Causal Response (ACR) parameter introduced in Angrist and Imbens (1995). Under Assumptions A1-A8 the ACR is:

$$\triangle^{ACR}(Z=0,Z=1) = \frac{E[Y|Z=1]-E[Y|Z=0]}{E[D|Z=1]-E[D|Z=0]} = \frac{\sum_{j=1}^{J-1}[\int_{\pi_j^0}^{\pi_j^1} \triangle_{j,j+1}^{MTE}(u_D)du_D]}{\sum_{j=1}^{J-1}(j+1)\times(\pi_j^1-\pi_j^0)}$$

By relaxing Assumptions A2 and A3 we could specify a set of identifying assumptions equivalent to the assumptions in Angrist and Imbens (1995) (See Vytlacil (2006) for a proof.), and we would still be allowing for EH. We do not estimate ACRs under these weaker assumptions, however, because they yield only one summary parameter that is quite difficult to interpret. By imposing the structure of the choice model in Section 2.1 we are able to allow for EH while at the same time decomposing the ACR into its contributing LATEs. These components of the ACR are considerably more interesting than the single ACR parameter by itself. Given our current econometric tools we believe assumptions A1-A8 strike the right balance of structure for interpretation.

3 Moving To Opportunity (MTO)

Moving To Opportunity (MTO) was inspired by the promising results of the Gautreaux program. Following a class-action lawsuit led by Dorothy Gautreaux, in 1976 the Supreme Court ordered the Department of Housing and Urban Development (HUD) and the Chicago Housing Authority (CHA) to remedy the extreme racial segregation experienced by public-housing residents in Chicago. One of the resulting programs gave families awarded Section 8 public housing vouchers the ability to use them beyond the territory of CHA, giving families the option to be relocated either to suburbs that were less than 30 percent black or to black neighborhoods in the city that were forecast to undergo “revitalization” (Polikoff (2006)).

The initial relocation process of the Gautreaux program created a quasi-experiment, and its results indicated housing mobility could be an effective policy. Relative to city movers, suburban movers from Gautreaux were more likely to be employed (Mendenhall et al. (2006)), and the children of suburban movers attended better schools, were more likely to complete high school, attend college, be employed, and had higher wages than city movers (Rosenbaum (1995)).[4]

MTO was designed to replicate these beneficial effects, offering housing vouchers to eligible households between September 1994 and July 1998 in Baltimore, Boston, Chicago, Los Angeles, and New York (Goering (2003)). Households were eligible to participate in MTO if they were low-income, had at least one child under 18, were residing in either public housing or Section 8 project-based housing located in a census tract with a poverty rate of at least 40%, were current in their rent payment, and all families members were on the current lease and were without criminal records (Orr et al. (2003)).

Families were drawn from the MTO waiting list through a random lottery. After being drawn, families were randomly allocated into one of three treatment groups. The *experimental* group was offered Section 8 housing vouchers, but were restricted to using them in census tracts with 1990 poverty rates of less than 10 percent. However, after one year had passed, families in the *experimental* group were then unrestricted in where they used their Section 8 vouchers. Families in this group were also provided with counseling and education through a local non-profit. Families in the *Section-8 only* comparison group were provided with no counseling, and were offered Section 8 housing vouchers without any restriction on their place of use. And families in the *control* group continued receiving project-based assistance.[5]

4 Data

The first source of data we use in our analysis is the MTO Interim Evaluation. The MTO Interim Evaluation contains variables listing the census tracts in which households lived at both the baseline and in 2002, the time the interim evaluation was conducted. These census tracts are used to merge the MTO sample with decennial census data from the National Historical Geographic Information System (NHGIS, Minnesota Population Center (2004)), which provide measures of neighborhood characteristics.

4.1 Variables

A variable measuring neighborhood quality is created as a linear combination of several neighborhood characteristics. Neighborhood characteristics measured by NHGIS variables are first transformed into percentiles of the national distribution from the 2000 census. Principal components analysis is then used to determine which single vector combines the most information about the national distribution of the poverty rate, the percent with high school degrees, the percent with BAs, the percent of single-headed households, the male Employed-to-Population Ratio (EPR), and the female unemployment rate. Table 1 shows that the resulting univariate index explains 69 per-

[4] It has also been found that suburban movers have much lower male youth mortality rates Votruba and Kling (2009) and tend to stay in high-income suburban neighborhoods many years after their initial placement (DeLuca and Rosenbaum (2003), Keels et al. (2005)).

[5] Section 8 vouchers pay part of a tenant's private market rent. Project-based assistance gives the option of a reduced-rent unit tied to a specific structure.

cent of the variance of these neighborhood characteristics, and that no additional eigenvector would explain more than 11 percent of the variance of these variables. Table 2 displays the coefficients relating each of these variables to the index vector. Note that the magnitudes of the coefficients for most variables are similar to the magnitude of the coefficient for poverty.

Quality of the baseline and 2002 neighborhood of residence is measured using 2000 Census data. This measurement does not allow us to account for expected neighborhood change, a factor that may contribute to the decision to move from one's baseline neighborhood. Perceived quality improvements, like new magnet schools in the baseline neighborhood, will reduce the benefit of moving. Due to this measurement issue, in our model, these components of the $\{C_j\}$ will be captured by the unobserved component of choice V.

Besides neighborhood quality, other baseline characteristic of the MTO households used in this model are whether the respondent had family living in their neighborhood of residence. Also included were baseline variables stating whether a member of the household was a victim of a crime in the previous 6 months and whether there were teenage children in the household. Other than site of residence, no other observable characteristics were not used in the estimation because the estimated ordered choice model coefficients on other observables all had p-values over 0.10.

Outcome variables for adults from the MTO Interim Evaluation include the self-reported total household income (all sources summed), the individual earnings in 2001 of the sample adult, the labor market status of the adult at the time of the interim survey (ie, Whether they were employed, unemployed, or not in the labor force at all.), welfare benefits, and the respondent's body mass index (BMI). Weights are used in constructing all estimates.[6]

4.2 Sample and Descriptive Statistics

We focus on the adults in the MTO Interim Evaluation sample. Figures 7 and 8 reproduce the broad result from Aliprantis (2012) that MTO induced small shares of adult participants into high quality neighborhoods. Similar results for school quality can be seen in Figure 9.

In order to facilitate the estimation of the ordered choice model, the ensuing analysis is focused on a sample that is restricted in three ways to ensure that we are focusing on a relatively homogeneous population. We first exclude Boston from the analysis because it is clearly an outlier relative to the other MTO sites. In Figure 8a we see that unlike all of the other sites, the baseline neighborhood quality in Boston was not confined to the first percentiles of the national distribution of neighborhood quality. Table 3 quantifies these differences precisely. We can see that the control group in Boston looks more like the experimental group in every other site. Observations from Boston comprised a little less than a quarter of the full sample.

Second, we drop all households living at baseline in a neighborhood above the tenth percentile

[6] Weights are used for two reasons. First, random assignment ratios varied both from site to site and over different time periods of sample recruitment. Randomization ratio weights are used to create samples representing the same number of people across groups within each site-period. This ensures neighborhood effects are not conflated with time trends. Second, sampling weights must be used to account for the sub-sampling procedures used during the interim evaluation data collection.

of the national distribution of quality. Returning to Table 3, we can see that the median baseline neighborhood quality for MTO participants was below the first percentile of the national distribution. For Chicago, Los Angeles, and New York City, nearly all participants lived at baseline in neighborhoods below the 10th percentile of the national distribution. In Baltimore, however, at baseline about 20 percent of participants lived in higher quality neighborhoods, driven mainly by the male EPR and the share of adults holding a BA in their neighborhoods. These individuals are also dropped from our sample.

Finally, we drop individuals who moved to neighborhoods in 2002 above the median of the national distribution of quality in 2000. We can see from Figure 10c that this is a small subsample to discard.

The resulting sample used in the analysis has approximately 2,500 adults (approximately 65 percent of the original adult sample). Figure 10d shows the distribution of the sample both at the baseline and in 2002. Seventy five percent of the sample lives in 2002 in a neighborhood below the 10th percentile of the 2000 national distribution of neighborhood quality.

5 Ordered Choice Model Estimates

5.1 Specification

In the ordered choice model from Section 2.1 the marginal benefit of choosing to move from treatment level j to $j+1$ is

$$D_j^* = \mu(X) + Z\gamma_j - C_j - V.$$

We specify

$$\mu(X) = \beta_1 X_1 + \cdots + \beta_m X_m \tag{12}$$

$$\gamma_j = \gamma_0 + \gamma_j \tag{13}$$

$$C_j = \delta_0 + \delta_j \tag{14}$$

where both γ_0 and δ_0 are site-specific, and we assume $V \sim$ iid $\mathcal{N}(0,1)$. X is composed of variables measuring baseline neighborhood quality and whether the respondent had family living in the neighborhood of residence, whether a member of the household was a victim of a crime in the previous 6 months and whether there were teenage children in the household.

The instrument Z is an indicator variable for assignment to the experimental voucher group in the MTO experiment: $Z_i = \mathbf{1}\{i \text{ is in experimental group}\}$.[7] Like Galiani et al. (2012), we interpret Z as the random assignment of a reduction in the cost to accessing a better quality neighborhood.

[7]We actually define binary instruments for both the experimental and Section 8 groups, (Z^{exp}, Z^{S8}). Much of the discussion in the paper focuses on comparing the experimental voucher with the control treatment for the sake of exposition. The Section 8 group is used in generating ordered choice model and parametric MTE estimates, but is not used in the estimation of semi-parametric LATEs.

An implication of the ordered choice model that was stated in Equation 3 is:

$$D = j \quad \Longleftrightarrow \quad \mu(X) + Z\gamma_j - C_j < V \leq \mu(X) + Z\gamma_{j-1} - C_{j-1}. \tag{15}$$

So if treatment levels are defined such that $D = j$ for $q^* \in [\underline{q}_j, \overline{q}_j]$, then the probability of observing $D = j$ is

$$Pr(D = j) \quad = \quad F_V[\mu(X) + Z\gamma(\underline{q}_j) - C(\underline{q}_j)] \; - \; F_V[\mu(X) + Z\gamma(\overline{q}_j) - C(\overline{q}_j)],$$

allowing us to write the log-likelihood of the data as

$$\mathcal{LL} = \sum_{i=1}^{N} \sum_{j=1}^{J} \mathbf{1}\{D_i = j\} \, ln(Pr(D_i = j)). \tag{16}$$

This ordered probit model is estimated with the gologit2 procedure in Stata.

5.2 The Identification of Unobservables

We now discuss how we use the parameter estimates from the discrete ordered choice model $\widehat{\theta}_J \equiv (\widehat{\beta}, \widehat{\gamma}, \widehat{\delta})$ to recover estimates of each individual's unobservable V. If we take the limit of the ordered choice model as $J \rightarrow \infty$, the benefit of moving to level $j+1$ from j, D^*_j, becomes the marginal benefit of moving to a neighborhood with higher quality. We denote the marginal benefit of moving to a higher quality neighborhood when in a neighborhood of quality q as

$$D^*(q) = \mu(X) + Z\gamma(q) - C(q) - V$$

where the cost and program effect components of marginal benefit are now allowed to be continuous functions of quality:

$$\mu(X) = \beta_1 X_1 + \cdots + \beta_m X_m \tag{17}$$

$$\gamma(q) = f(q) \tag{18}$$

$$C(q) = g(q). \tag{19}$$

Given estimates of $\gamma(\overline{q}_j) = \gamma_j$ from the ordered choice model, we linearly interpolate

$$\gamma(q) = \gamma_0 + \gamma_j + (q - \overline{q}_j)\left(\frac{\gamma_{j+1} - \gamma_j}{\overline{q}_{j+1} - \overline{q}_j}\right) \quad \text{for } q \in (\overline{q}_j, \overline{q}_{j+1}) = (\underline{q}_{j+1}, \overline{q}_{j+1})$$

and an analogous function g for cost. The piece-wise linear functions created by interpolating between the $\{\gamma_j\}$ and $\{C_j\}$ located at the $\{\overline{q}_j\}_{j=1}^{J-1}$ allow for great flexibility in the shape these functions take. This flexibility is the reason for estimating an ordered choice model rather than a continuous model. We set the location of the knots $\{\overline{q}_j\}_{j=1}^{J-1}$ to the cutoffs separating the deciles of the sample distribution of neighborhood quality.

If we assume D^* is strictly decreasing and households maximize their utility at q^*, then q^* satisfies the following first order condition equivalent to conditions in Equations 2 and 3 in the multilevel case:

$$D^*(q^*) = 0. \tag{20}$$

Given the estimated parameters and functions $\theta \equiv (\widehat{\beta}, \widehat{\gamma}(q), \widehat{C}(q))$ constructed from the discrete choice model parameters $\widehat{\theta}_J$, we use the FOC in Equation 20 to identify unobservables as:[8]

$$\widehat{V} = \widehat{\mu}(X) + Z\widehat{\gamma}(q^*) - \widehat{C}(q^*).$$

For each individual we also estimate

$$\widehat{\pi}_q^0(X) = \Phi[\widehat{\beta}_1 X_1 + \cdots + \widehat{\beta}_m X_m - \widehat{C}(q_j)] \tag{21}$$

$$\widehat{\pi}_q^1(X) = \Phi[\widehat{\beta}_1 X_1 + \cdots + \widehat{\beta}_m X_m - \widehat{C}(q) + \widehat{\gamma}(q)] \tag{22}$$

for several values of q. Note that with the estimated parameters θ we can estimate $\widehat{\pi}_q^Z$ for any q and Z. These define the intervals of u_D at which individuals will select into neighborhoods of a specific quality range, given their observable characteristics and voucher assignment. They also allows us to determine whether an individual is a complier when given a voucher regardless of the margins of choice by which a discrete treatment is defined.

5.3 Ordered Choice Model Estimation Results

As a first check on model fit, Figure 11 shows the unconditional average π_q^Z from the data together with fractional polynomial fitting of $\widehat{\pi}_q^Z$ estimates as functions of $\widehat{\mu}(X)$. We can see that the model fits the data well.

As measured by the estimated $\widehat{\mu}(X)$, the distributions of observable characteristics conditional on voucher status confirm that voucher assignment was successfully randomized (Figure 12a). The distribution of unobservables follows a standard normal distribution very closely (Figure 12b), and the distributions of unobservables conditional on experimental group status look similar (Figure 12c).

Figures 13a and 13b show that MTO had only modest effects on neighborhood quality. These Figures show the joint distribution of $(\widehat{\mu}(X), \widehat{u}_D)$ along with fractional polynomial fittings of $\widehat{\pi}_q^Z$ for several quantiles q as functions of $\widehat{\mu}(X)$. Even for the experimental group, the vast majority of MTO volunteers remained in a neighborhood in the first decile of the national distribution of quality. Even while remembering that our sample does not include those who moved to neighborhoods above the national median, we still see that in our sample (as in the entire MTO population) only a small share of volunteers were induced across the 30th percentile of neighborhood quality.

[8]This is a notable departure from HUV. In the binary case, HUV define the support over which MTEs can be identified as the intersection of the sets of predicted values of the propensity score obtained from the treated and untreated samples. With linear interpolation we obtain estimates of U_D throughout the entire interval $(0, 1)$.

When examining the limited mobility induced by MTO, it is important to remember that the participants in the program were not a random sample from the US population. At baseline, the MTO volunteers were representative of the one percent living in the left tail of the national distribution of neighborhood quality (Figure 8a). Although MTO volunteers tended to live in better neighborhoods by the time of the interim evaluation, it was still the case that less than 15 percent of the control group in our sample lived in neighborhoods of quality above the 10th percentile of the national distribution (Figure 10d). Given these distributions, it seems reasonable to note the mobility that MTO *did* induce. The cutoffs for moving across the 20th and 30th percentiles of the national distribution for the experimental group ($\widehat{\pi}_{20}^{1}(X)$ and $\widehat{\pi}_{30}^{1}(X)$) are at times both below the cutoff for the 10th percentile for the control group ($\widehat{\pi}_{10}^{0}(X)$).

Perhaps the most interesting result from the estimation of the ordered choice model is the strong heterogeneity in the effects of the program as displayed in Figure 14. Those sites with the highest marginal cost of increasing neighborhood quality in the absence of the program (Los Angeles and New York City, see Figure 14a.) were also the sites where the MTO experimental voucher had the largest effect on the marginal benefit of moving (Figure 14c). And not only were effects heterogeneous by site, but also by voucher type, which is important for the design of future housing mobility programs (Galiani et al. (2012)). The standard Section 8 housing voucher was more effective at reducing moving costs than the experimental voucher at very low levels of neighborhood quality, but this relationship reverses somewhere before the fifth percentile of the national distribution of neighborhood quality.

Table 4 reports the estimated coefficients in $\mu(X)$. There are large differences between sites, with Los Angeles and New York City being significantly worse than Baltimore and Chicago. We can also see that all four observables in our model make households more likely to move to a higher quality neighborhood: someone being a victim of a crime, having no teens in the household, having no family in the baseline neighborhood, and living in a higher quality neighborhood at baseline. Of these observables, the effect of baseline neighborhood quality changes the most over its range.

6 Neighborhood Effects Estimates

6.1 Definition of Treatment

We previously estimated the ordered choice model on a discretized neighborhood quality variable with cutoffs determined by the MTO sample distribution. Through interpolation of the estimated choice model parameters, we recovered the continuous components of the latent index. This allows us to estimate the marginal benefit of moving across any margin of quality and thus, we are able to estimate MTEs and LATEs for any discretization of quality.

When estimating parameters we choose to define treatment levels in terms of deciles of the national distribution:

$$D_i = \begin{cases} 1 & \text{if } q_i^* \in [0, 10); \\ \vdots & \vdots \\ 5 & \text{if } q_i^* \in [40, 50). \end{cases}$$

We choose this discretization not because we believe treatment should have an effect when crossing these particular thresholds of neighborhood quality, but because we believe it offers the best balance between theoretical ideal and practical necessity. The model assumes that moves within a given level of treatment will not have effects on outcomes. Even if they do, it is enough to assume that individuals do not select *within* treatment levels based on rich information regarding neighborhood quality.[9] If these assumptions do not hold within entire deciles of quality, the effects from such moves will likely enter the estimation results through the U_j. Theoretically, one way to handle this issue would be to increase the number of bins until moves within a given level do not have effects on outcomes. Another way to handle this problem would be to reformulate the model to accommodate a continuous treatment (Florens et al. (2008)).

Due to the limited mobility induced by MTO, we believe deciles of quality offer the smallest window on which it is feasible to estimate MTEs and LATEs from MTO. Referring back to Figures 13a and 13b, it is not surprising that under this definition of treatment either single-level LATEs or combinations of several level LATEs are imprecisely estimated. This is especially true as one moves to higher treatment levels. As we will see, even this discretization leaves us with undesirably small sample sizes of compliers. As a result, the only MTEs and LATEs we attempt to estimate are $\Delta_{1,2}^{MTE}$ and $\triangle_{1,2}^{LATE}$.

6.2 Parametric Specification of Potential Outcomes and Estimation Algorithm

While semi-parametric LATE estimators are identified only over the set of compliers, we are able to extend the support over which MTEs are estimated by imposing parametric assumptions. Define the intervals $\mathcal{Y}$ and $\mathcal{M}$ such that $E[Y(D=j)|\mu(X), u_D] \subset \mathcal{Y} \subset \mathbb{R}$ for all $j \in \{1, \ldots, J\}$ and $\mu(X) \subset \mathcal{M} \subset \mathbb{R}$. Then as illustrated in Figure 4, we assume $E[Y(D=j)|\mu(X), u_D] = f(\mu(X), u_D)$ defines a surface in the Cartesian product $\mathcal{Y} \times \mathcal{M} \times [0,1]$. We assume this surface is a plane in $\mathcal{Y} \times \mathcal{M} \times [0,1]$ that we parameterize for each j by

$$E[Y(D=j)|\mu(X), u_D] = \beta_0^j + \beta_1^j \mu(X) + \beta_2^j u_D. \tag{23}$$

Figure 15a shows the support of $E[Y(D=j)|\mu(X), u_D]$ in $\mathcal{M} \times [0,1]$ for $D \in \{1,2,3,4\}$. The most important observation from this Figure is so obvious that it may be overlooked; namely, that most individuals select into $D=1$. We can also see that in order to select into a level of treatment higher than 1, an individual must have had a high $\mu(X)$ and low u_D.

Since the focus of this analysis is on differences between $E[Y(D=j)|\mu(X), u_D]$ for $D \in \{1,2\}$,

[9] This can be seen as a stronger version of the central identifying assumption in Bayer et al. (2008).

Figures 15b and 15c show the support of $E[Y(D=j)|\mu(X),u_D]$ in $\mathcal{M}\times[0,1]$ for these levels of treatment. One immediate difference we can see from the previous figure is that the combination of a very high $\mu(X)$ and a very low u_D leads individuals to select into levels of treatment above 1 or 2. Thus this region is not in the support of $E[Y(D=j)|\mu(X),u_D]$ for $D\in\{1,2\}$. The area of overlapping or near-overlapping support for $E[Y(D=1)|\mu(X),u_D]$ and $E[Y(D=2)|\mu(X),u_D]$ plays a key role in our estimation algorithm.

Recall that estimation of the ordered choice model has allowed us to estimate counterfactual propensity scores with and without subsidy, π_j^1, π_j^0, for specific quality transitions $D=j$ to $D=j+1$. These define the intervals of u_D at which individuals will select into neighborhoods of a specific quality level, given their observable characteristics and voucher assignment, and thus, are key in defining the support sets over which we will estimate Equation 23 for each j. Let individuals be defined by their observable and idiosyncratic indices of choice, as well as their voucher assignment: $(\mu(X),u_D,Z)$. The algorithm for estimating MTEs specific to transition 1-2 is described below:

Step 1 Select individuals with u_D's such that, given $\mu(X)$ and Z, they maximize utility at quality level $D=1$:

$$I_1=\left\{(\mu(X),u_D,0)|u_D>\pi_1^0(X)\right\}\ \cup\ \left\{(\mu(X),u_D,1)|u_D>\pi_1^1(X)\right\}.$$

Step 2 Select individuals with u_D's such that, given $\mu(X)$ and Z, they maximize utility at quality level $D=2$:

$$I_2=\left\{(\mu(X),u_D,0)|u_D\in[\pi_2^0(X),\ \pi_1^0(X)]\right\}\ \cup\ \left\{(\mu(X),u_D,1)|u_D\in[\pi_2^1(X),\ \pi_1^1(X)]\right\}.$$

Step 3 Trim and/or expand the support sets to obtain regions TI_1 and TI_2 with the largest region of overlap subject to the constraint that the parametric assumptions from Equation 23 are reasonable for each j. Given the distribution of $\mu(X)$ in our sample (Figure 12a), we drop individuals with $\mu(X)$ above the 95th percentile for both TI_1 and TI_2. Given the selection patterns in our sample (Figure 15), we also impose an upper bound on u_D of 0.4 for both sets. Thus we trim and expand I_1 and I_2 to obtain $TI_1=TI_2=[-0.75,0.45]\times[0,0.4]$.

Step 4 Estimate the expected value of outcome Y_j at neighborhood quality level $D=j$ as a linear function of $\mu(X)$ and u_D over each set TI_j for $j\in\{1,2\}$:

$$E[Y(D=j)|\mu(X),u_D]=\beta_0^j+\beta_1^j\mu(X)+\beta_2^j u_D.$$

Step 5 Determine the area of common support $\mathcal{S}\subset\mathcal{M}\times[0,1]$ over which parametric assumptions will be extended and the MTE function estimated. We set $\mathcal{S}=TI_1=TI_2$. Looking at Figure 15c, in our application we extend the parametric assumptions from Equation 23 for both $j=1$ and $j=2$ over the support set $\mathcal{S}=[-0.75,0.45]\times[0,0.4]$.

Step 6 Use the parameter estimates obtained in Step 4 to estimate the MTE function

$$\Delta_{1,2}^{MTE}(\mu(X), u_D) = (\beta_0^2 - \beta_0^1) + (\beta_1^2 - \beta_1^1)\mu(X) + (\beta_2^2 - \beta_2^1)u_D$$

over $\mathcal{S}$ as:

$$\widehat{\Delta}_{1,2}^{MTE}(\mu(X), u_D) = (\widehat{\beta}_0^2 - \widehat{\beta}_0^1) + (\widehat{\beta}_1^2 - \widehat{\beta}_1^1)\widehat{\mu}(X) + (\widehat{\beta}_2^2 - \widehat{\beta}_2^1)\widehat{u}_D.$$

Step 7 Repeat the following steps T times:

Step a Sample with replacement

Step b Estimate the ordered choice model on the sample

Step c Repeat Steps 1-6 on the sample

Construct standard errors using the T parameter estimates. We set $T = 1,000$.

Evaluating $\Delta_{1,2}^{MTE}(\mu(X), u_D)$ at a fixed value of $\mu(X)$ allows us to trace the MTE over the quantiles of u_D. Likewise, we can see how the MTE varies over the range of $\mu(X)$ for any fixed value of u_D. Note that the MTEs computed via this algorithm apply to more individuals than the compliers, $I_{1,2} = \{(\mu(X), u_D, Z)|u_D \in [\pi_1^0(X),\, \pi_1^1(X)]\}$, over whom we estimate semi-parametric LATEs. Compliers are individuals with values of u_D such that, given $\mu(X)$, they would be induced into a $D = 1$ to $D = 2$ transition by manipulating $Z = 0$ to $Z = 1$ (ie, by voucher assignment). Thus the semi-parametric LATE for transition 1-2 is estimated as the expected difference in outcomes in the experimental and control groups for individuals in $I_{1,2} \subset \mathcal{S}$ (See Appendix B for details.).

Identification of MTE and LATE parameters over the larger set $\mathcal{S}$ is achieved through the parametric assumptions in Equation 23. Looking at Figure 15c, since $\mathcal{S} = [-0.75, 0.45] \times [0, 0.4]$, we can see that identification of $E[Y(D = j)|\mu(X), u_D]$ is based off of functional form assumptions for high u_D when $j = 2$. Similarly, when $j = 1$, identification of $E[Y(D = j)|\mu(X), u_D]$ comes from parametric assumptions for high $\mu(X)$ and low u_D.

6.3 Neighborhood Effects Estimation Results

Parametric MTE estimates are presented in Figures 16 and 17. Focusing first on Figure 16, we can see how MTE estimates vary over observable characteristics at the median u_D for which MTEs were estimated ($u_D = 0.2$). What we see from these MTEs is that even conditioning on unobservables, moving to a higher quality neighborhood had highly heterogeneous effects depending on an individual's observable characteristics as summarized by $\mu(X)$.

Looking at Figure 16a, we can see that the heterogeneity in MTEs across observables is quite pronounced. The MTE on household income at the lowest values of $\mu(X)$ is estimated to be over \$5,000, while the point estimate of the MTE at the highest values of $\mu(X)$ is actually negative. Figure 16d shows similar heterogeneity with respect to employment. As reported in Table 4, the covariates influencing $\mu(X)$ the most are baseline quality and site at which an individual resided. Thus, one way of interpreting the heterogeneity reported in Figure 16 is that labor market outcomes were affected most positively for individuals in Los Angeles and New York City, the two sites with the most mobility induced by MTO.

Figure 17 shows MTE parameter estimates along with 90 percent confidence intervals over the range of u_D at the median value of $\widehat{\mu}(X)$ for which MTEs were estimated, $\widehat{\mu}(X) = -0.12$. We can see that even accounting for observables, effects were strongly heterogeneous by unobservables. We take these results as evidence in favor of selection occurring in a model under essential heterogeneity rather than strong ignorability. Even under the more relaxed essential heterogeneity conditions, variation in conditional effects suggests the difficulty of measuring the effect of neighborhood quality with a one-dimensional index.

After first noticing their strong heterogeneity, three other aspects of the MTE estimates reported in Figures 16 and 17 merit attention. The first notable characteristic of the MTE estimates is their large magnitude. For example, effects on household income represent a large share of the sample average (Table 5). Some effects on adult BMI are extremely large, and effects on employment, labor force participation, and TANF receipt are all also very large over some regions of $\mathcal{S}$.

A second notable characteristic of the MTEs is their imprecision. Few point estimates are statistically significant, even for very large effects. The imprecision of the MTE estimates is hardly a surprise given the limited mobility across neighborhood quality as discuss in Sections 4 and 5. We also believe that the lack of structure in our model could be contributing to this imprecision. It is important to note that the linear assumptions on $E[Y(D = j)|\mu(X), U_D]$ not only help to identify MTEs, but also increase the precision of their estimates. The semi-parametric LATE estimates presented in Appendix B that make no parametric assumptions about $E[Y(D = j)|\mu(X), U_D]$ tend to be more imprecise than our MTE estimates.

A final notable characteristic of the MTE estimates is the precise form taken by the effect heterogeneity. Why does the MTE function exhibit the type of heterogeneity that it does? Thinking of the canonical example of MTEs from educational attainment, one would expect the MTE function to be downward-sloping in U_D, which is an empirical finding from Carneiro et al. (2011). Yet in MTO those individuals most likely to select into treatment were those individuals with the *lowest* labor market returns from doing so. The individuals who benefit most from improving neighborhood quality are those who are most difficult to move to high quality neighborhoods.

Ultimately, our model is limited in its structure to provide an explanation of the heterogeneity of estimates. It is not clear how to interpret MTE heterogeneity since individuals can exhibit low $\mu(X)$'s and high u_D's for a number of reasons. Differences in access to other social programs and opportunities at both the site (captured in $\mu(X)$) or individual (captured in u_D) levels could explain our results. Another explanation for this form of heterogeneity that we find plausible is violence. Effects on outcomes other than safety may only have a limited influence on selection. Finally, information could play a key role in generating this heterogeneity. So even if individuals make their selection of neighborhood quality based on their perceptions of idiosyncratic gains, other factors outside their information set may ultimately influence outcomes. That these factors are not explicitly modeled in our analysis points to the limits of our methodology to fully understand effects from neighborhood quality. We believe future research into the sources of MTE heterogeneity could be fruitful.

7 Conclusion

This paper proposed and implemented a new approach to identifying MTE and LATE parameters that exploits the identification of unobservables in a multi-level model of treatment. We developed this estimator in order to identify effects from neighborhood quality using MTO voucher assignment as an instrumental variable. Neighborhood effects were defined as differences in potential outcomes when residing in neighborhoods of varying quality. We relaxed two key identifying assumptions in the analyses: that MTEs are homogenous over unobservables, and that poverty is a proxy for quality. Neighborhood externalities were captured in the resulting model through an index of neighborhood quality we created. Because overall changes in neighborhood quality induced by the voucher were small, neighborhood effects were imprecisely estimated. However, one could imagine that with other instruments, this approach could more precisely identify neighborhood quality effects under essential heterogeneity.

Estimation results from the ordered choice model indicated that the Moving to Opportunity (MTO) housing mobility experiment had strongly heterogeneous effects across sites. As compared to Baltimore and Chicago, Los Angeles and New York City had higher transition costs to moving to higher quality neighborhoods without a subsidy. The MTO voucher assignment more or less eliminated those differences in cost across sites: Los Angeles and New York City displayed the largest difference in neighborhood quality across control and treated groups by 2002.

Parametric assumptions on expected outcomes allowed us to estimate MTEs over a larger subset of individuals than just the compliers pertaining to LATEs. Nevertheless, the low mobility induced by the experiment still made these parameters pertain to a restricted set subset of individuals, and rendered them imprecisely estimated.

Estimated effects of neighborhood quality on employment and household income were heterogeneous and of large magnitudes. For example, even conditioning on having the median unobservable, the MTE on household income at one extreme of observables was estimated to be over $5,000, while the MTE at the other extreme was actually negative. Effects were similarly heterogeneous by unobservables after accounting for observables.

The heterogeneity we found in MTEs indicates that improving neighborhood quality would have the most positive effects on individuals with either observed or unobserved characteristics that would make them the least likely to select into higher quality neighborhoods without a voucher. We hypothesized that other social programs, violence, and information could all play a role in driving this heterogeneity.

Our model lacked enough structure to make clear conclusions about MTE heterogeneity. However, we can say that heterogeneity over observables suggests that labor market outcomes were affected most positively for individuals at the sites in which larger changes in neighborhood quality were induced by MTO. And we interpreted MTE heterogeneity over unobservables to suggest the need for richer models of how neighborhood characteristics impact outcomes.

References

Aliprantis, D. (2012). Assessing the evidence on neighborhood effects from Moving to Opportunity. *Federal Reserve Bank of Cleveland Working Paper 11-22r*.

Angrist, J. D. and G. W. Imbens (1995). 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.

Bayer, P., S. L. Ross, and G. Topa (2008). Place of work and place of residence: Informal hiring networks and labor market outcomes. *Journal of Political Economy 116*(6), 1150–1196.

Björklund, A. and R. Moffitt (1987). The estimation of wage gains and welfare gains in self-selection models. *The Review of Economics and Statistics 69*(1), pp. 42–49.

Carneiro, P., J. J. Heckman, and E. J. Vytlacil (2011). Estimating marginal returns to education. *American Economic Review 101*(6), 2754–2781.

Clampet-Lundquist, S. and D. S. Massey (2008). Neighborhood effects on economic self-sufficiency: A reconsideration of the Moving to Opportunity experiment. *American Journal of Sociology 114*(1), 107–143.

DeLuca, S. and J. E. Rosenbaum (2003). If low-income blacks are given a chance to live in white neighborhoods, will they stay? Examining mobility patterns in a quasi-experimental program with administrative data. *Housing Policy Debate 14*(3), 305–345.

Florens, J., J. J. Heckman, C. Meghir, and E. Vytlacil (2008). Identification of treatment effects using control functions in models with continuous, endogenous treatment and heterogeneous effects. *Econometrica 76*(5), 1191–1206.

Galiani, S., A. Murphy, and J. Pantano (2012). Estimating neighborhood choice models: Lessons from the Moving to Opportunity experiment. *Mimeo., Washington University in St. Louis*.

Goering, J. (2003). The impacts of new neighborhoods on poor families: Evaluating the policy implications of the Moving to Opportunity demonstration. *Economic Policy Review 9*(2).

Heckman, J. J., S. Urzúa, and E. Vytlacil (2006). Understanding Instrumental Variables in models with essential heterogeneity. *The Review of Economics and Statistics 88*(3), 389–432.

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

Hirano, K. and G. W. Imbens (2005). The propensity score with continuous treatments. In A. Gelman and X.-L. Meng (Eds.), *Applied Bayesian Modeling and Causal Inference from Incomplete-Data Perspectives: An Essential Journey with Donald Rubin's Statistical Family*, pp. 73–84. John Wiley & Sons.

Imai, K. and D. A. van Dyk (2004). Causal inference with general treatment regimes: Generalizing the propensity score. *Journal of the American Statistical Association 99*(467), 854–866.

Imbens, G. W. (2004). Nonparametric estimation of average treatment effects under exogeneity: A review. *The Review of Economics and Statistics 86*(1), pp. 4–29.

Keels, M., G. J. Duncan, S. Deluca, R. Mendenhall, and J. Rosenbaum (2005). Fifteen years later: Can residential mobility programs provide a long-term escape from neighborhood segregation, crime, and poverty? *Demography 42*(1), pp. 51–73.

Kling, J. R., J. B. Liebman, and L. F. Katz (2007). Experimental analysis of neighborhood effects. *Econometrica 75*(1), 83–119.

Ludwig, J. and J. R. Kling (2007). Is crime contagious? *Journal of Law and Economics 50*(3), 491–518.

Ludwig, J., J. B. Liebman, J. R. Kling, G. J. Duncan, L. F. Katz, 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), 144–188.

Mendenhall, R., S. DeLuca, and G. Duncan (2006). Neighborhood resources, racial segregation, and economic mobility: Results from the Gautreaux program. *Social Science Research 35*(4), 892–923.

Minnesota Population Center (2004). *National Historical Geographic Information System* (Pre-release Version 0.1 ed.). Minneapolis, MN: University of Minnesota. http://www.nhgis.org.

Orr, L. L., J. D. Feins, R. Jacob, E. Beecroft, L. Sanbonmatsu, L. F. Katz, J. B. Liebman, and J. R. Kling (2003). *Moving to Opportunity: Interim Impacts Evaluation.* Washington, DC: US Department of Housing and Urban Development, Office of Policy Development and Research.

Polikoff, A. (2006). *Waiting for Gautreaux.* Northwestern University Press.

Rosenbaum, J. E. (1995). Changing the geography of opportunity by expanding residential choice: Lessons from the Gautreaux program. *Housing Policy Debate 6*(1), 231–269.

Rubinowitz, L. S. and J. E. Rosenbaum (2000). *Crossing the Class and Color Lines: From Public Housing to White Suburbia.* University of Chicago Press.

Sampson, R. J. (2008). Moving to inequality: Neighborhood effects and experiments meet social structure. *American Journal of Sociology 114*(1), 189–231.

Sanbonmatsu, L., J. R. Kling, G. J. Duncan, and J. Brooks-Gunn (2006). Neighborhoods and academic achievement: Results from the Moving to Opportunity experiment. *The Journal of Human Resources 41*(4), 649–691.

Votruba, M. E. and J. R. Kling (2009). Effects of neighborhood characteristics on the mortality of black male youth: Evidence from Gautreaux, Chicago. *Social Science & Medicine 68*(5), 814–823.

Vytlacil, E. (2006). Ordered discrete-choice selection models and local average treatment effect assumptions: Equivalence, nonequivalence, and representation results. *The Review of Economics and Statistics 88*(3), 578–581.

Wilson, W. J. (1987). *The Truly Disadvantaged: The Inner City, the Underclass, and Public Policy.* University of Chicago.

8 Appendix A: Identification for Fixed $X = x$

To discuss the identification and estimation of LATEs, we start by deriving an expression for $E[Y|Z=1] - E[Y|Z=0]$ general enough to allow for any ordering relationship between $\boldsymbol{\pi}^0(X)$ and $\boldsymbol{\pi}^1(X)$. Assumption A5 requires that $\pi_j^1(X) \geq \pi_j^0(X)$ for all j and $\pi_j^1(X) > \pi_j^0(X)$ for some j, since the instrument monotonically increases individuals' latent index. And from A4 the variation in average outcomes induced by the instrument can be attributed to changing $\boldsymbol{\pi}^{Z=0}$ to $\boldsymbol{\pi}^{Z=1}$ (from this point forward we keep conditioning on X implicit for the sake of exposition):

$$
\begin{aligned}
E[Y|Z=1] - E[Y|Z=0] &= \sum_{j=1}^{J} \int_0^1 \mathbf{1}\{\pi_j^1 \leq u_D < \pi_{j-1}^1\} E[Y(D=j)|u_D] du_D \qquad (24) \\
&\quad - \sum_{j=1}^{J} \int_0^1 \mathbf{1}\{\pi_j^0 \leq u_D < \pi_{j-1}^0\} E[Y(D=j)|u_D] du_D \\
&= \sum_{j=1}^{J} \Bigg\{ \int_{\pi_j^1}^{\pi_{j-1}^1} E[Y(D=j)|u_D] du_D \qquad (25) \\
&\quad - \int_{\pi_j^1}^{\pi_{j-1}^1} \sum_{k=1}^{J} \Big[\mathbf{1}\{\pi_k^0 \leq u_D < \pi_{k-1}^0\} E[Y(D=k)|u_D] \Big] du_D \Bigg\}.
\end{aligned}
$$

Given equation 6 and since

$$
E[Y(D=j)|u_D] - E[Y(D=j-m)|u_D] = \sum_{k=j}^{j-m+1} \Big\{ E[Y(D=k)|u_D] - E[Y(D=k-1)|u_D] \Big\}
$$

we can rewrite Equation 24 as:

$$
\begin{aligned}
E[Y|Z=1] - E[Y|Z=0] &= \sum_{j=1}^{J-1} \int_{\pi_j^0}^{\pi_j^1} \triangle_{j,j+1}^{MTE}(u_D) du_D \\
&= \sum_{j=1}^{J-1} \Big\{ \triangle_{j,j+1}^{LATE}(\pi_j^0, \pi_j^1) \quad [\pi_j^1 - \pi_j^0] \Big\} \qquad (26)
\end{aligned}
$$

These expressions can be seen in Figures 2-3 for two examples with $J = 3$.

Now suppose we tighten Assumption A6 to:

A6* $\pi_j^1 > \pi_{j-1}^0$ for all $j \in \{1, \ldots, J-1\}$.

Example I in Figure 2 shows such an ordering when $J = 3$. Under A6* the right hand side of

Equation 26 can be derived quickly since

$$\begin{aligned}
E[Y|Z=1] &= \sum_{j=1}^{J} \int_{\pi_j^1}^{\pi_{j-1}^1} E[Y(D=j)|u_D]du_D \\
&= \int_{\pi_1^1}^{1} E[Y(D=1)|u_D]du_D \\
&\quad + \sum_{j=1}^{J-1} \left\{ \int_{\pi_j^0}^{\pi_j^1} E[Y(D=j+1)|u_D]du_D + \int_{\pi_{j+1}^1}^{\pi_j^0} E[Y(D=j+1)|u_D]du_D \right\}
\end{aligned}$$

and

$$\begin{aligned}
E[Y|Z=0] &= \sum_{j=0}^{J} \int_{\pi_j^0}^{\pi_{j-1}^0} E[Y(D=j)|u_D]du_D \\
&= \int_{\pi_1^1}^{1} E[Y(D=1)|u_D]du_D \\
&\quad + \sum_{j=1}^{J-1} \left\{ \int_{\pi_j^0}^{\pi_j^1} E[Y(D=j)|u_D]du_D + \int_{\pi_{j+1}^1}^{\pi_j^0} E[Y(D=j+1)|u_D]du_D \right\}.
\end{aligned}$$

Thus the difference in expected outcomes due to changes in the instrument is the sum of integrated MTEs:

$$\begin{aligned}
E[Y|Z=1] - E[Y|Z=0] &= \sum_{j=1}^{J-1} \left[\int_{\pi_j^0}^{\pi_j^1} E[Y(D=j+1)|u_D]du_D - \int_{\pi_j^0}^{\pi_j^1} E[Y(D=j)|u_D]du_D \right] \\
&= \sum_{j=1}^{J-1} \int_{\pi_j^0}^{\pi_j^1} \triangle_{j,j+1}^{MTE}(u_D)du_D = \sum_{j=1}^{J-1} \left\{ \triangle_{j,j+1}^{LATE}(\pi_j^0, \pi_j^1) \quad [\pi_j^1 - \pi_j^0] \right\}. \qquad (27)
\end{aligned}$$

Since we can recover $\boldsymbol{\pi}^Z = [\pi_1^Z, \pi_2^Z, \ldots, \pi_{J-1}^Z]$ and we can tell in which $[\pi_j^0, \pi_j^1]$ interval u_D lies from the data, we can estimate these LATEs using the semi-parametric estimator discussed in Appendix B. The variation in treatment induced by the instrument identifies:

$$\triangle_{1,2}^{LATE}(\pi_1^0, \pi_1^1) = \frac{\int_{\pi_1^0}^{\pi_1^1} \triangle_{1,2}^{MTE}(u_D)du_D}{\pi_1^1 - \pi_1^0} = E[Y|u_D \in [\pi_1^0, \pi_1^1], Z=1] - E[Y|u_D \in [\pi_1^0, \pi_1^1], Z=0]$$

and

$$\triangle_{2,3}^{LATE}(\pi_2^0, \pi_2^1) = \frac{\int_{\pi_2^0}^{\pi_2^1} \triangle_{2,3}^{MTE}(u_D)du_D}{\pi_2^1 - \pi_2^0} = E[Y|u_D \in [\pi_2^0, \pi_2^1], Z=1] - E[Y|u_D \in [\pi_2^0, \pi_2^1], Z=0].$$

An important issue to remember is that all of the preceding has been implicitly conditioning on observable characteristics X. Figure 4 shows that Example I in Figure 2 is just one cross section taken from an interval of observable characteristics.

Now suppose that we drop A6* and replace it with the less restrictive original Assumption A6. In this case it is possible that $\pi_j^1 > \pi_{j-1}^0$ for some j. Let $u_D \in [\pi_m^1, \pi_{m-1}^1]$ and $u_D \in [\pi_n^0, \pi_{n-1}^0]$ for some $m, n \in \{1, \ldots, J-1\}$ where $m > n$. Then Equation 26 implies

$$E[Y|u_D, Z=1] - E[Y|u_D, Z=0] = \sum_{j=n}^{m-1} \triangle_{j,j+1}^{MTE}(u_D).$$

Thus if $a = \max\{\pi_m^1, \pi_n^0\}$ and $b = \min\{\pi_{m-1}^1, \pi_{n-1}^0\}$, we can identify:

$$E[Y|u_D \in [a,b], Z=1] - E[Y|u_D \in [a,b], Z=0] = \frac{\int_a^b \sum_{j=n}^{m-1} \triangle_{j,j+1}^{MTE}(u_D) du_D}{b-a} \tag{28}$$

This scenario highlights that the precise LATEs identified will be determined by the exogenous variation in the choice probabilities $\boldsymbol{\pi}^{\boldsymbol{Z}}$ induced by the instrument. Note that if $\pi_j^1 > \pi_{j-1}^0$ for some j, the corresponding LATE parameter is still separately identified over the interval

$$\Big[\max\{\pi_j^0, \pi_{j+1}^1\}, \min\{\pi_{j-1}^0, \pi_j^1\}\Big].$$

But if $\max\{\pi_j^0, \pi_{j+1}^1\} \neq \pi_j^0$ or $\min\{\pi_{j-1}^0, \pi_j^1\} \neq \pi_j^1$, then the level-specific LATE parameters will not be separately identified over the entire interval $[\pi_j^0, \pi_j^1]$.

Comparing Example I and Example II in Figure 3 helps to illustrate how the ordering of the π_j^Z determines identification. Since $\pi_2^1 > \pi_0^1$ in Example II, the instrument identifies

$$\triangle_{1,2}^{LATE}(\pi_2^1, \pi_1^1) = \frac{\int_{\pi_2^1}^{\pi_1^1} \triangle_{1,2}^{MTE}(u_D) du_D}{\pi_1^1 - \pi_2^1} = E[Y|u_D \in [\pi_2^1, \pi_1^1], Z=1] - E[Y|u_D \in [\pi_2^1, \pi_1^1], Z=0]$$

However, over the interval $[\pi_1^0, \pi_2^1] = [\pi_1^0, \min\{\pi_2^1, \pi_1^1\}]$ we cannot separately identify each level-specific LATE. Instead the instrument identifies:

$$\begin{aligned}
\triangle_{1,3}^{LATE}(\pi_1^0, \pi_2^1) &= \triangle_{1,2}^{LATE}(\pi_1^0, \pi_2^1) + \triangle_{2,3}^{LATE}(\pi_1^0, \pi_2^1) \\
&= E[Y|u_D \in [\pi_1^0, \pi_2^1], Z=1] - E[Y|u_D \in [\pi_1^0, \pi_2^1], Z=0] \\
&= \frac{\int_{\pi_1^0}^{\pi_2^1} \left(\triangle_{1,2}^{MTE}(u_D) + \triangle_{2,3}^{MTE}(u_D)\right) du_D}{\pi_2^1 - \pi_1^0}.
\end{aligned}$$

But over the interval $[\pi_2^0, \pi_1^0] = [\pi_2^0, \min\{\pi_1^0, \pi_2^1\}]$ the instrument does again separately identify the LATE parameter:

$$\triangle_{2,3}^{LATE}(\pi_2^0, \pi_1^0) = E[Y|u_D \in [\pi_2^0, \pi_1^0], Z=1] - E[Y|u_D \in [\pi_2^0, \pi_1^0], Z=0] = \frac{\int_{\pi_2^0}^{\pi_1^0} \triangle_{2,3}^{MTE}(u_D) du_D}{\pi_1^0 - \pi_2^0}.$$

Figure 19 shows these LATEs graphically.

9 Appendix B: Semi-Parametric LATEs

9.1 Semi-Parametric LATE Estimator

One approach to estimating LATEs is to place no parametric assumptions on the functions $E[Y(D=j)|\mu(X), U_D]$. To justify such a semi-parametric LATE estimator, we keep conditioning on $\mu(X)$ implicit for the sake of exposition. We assume $\pi_j^0 < \pi_j^1 < \pi_{j-1}^0$ and define $U_{Dj}^* \equiv \{U_D | U_D \in [\pi_j^0, \pi_j^1]\}$ so that $U_{Dj}^* \sim U[\pi_j^0, \pi_j^1]$. By definition

$$\triangle_{j,j+1}^{MTE}(u_D) \equiv E[Y_{j+1} - Y_j | U_D = u_D], \quad \text{and}$$
$$\triangle_{j,j+1}^{LATE}(\pi_j^0, \pi_j^1) \equiv \frac{1}{\pi_j^1 - \pi_j^0} \int_{\pi_j^0}^{\pi_j^1} \triangle_{j,j+1}^{MTE}(u_D) du_D.$$

Thus

$$\begin{aligned}
\triangle_{j,j+1}^{LATE}(\pi_j^0, \pi_j^1) &= \frac{1}{\pi_j^1 - \pi_j^0} \int_{\pi_j^0}^{\pi_j^1} \triangle_{j,j+1}^{MTE}(u_D) du_D \\
&= \frac{1}{\pi_j^1 - \pi_j^0} \int_{\pi_j^0}^{\pi_j^1} \{E[Y(D=j+1)|u_D] - E[Y(D=j)|u_D]\} du_D \\
&= E_{U_{Dj}^*}\{E[Y(D=j+1)|u_{Dj}^*] - E[Y(D=j)|u_{Dj}^*]\} \\
&= E_{U_{Dj}^*}\{E[Y(D=j+1)|u_{Dj}^*]\} - E_{U_{Dj}^*}\{E[Y(D=j)|u_{Dj}^*]\} \\
&= E_{U_{Dj}^*}\{E[Y(D=j+1)|u_{Dj}^*, Z=1]\} - E_{U_{Dj}^*}\{E[Y(D=j)|u_{Dj}^*, Z=0]\}, \qquad (29)
\end{aligned}$$

where 29 follows from random assignment since $Z \perp\!\!\!\perp (U_1, \ldots, U_J, U_D)$.

The sample analogue to 29 is

$$\begin{aligned}
\widehat{\triangle}_{j,j+1}^{LATE}(\pi_j^0, \pi_j^1) = &\frac{\sum_{i=1}^N \mathbf{1}\{\widehat{u}_{Di} \in [\widehat{\pi}_{ji}^0, \widehat{\pi}_{ji}^1]\}}{\sum_{i=1}^N \mathbf{1}\{\widehat{u}_{Di} \in [\widehat{\pi}_{ji}^0, \widehat{\pi}_{ji}^1]\}} \left\{ \frac{\sum_{j=1}^N Z_j \mathbf{1}\{\widehat{u}_{Dj} = \widehat{u}_{Di}\} Y_j}{\sum_{j=1}^N Z_j \mathbf{1}\{\widehat{u}_{Dj} = \widehat{u}_{Di}\}} \right\} \\
&- \frac{\sum_{i=1}^N \mathbf{1}\{\widehat{u}_{Di} \in [\widehat{\pi}_{ji}^0, \widehat{\pi}_{ji}^1]\}}{\sum_{i=1}^N \mathbf{1}\{\widehat{u}_{Di} \in [\widehat{\pi}_{ji}^0, \widehat{\pi}_{ji}^1]\}} \left\{ \frac{\sum_{j=1}^N (1-Z_j) \mathbf{1}\{\widehat{u}_{Dj} = \widehat{u}_{Di}\} Y_j}{\sum_{j=1}^N (1-Z_j) \mathbf{1}\{\widehat{u}_{Dj} = \widehat{u}_{Di}\}} \right\}.
\end{aligned}$$

Assuming $\widehat{u}_{Dj} \neq \widehat{u}_{Di}$ for all $i \neq j$ in our sample, it follows that:

$$\widehat{\triangle}_{j,j+1}^{LATE}(\pi_j^0, \pi_j^1) = \frac{\sum_{i=1}^N \mathbf{1}\{\widehat{u}_{Di} \in [\widehat{\pi}_{ji}^0, \widehat{\pi}_{ji}^1]\}\ Z_i Y_i}{\sum_{i=1}^N \mathbf{1}\{\widehat{u}_{Di} \in [\widehat{\pi}_{ji}^0, \widehat{\pi}_{ji}^1]\}\ Z_i} - \frac{\sum_{i=1}^N \mathbf{1}\{\widehat{u}_{Di} \in [\widehat{\pi}_{ji}^0, \widehat{\pi}_{ji}^1]\}\ (1-Z_i) Y_i}{\sum_{i=1}^N \mathbf{1}\{\widehat{u}_{Di} \in [\widehat{\pi}_{ji}^0, \widehat{\pi}_{ji}^1]\}\ (1-Z_i)}. \qquad (30)$$

Given estimates $\widehat{\boldsymbol{\pi}}$ and $\widehat{u}_D$ from the ordered choice model we can construct empirical estimates of LATEs without placing any parametric assumptions on $E[Y(D=j)|\mu(X), U_D]$ using the estimator in Equation 30 when $\pi_j^0 < \pi_j^1 < \pi_{j-1}^0$. As discussed in Appendix A, the precise LATEs identified will be determined by the exogenous variation in the choice probabilities $\boldsymbol{\pi}^Z$ induced by the instrument.

9.2 Semi-Parametric LATE Estimation Results

We also present semi-parametric LATE estimation results in Table 5. We do not analyze these results in great depth due to their imprecision, which is itself due to the small sample of compliers ($n \approx 180$). Table 5 shows a semi-parametric estimator for $\triangle_{1,2}^{LATE}$ derived in the Appendix for several adult outcomes that are averaged over all observables X together with estimates of the average interval $(\max\{\widehat{\pi}_2^1, \widehat{\pi}_1^0\}, \widehat{\pi}_1^1)$ to which the LATE pertains. Estimates of the program effects $\triangle^{ITT}$ on the estimation sample are displayed together with these neighborhood effects to give some context to these results. The most notable feature of these estimates is that they have very large standard errors (computed using 1,000 bootstrap replications). Another feature of these estimates is that the neighborhood effects pertain on average to a very small subset of the sample, with this 11.2 percent being located among those households with low unobservables U_D.

Figures

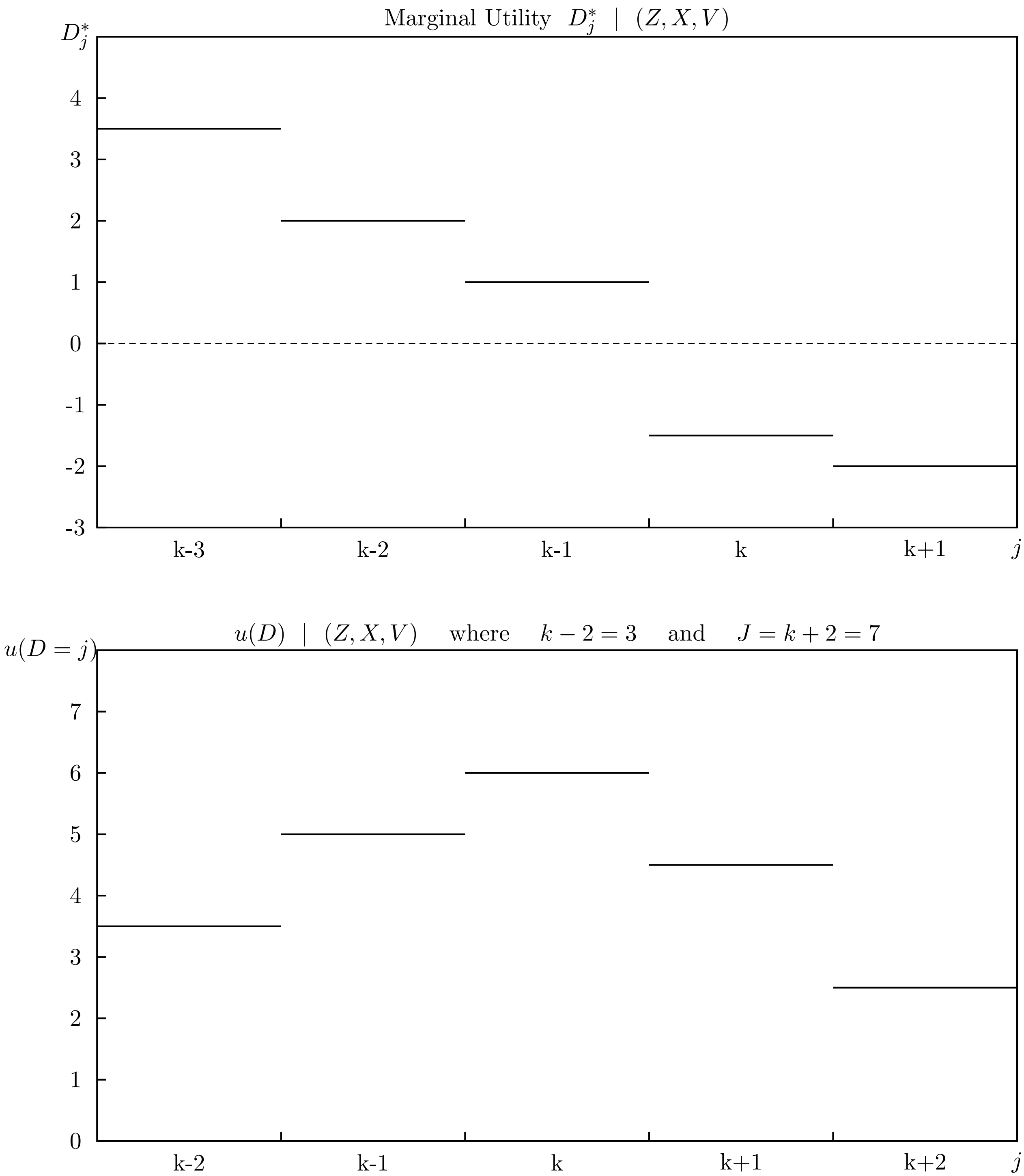


Figure 1: The First Stage Ordered Choice Model

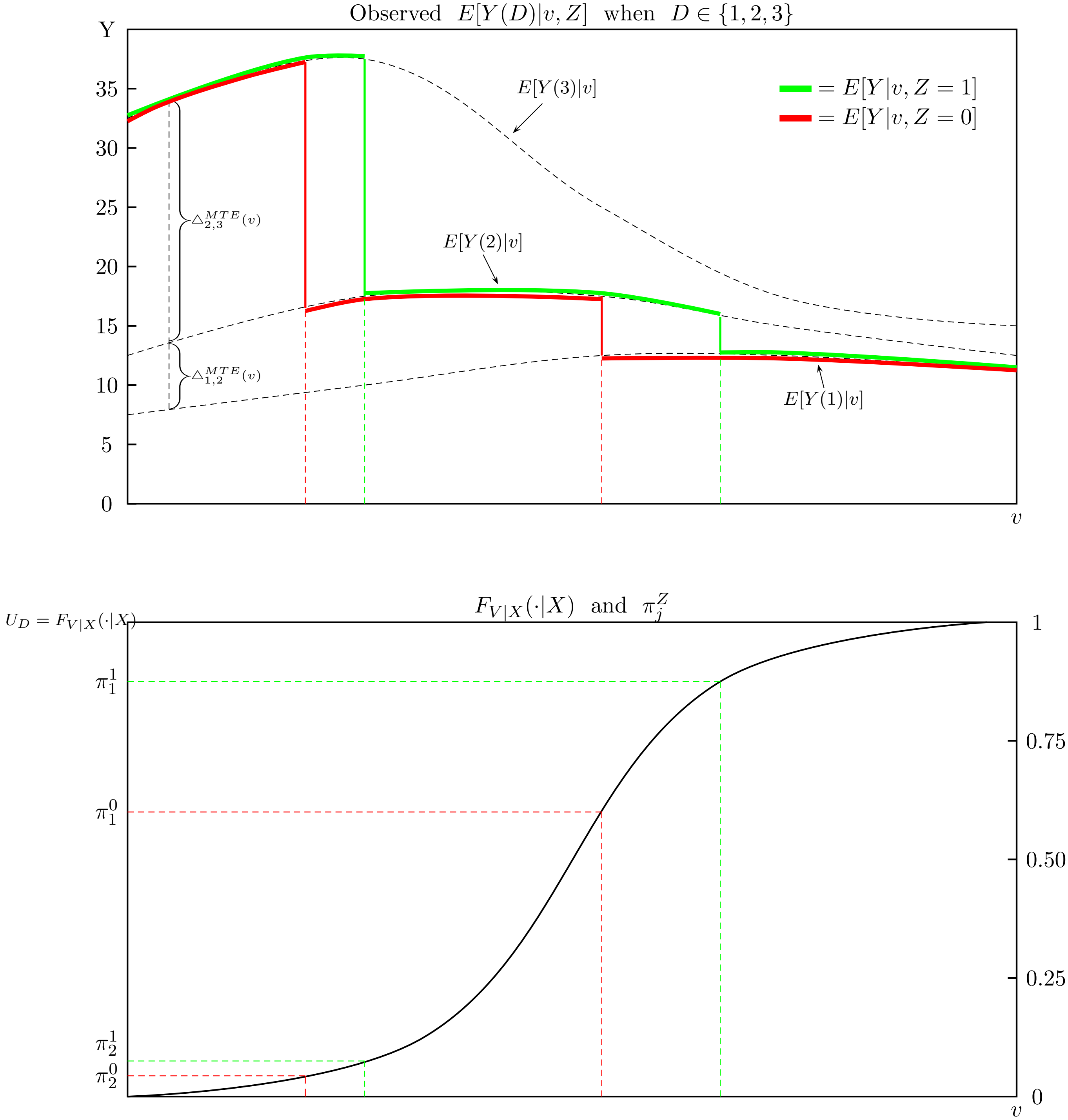
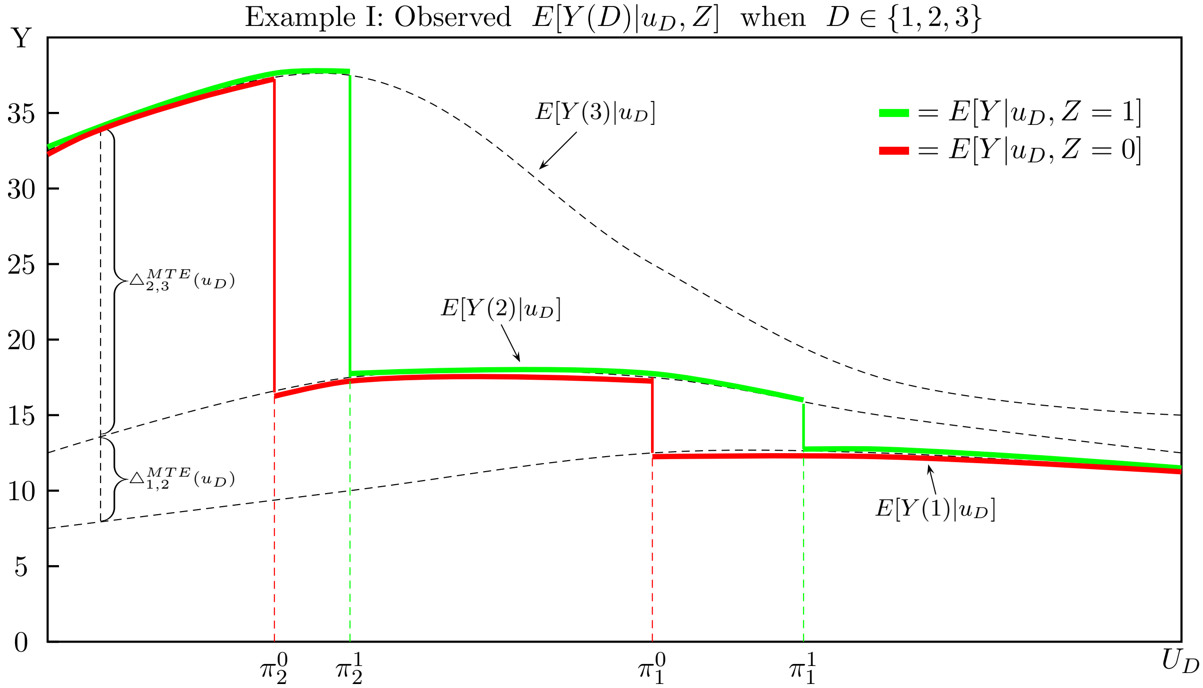


Figure 2: Example I: Potential Outcomes and Marginal Treatment Effects Given $\mu(X)$

Example I: Observed $E[Y(D)|u_D, Z]$ when $D \in \{1, 2, 3\}$



Example II: Observed $E[Y(D)|u_D, Z]$ when $D \in \{1, 2, 3\}$

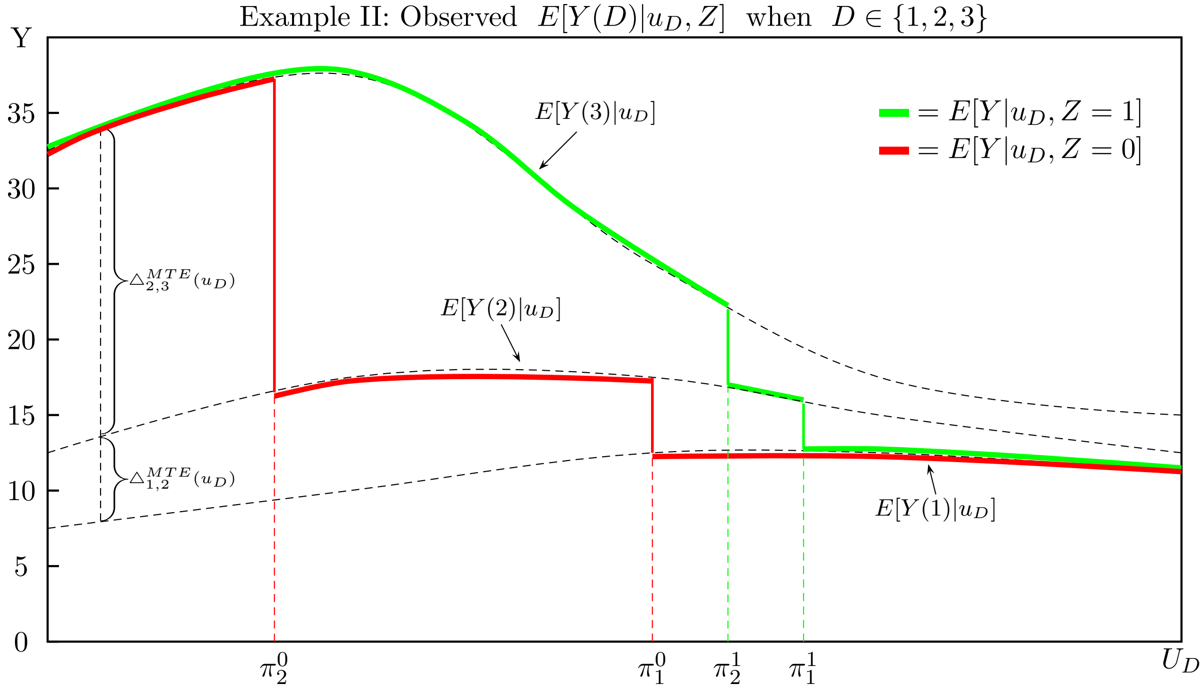


Figure 3: Examples I and II: Potential Outcomes and Marginal Treatment Effects Given $\mu(X)$

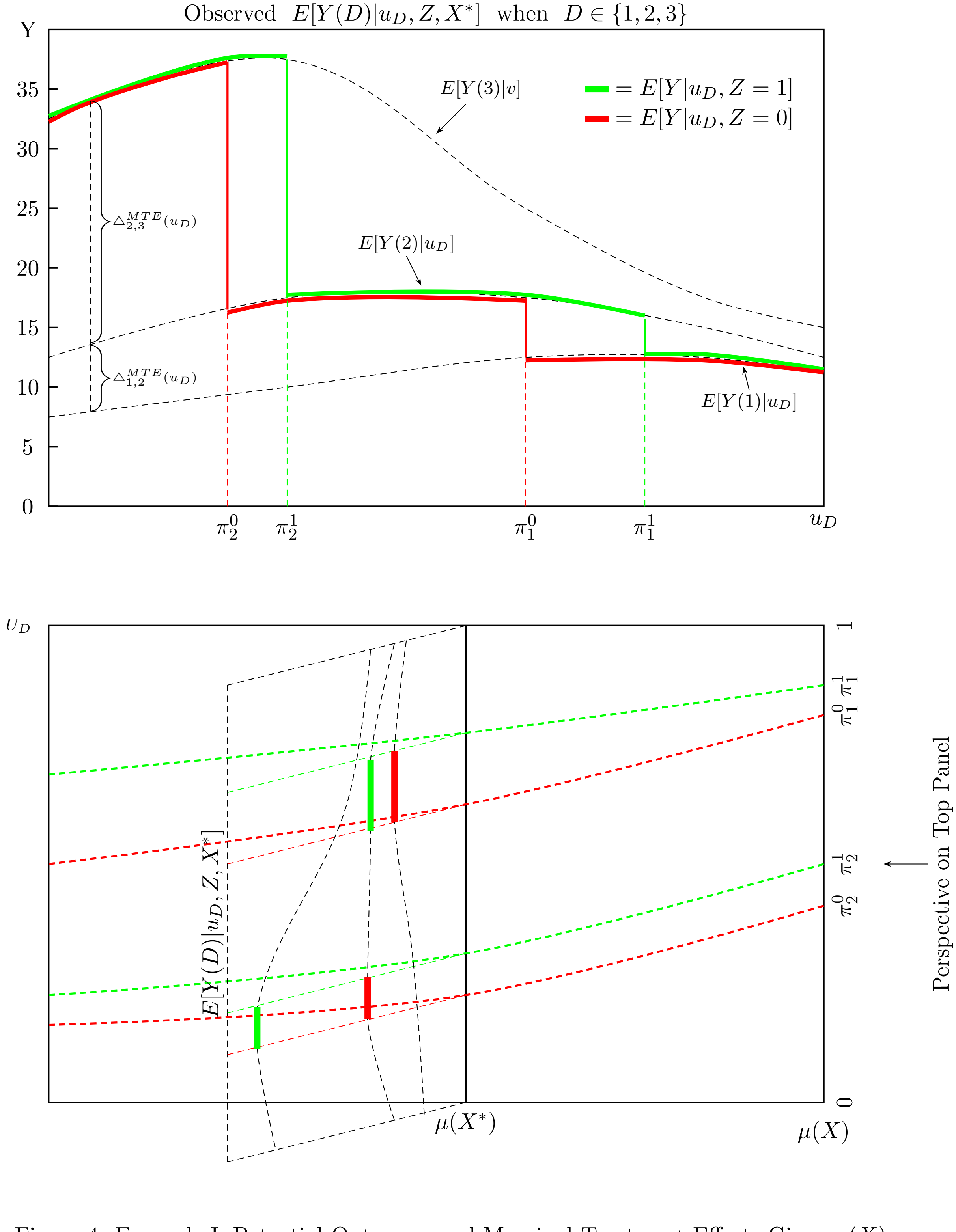
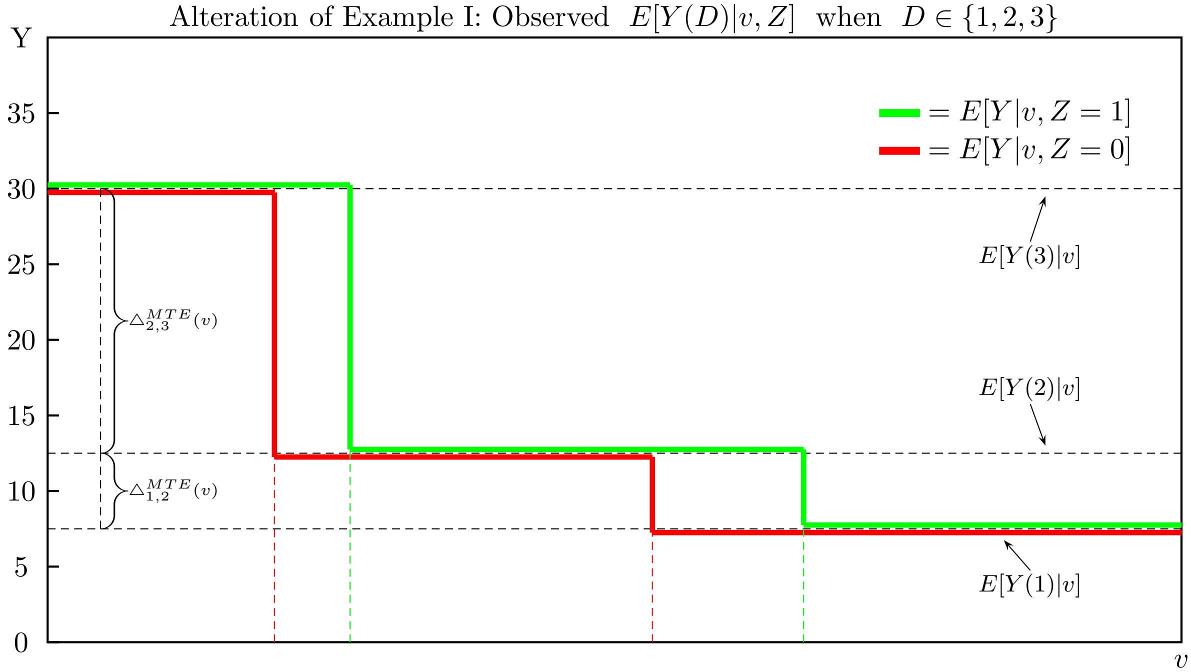


Figure 4: Example I: Potential Outcomes and Marginal Treatment Effects Given $\mu(X)$



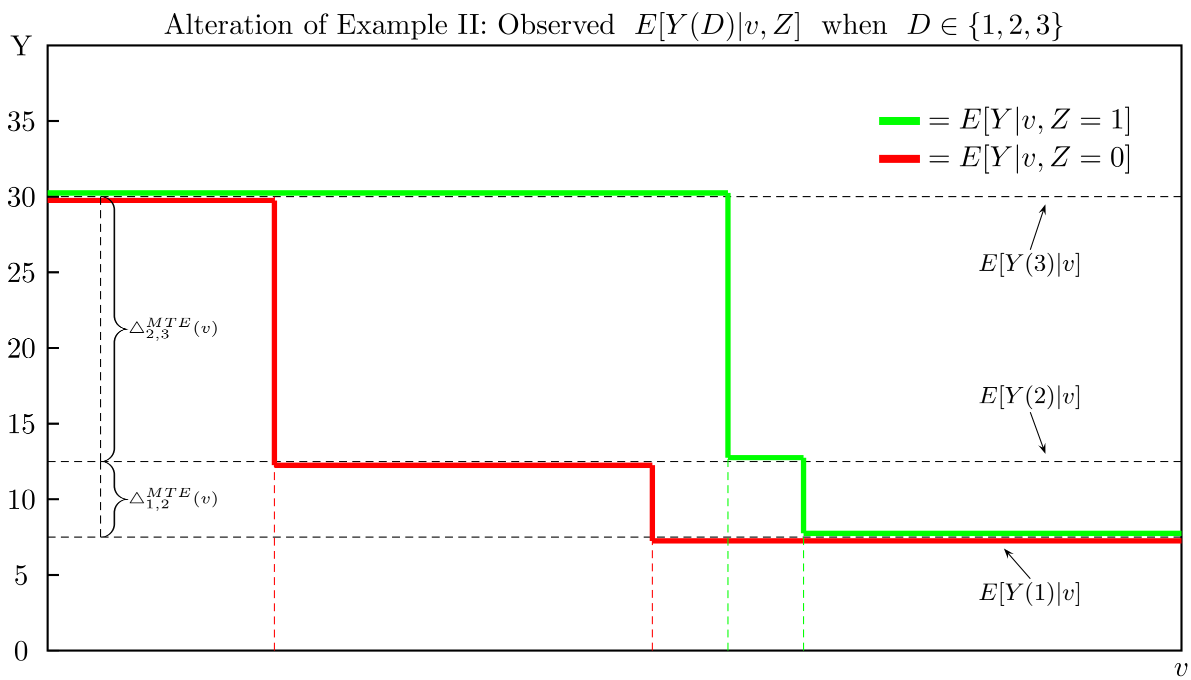


Figure 5: Altering Examples I and II to Satisfy Strong Ignorability

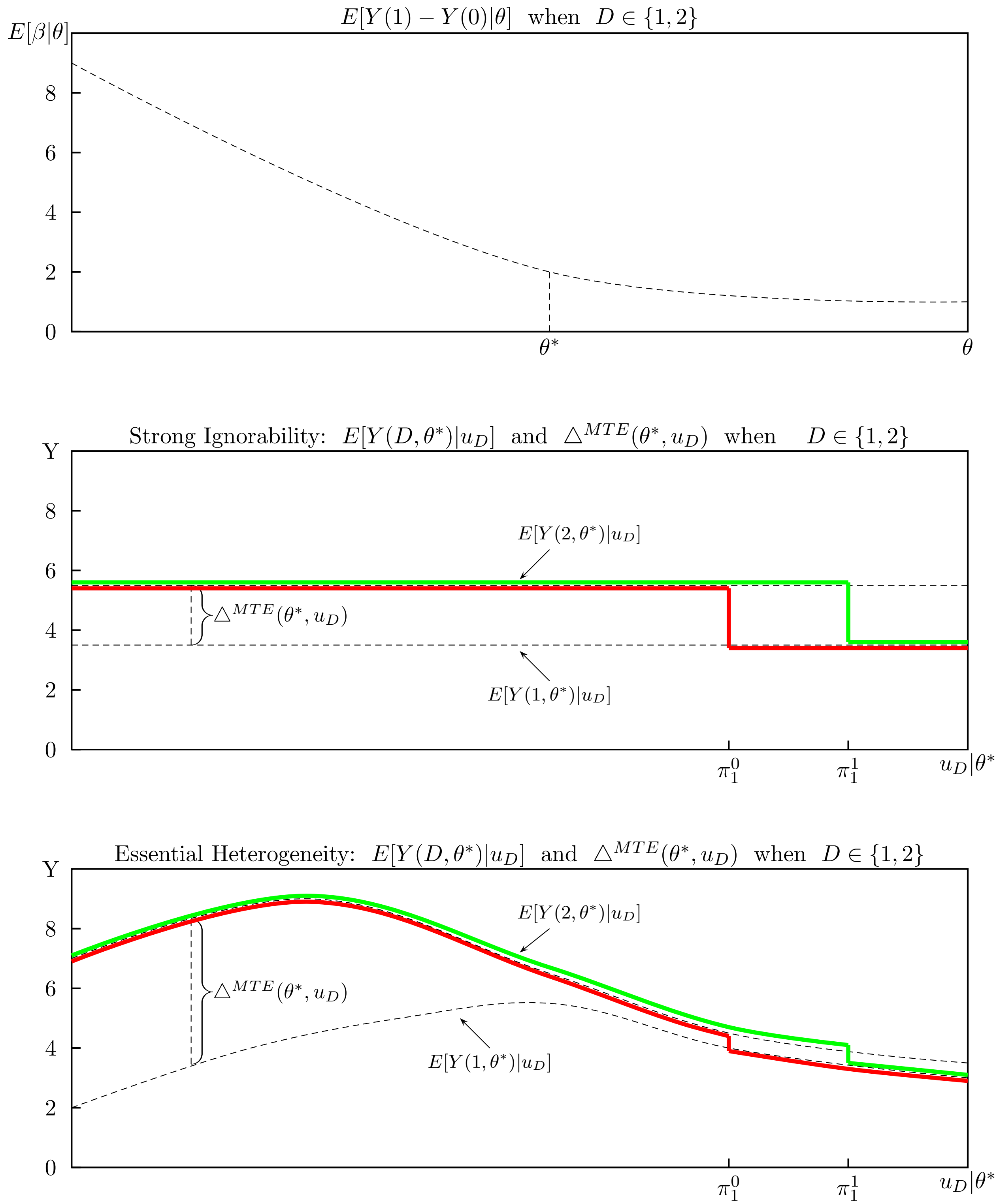


Figure 6: Binary Example with and without Strong Ignorability

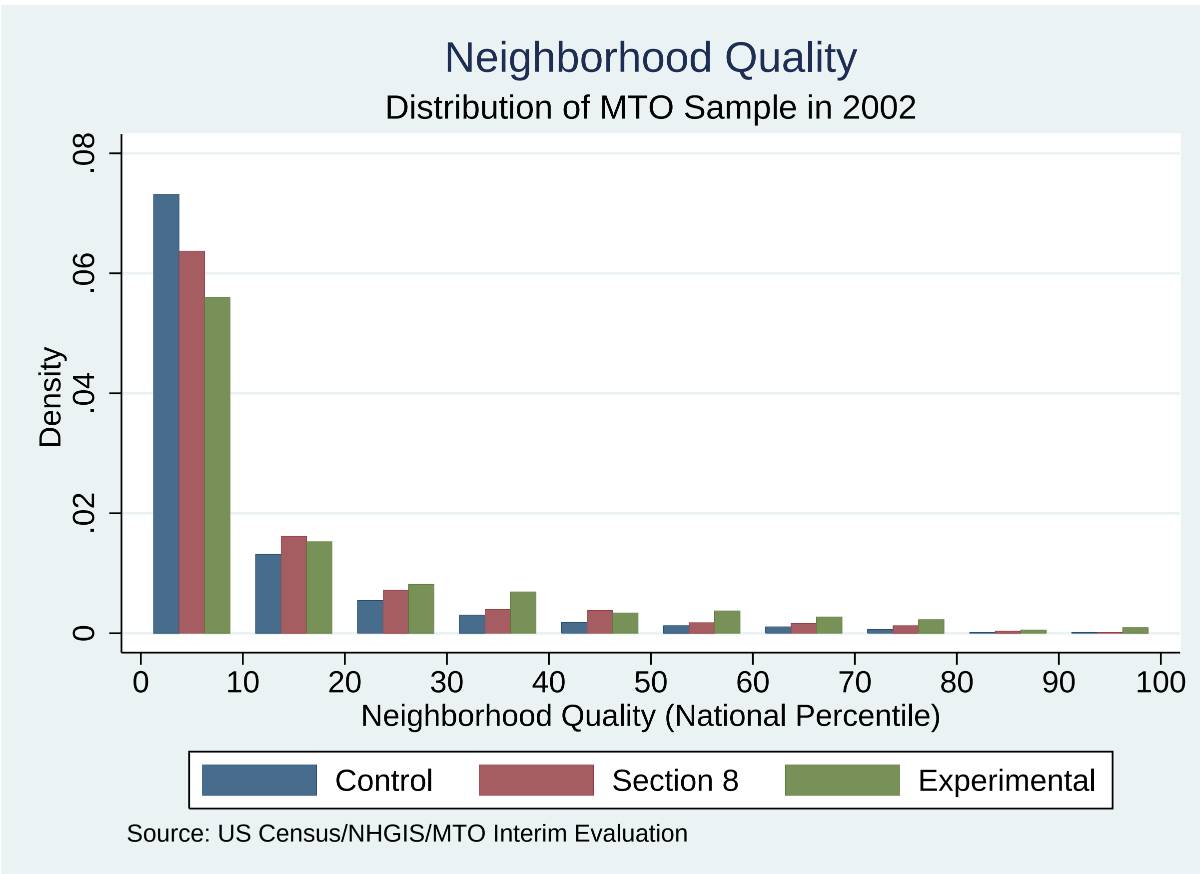
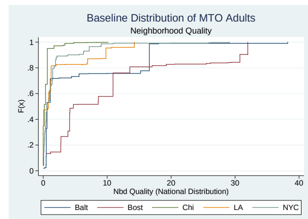
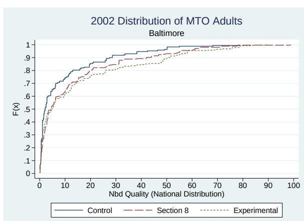


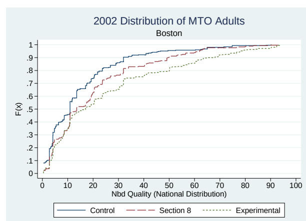
Figure 7: Neighborhood Quality of MTO Adults



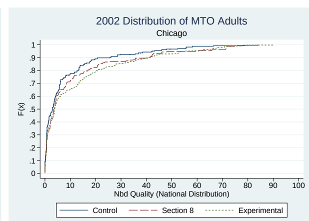
(a) Baseline Quality



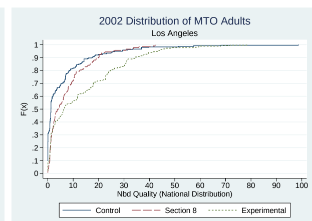
(b) Baltimore



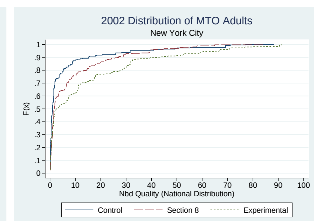
(c) Boston



(d) Chicago

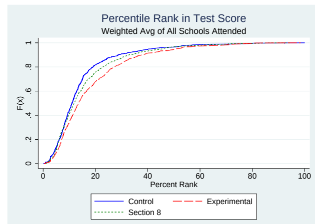


(e) LA

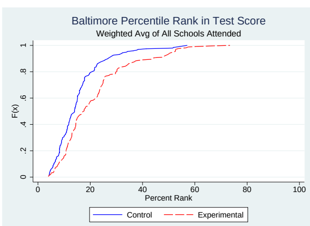


(f) NYC

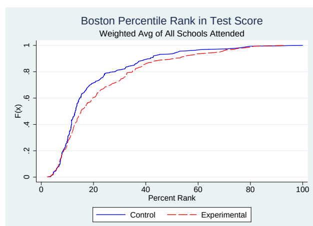
Figure 8: Neighborhood Quality by Site



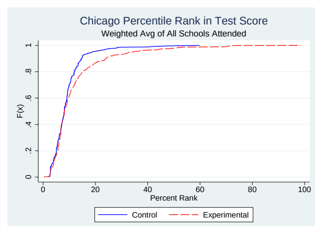
(a) All Sites Pooled



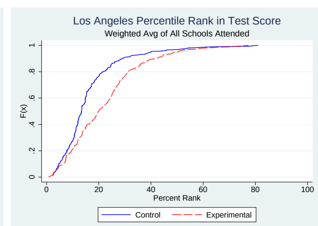
(b) Baltimore



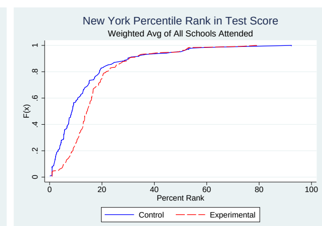
(c) Boston



(d) Chicago

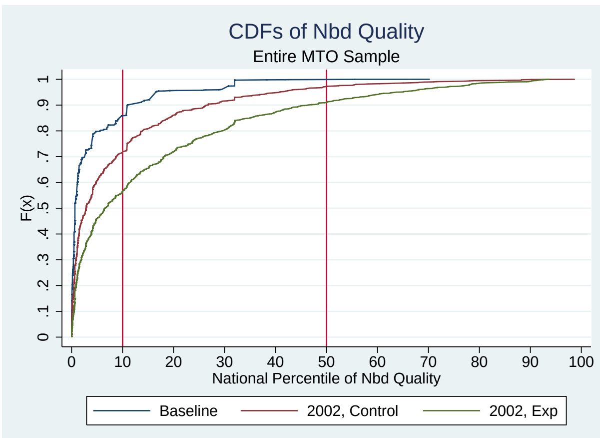


(e) Los Angeles

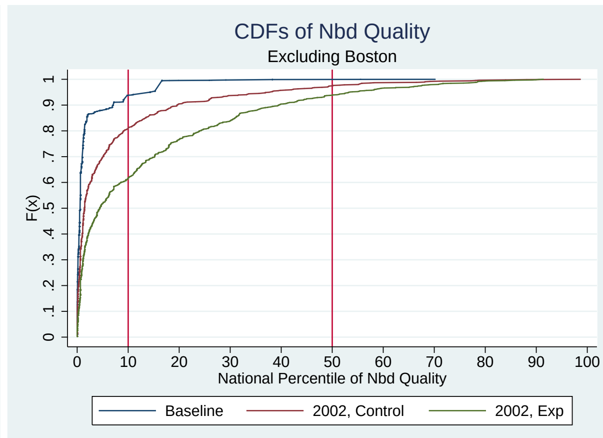


(f) New York City

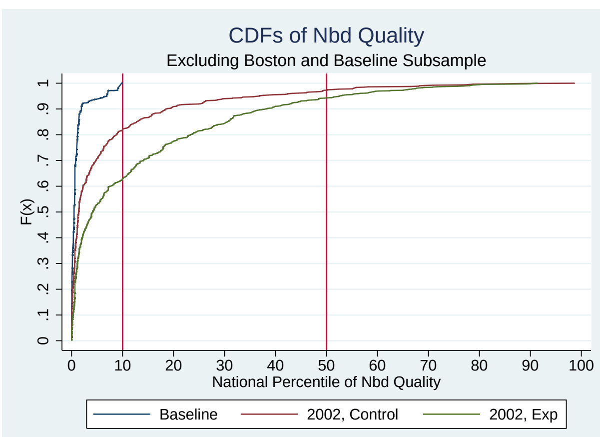
Figure 9: School Ranking on State Tests, Weighted Average Percentile over all Schools Attended (by Site)



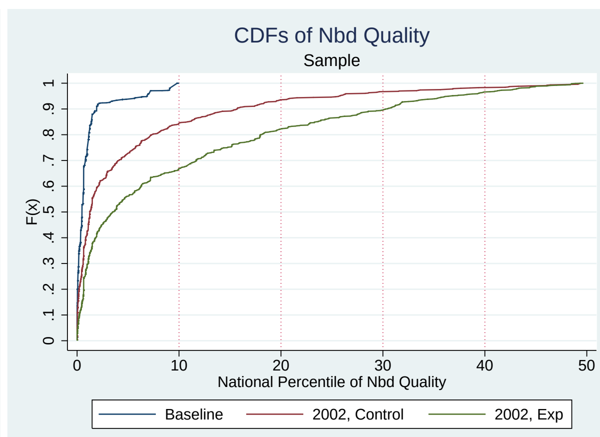
(a) MTO Distributions



(b) Distributions Excluding Boston

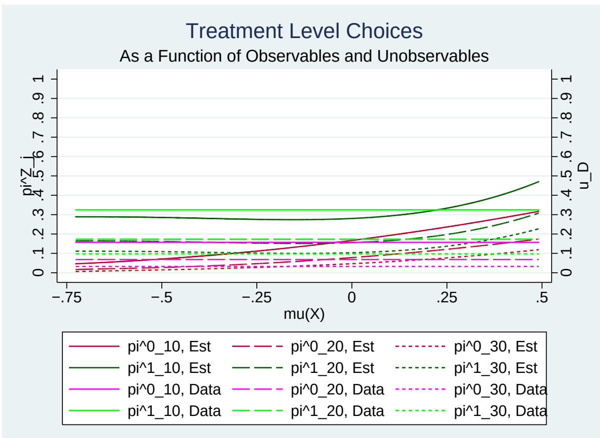


(c) Distributions Excluding Boston and Baltimore Subsample

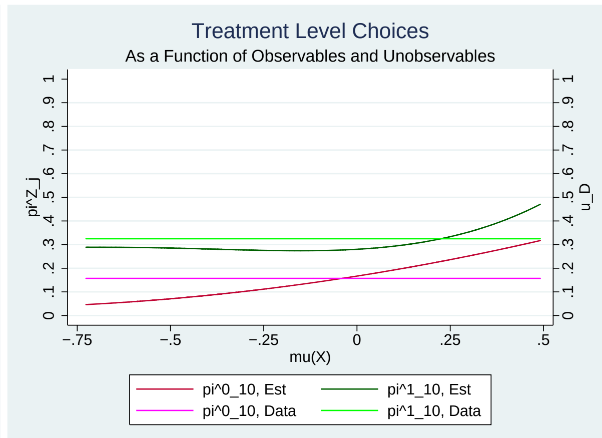


(d) Sample Distributions

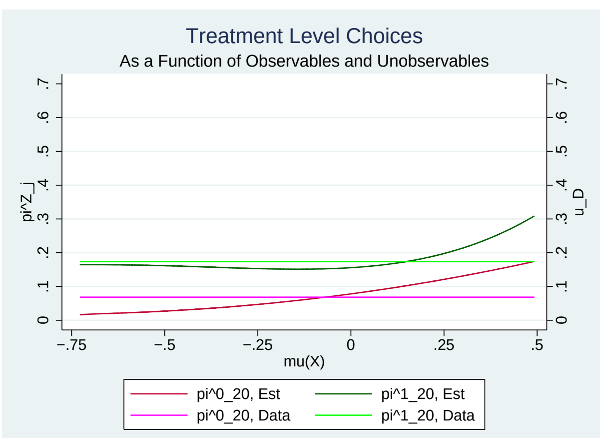
Figure 10: Neighborhood Quality



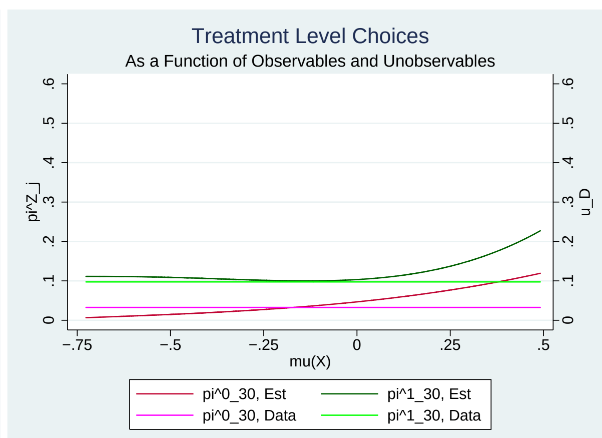
(a) $\widehat{\pi^Z}(X)$ Estimated from the Model and Unconditional π^Z from the Data



(b) $\widehat{\pi}_1^Z(X)$ and π_1^Z

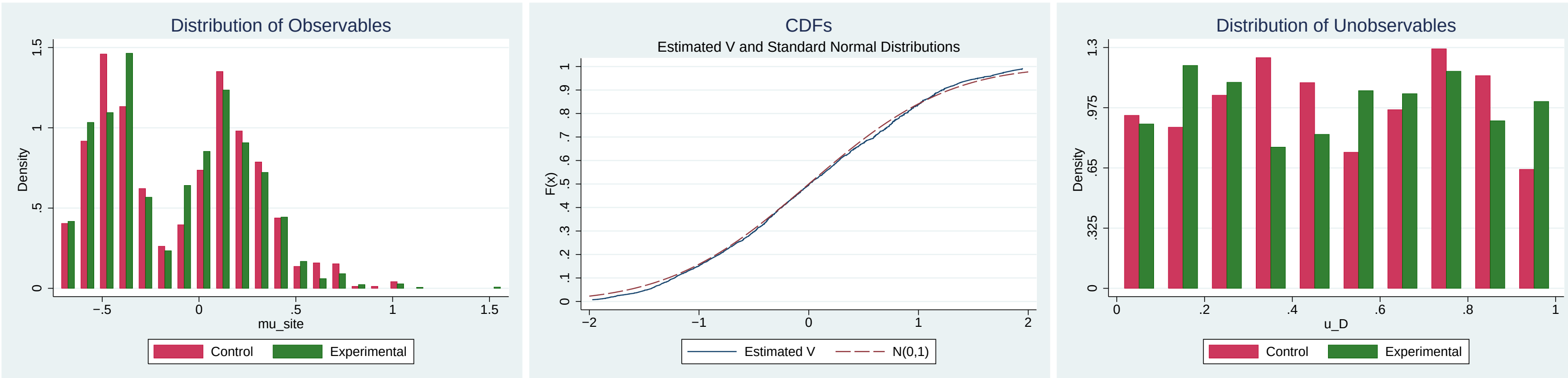


(c) $\widehat{\pi}_2^Z(X)$ and π_2^Z



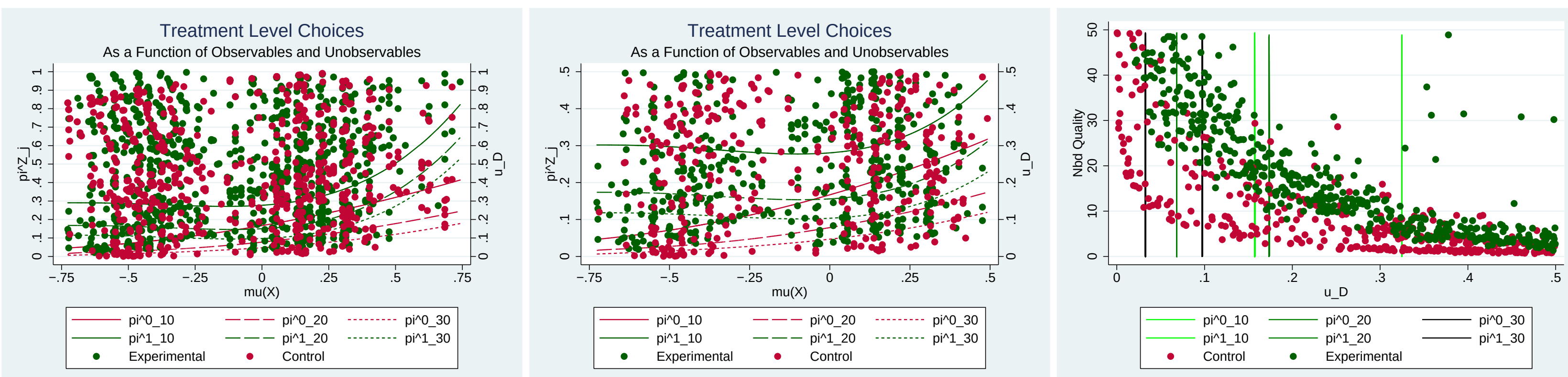
(d) $\widehat{\pi}_3^Z(X)$ and π_3^Z

Figure 11: Model Fit: $\widehat{\pi}_q^Z(X)$ Estimated from the Model and Average π_q^Z from the Data



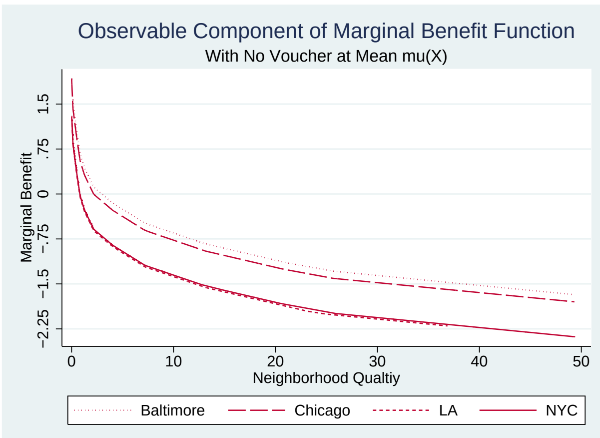
(a) Observables (b) $\widehat{F}_V$ and Φ (c) $\widehat{U}_D|Z$

Figure 12: The Distributions of Observables and Unobservables

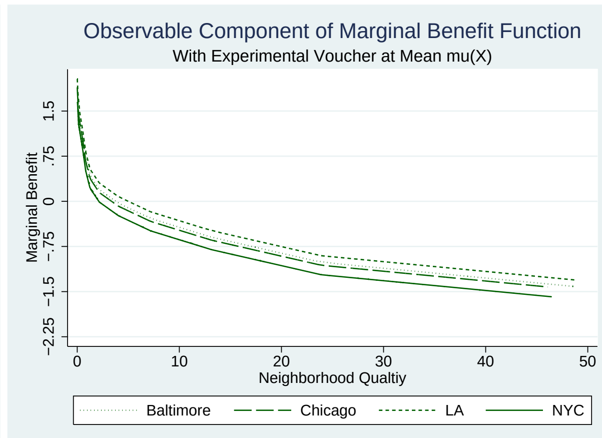


(a) Joint Distribution of $(\widehat{\mu}(X), \widehat{u}_D)$ with $\widehat{\pi}_q^Z$ (b) Joint Distribution of $(\widehat{\mu}(X), \widehat{u}_D)$ with $\widehat{\pi}_q^Z$ (c) $\widehat{u}_D$ and Quality

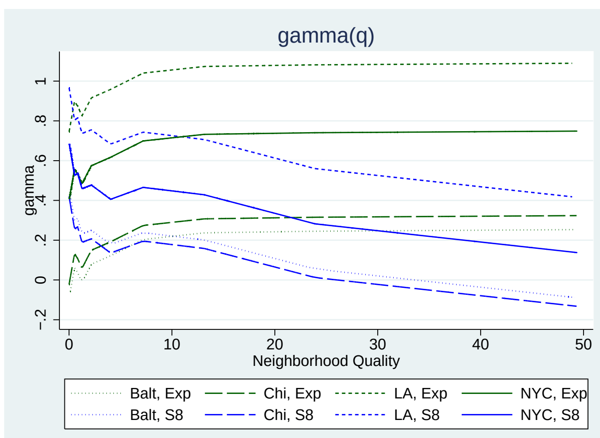
Figure 13: Observables, Unobservables, and Treatment Level



(a) $\widehat{\mu}(X) - \widehat{C}(q)$ by Site

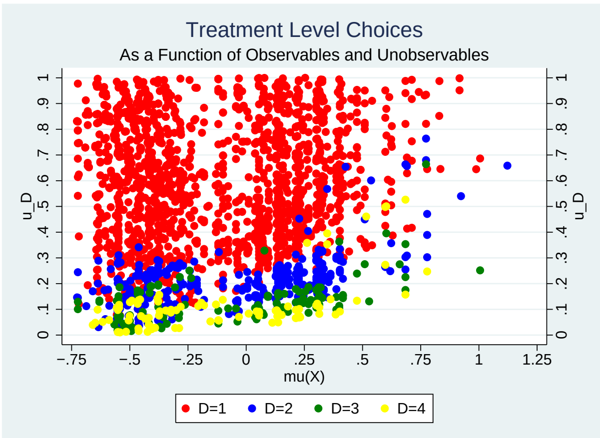


(b) $\widehat{\mu}(X) - \widehat{C}(q) + \widehat{\gamma}(q)$ by Site

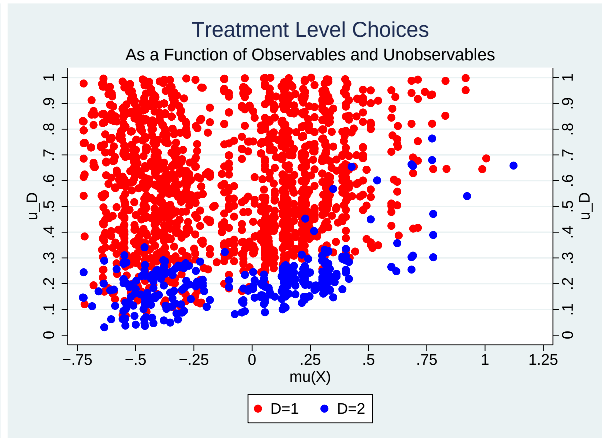


(c) $\widehat{\gamma}(q)$ by Site

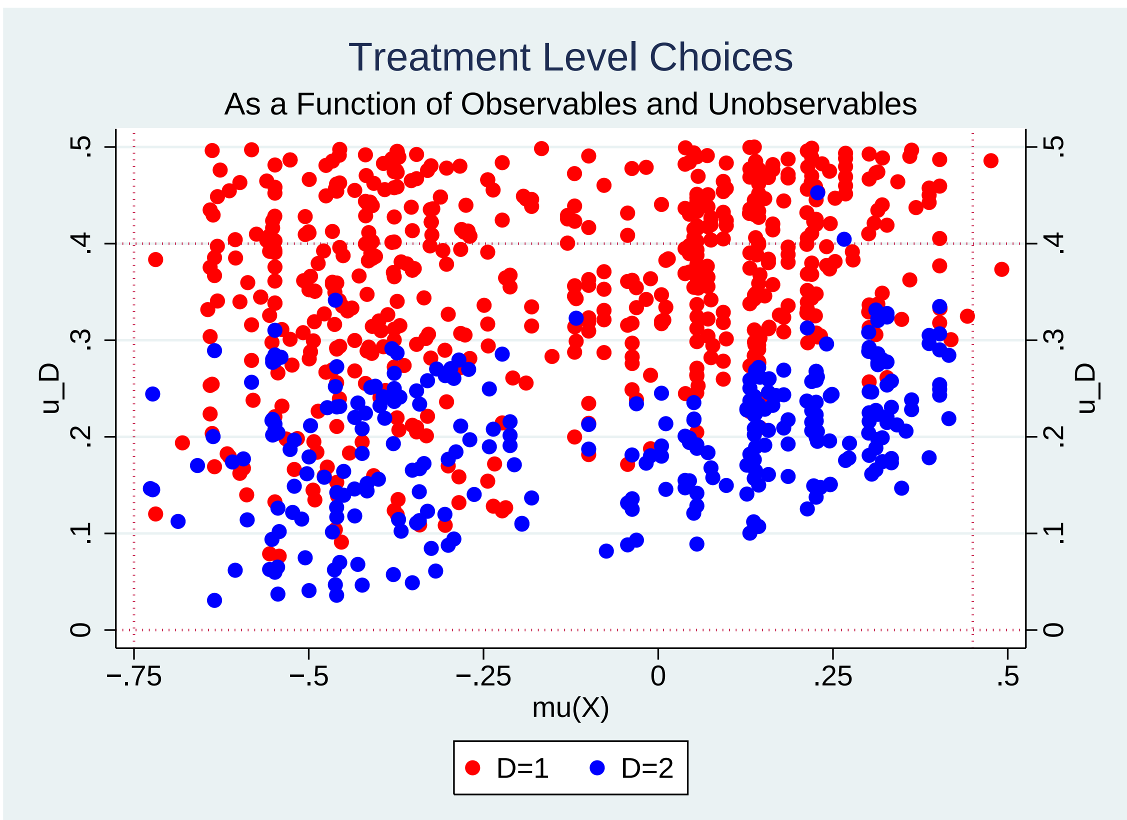
Figure 14: Heterogeneity in Effects of MTO Voucher on Neighborhoods Quality



(a) Common Support

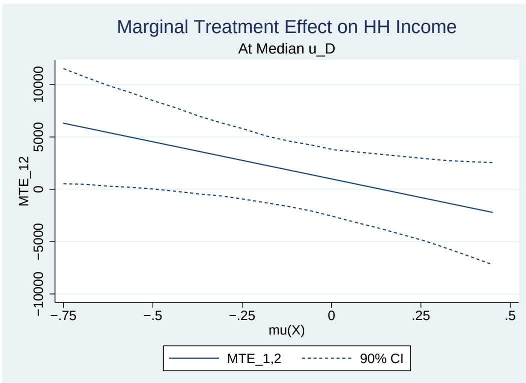


(b) Common Support

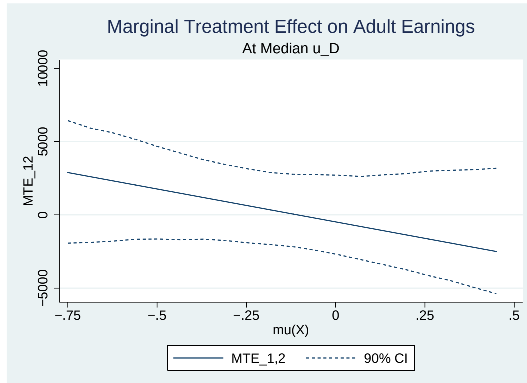


(c) Common Support $\mathcal{S}$

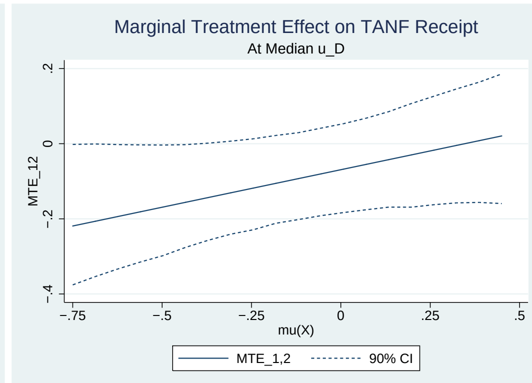
Figure 15: Area of Common Support $\mathcal{S} \subset \mathcal{M} \times [0, 1]$



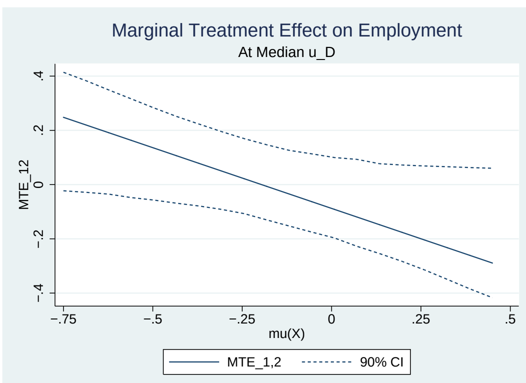
(a) HH Income



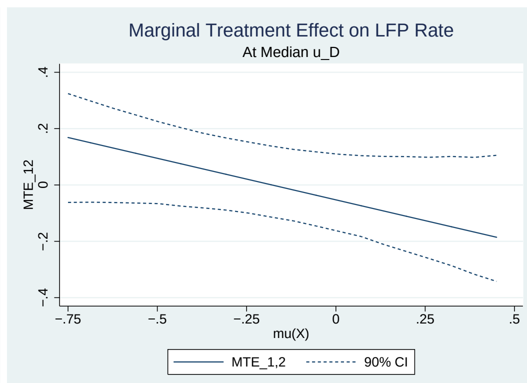
(b) Earnings



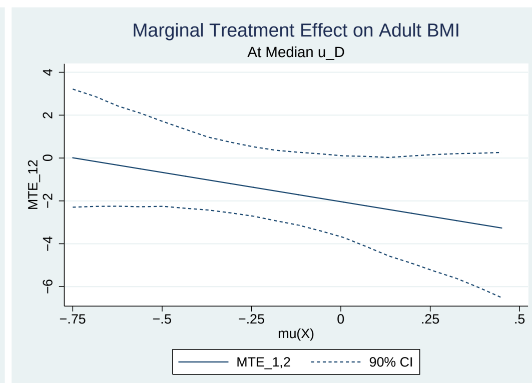
(c) Received TANF Benefits



(d) Employed



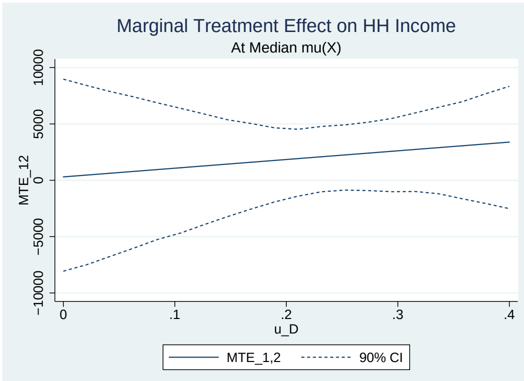
(e) Labor Force Participation



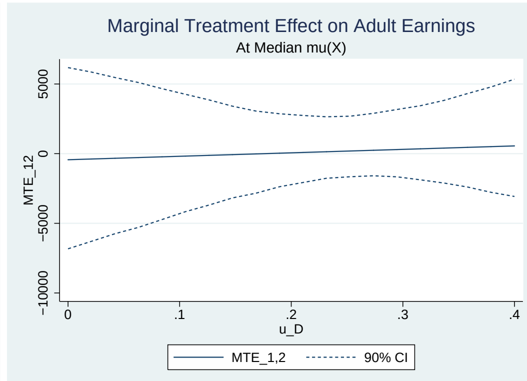
(f) BMI

Figure 16: Marginal Treatment Effects from D=1 to D=2 at Median u_D, 0.2

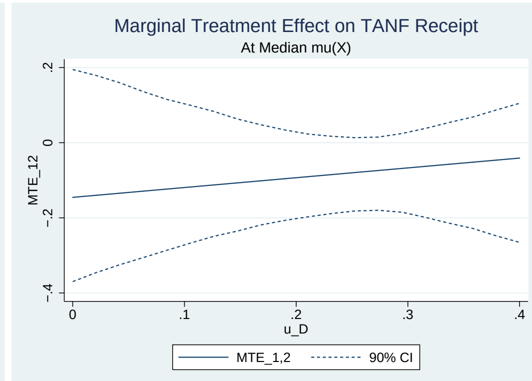
Confidence intervals are computed using 1,000 bootstrap replications



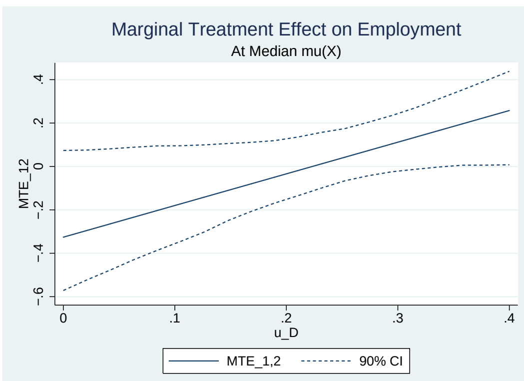
(a) HH Income



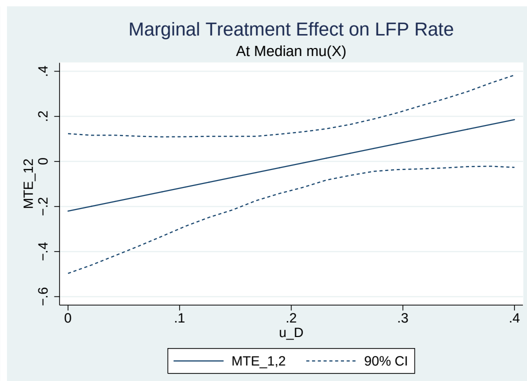
(b) Earnings



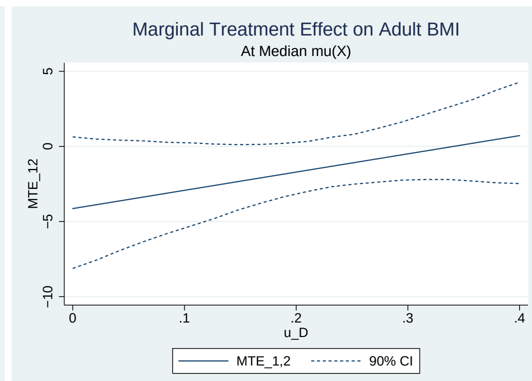
(c) Received TANF Benefits



(d) Employed



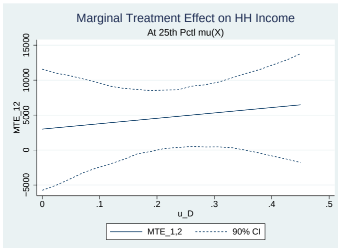
(e) Labor Force Participation



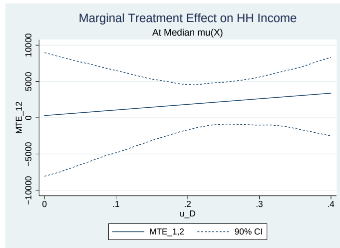
(f) BMI

Figure 17: Marginal Treatment Effects from D=1 to D=2 at Median $\mu(X)$, –0.12

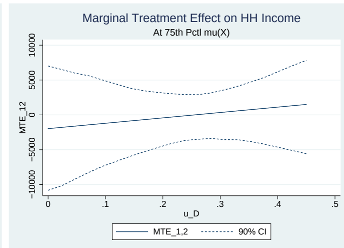
Confidence intervals are computed using 1,000 bootstrap replications



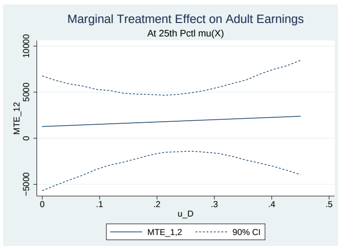
(a) HH Income



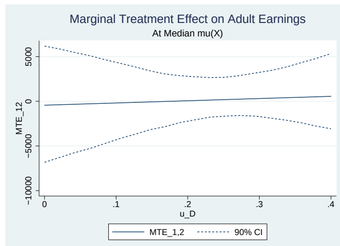
(b) HH Income



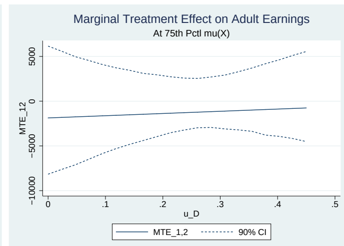
(c) HH Income



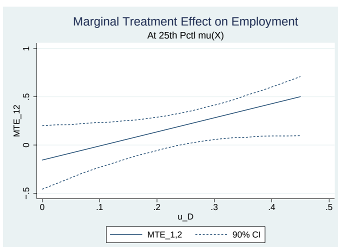
(d) Earnings



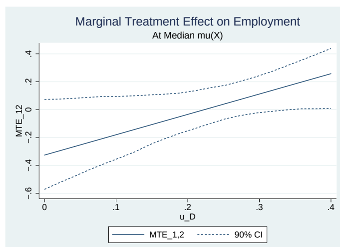
(e) Earnings



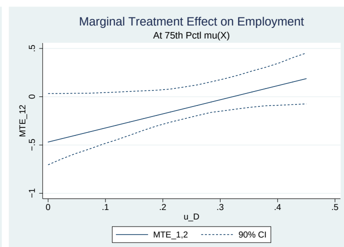
(f) Earnings



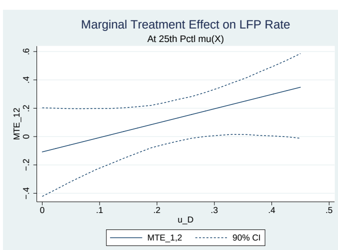
(g) Employed



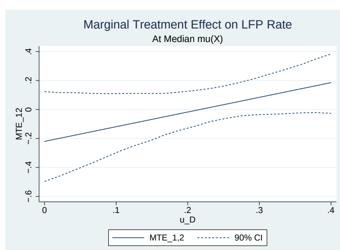
(h) Employed



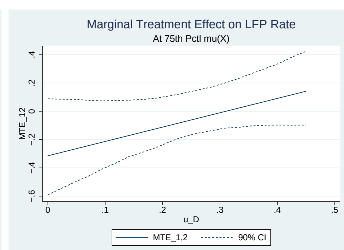
(i) Employed



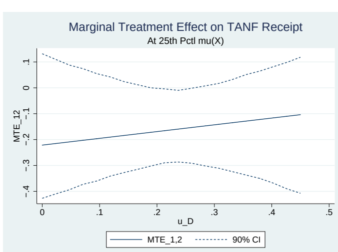
(j) Labor Force Participation



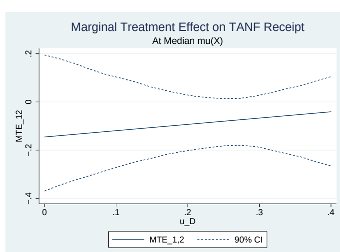
(k) Labor Force Participation



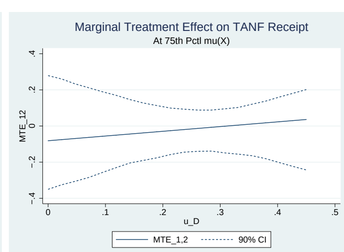
(l) Labor Force Participation



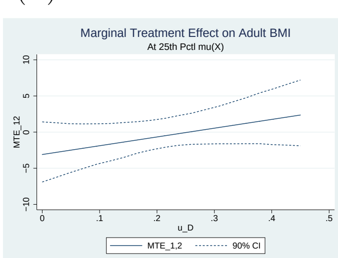
(m) Received TANF Benefits



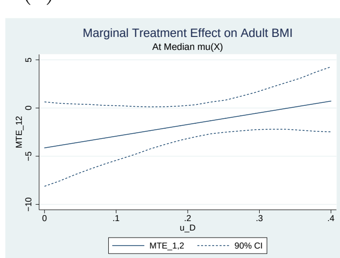
(n) Received TANF Benefits



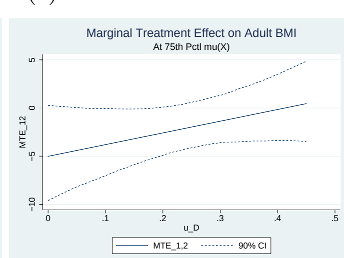
(o) Received TANF Benefits



(p) Adult BMI (Raw)



(q) Adult BMI (Raw)



(r) Adult BMI (Raw)

Figure 18: Marginal Treatment Effects from D=1 to D=2 at Low, Median, and High $\mu(X)$
Confidence intervals are computed using 1,000 bootstrap replications

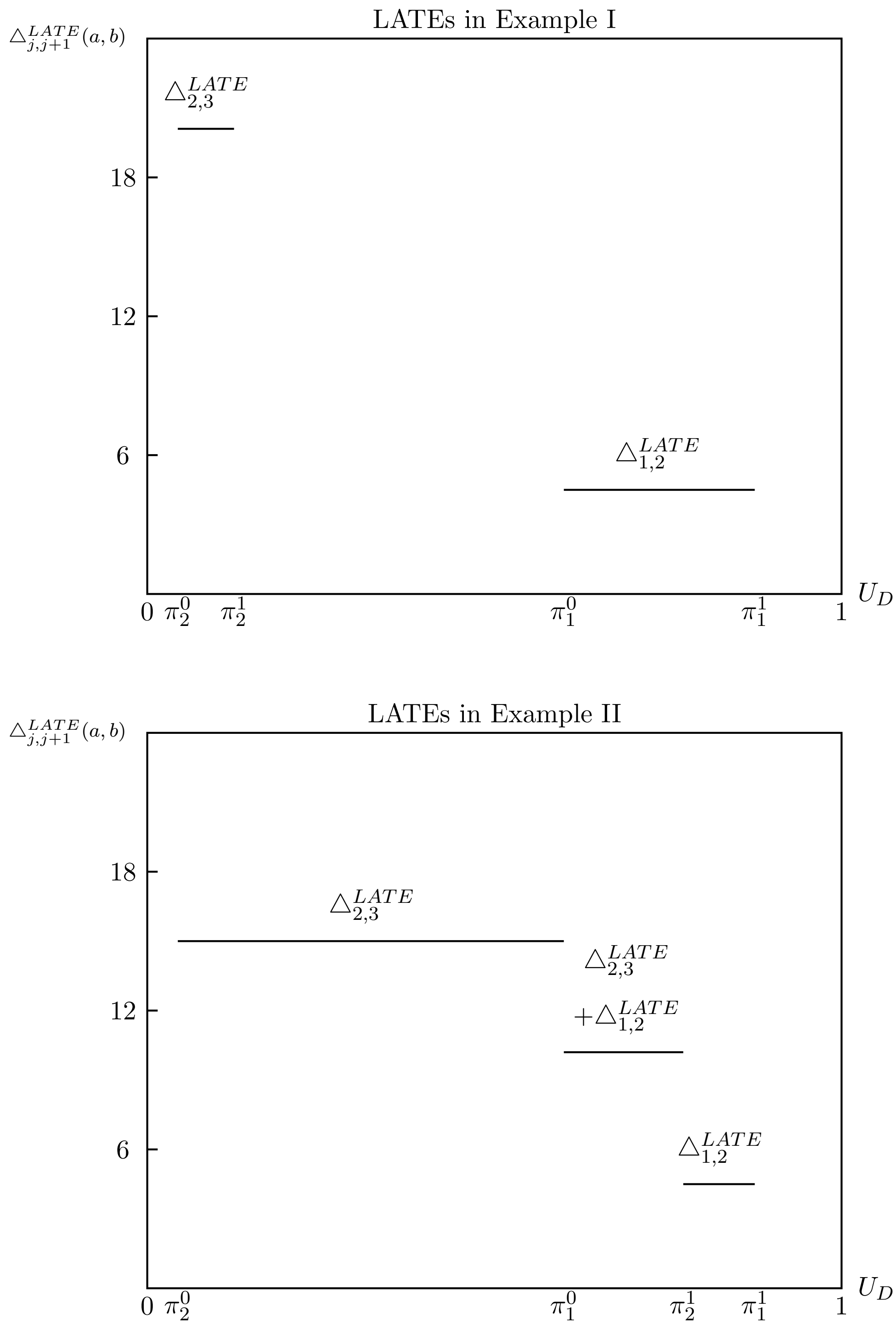


Figure 19: Examples I and II: Generalized Local Average Treatment Effects Given $\mu(X)$

Tables

Table 1: Proportion of Variance Explained by Principal Components Eigenvectors

Eigenvector	Eigenvalue	Proportion of Variance
1	4.14	0.69
2	0.67	0.11
3	0.51	0.08
4	0.35	0.06
5	0.22	0.04
6	0.12	0.02

Table 2: Principal Components Analysis: First Eigenvector Coefficients

Variable	Coefficient
Poverty Rate	-0.45
HS Graduation Rate	0.44
BA Attainment Rate	0.40
Percent Single-Headed HHs	-0.36
Male EPR	0.41
Female Unemployment Rate	-0.39

Table 3: Interim Neighborhood Quality by Site

	Mean			Median		
Site	Control	Section 8	Experimental	Control	Section 8	Experimental
Baltimore	8.8	12.8	15.1	2.2	4.4	4.9
Boston	15.1	20.3	25.9	10.9	13.5	17.0
Chicago	8.3	11.2	12.5	2.9	3.6	3.9
Los Angeles	6.0	7.6	13.2	1.2	3.9	6.3
New York City	5.7	8.2	13.7	1.0	1.7	2.6

Table 4: Ordered Choice Model Parameter Estimates

X_k	$\widehat{\beta}_k$	SE
Baseline Characteristics		
HH Member Victim	0.09	0.05
No Teens in HH	0.17	0.05
No Family in Nbd	0.08	0.05
Nbd Quality	0.12	0.01
Site Fixed Effects		
Baltimore	0	–
Chicago	–0.12	0.12
Los Angeles	–0.73	0.13
New York City	–0.70	0.12

Table 5: Adult LATE Estimates Averaged over all $\mu(X)$

Outcome	Program Effects		Neighborhood Quality Effects			
	$\widehat{\triangle}^{ITT}$	Control Mean	$\widehat{\triangle}^{LATE}_{1,2}(\max\{\widehat{\pi}^1_{20}, \widehat{\pi}^0_{10}\}, \widehat{\pi}^1_{10})$	Control Mean	$\max\{\widehat{\pi}^1_{20}, \widehat{\pi}^0_{10}\}$	$\widehat{\pi}^1_{10}$
Labor Market						
Household Income ($)	856	14,297	3,396	14,319	21.0	31.9
	(594)	(447)	(2,530)		(1.3)	(1.6)
Earnings ($)	–1	8,366	1,413	8,614		
	(512)	(384)	(1,928)			
Employed (%)	3.8	50.5	4.2	56.9		
	(2.5)	(1.9)	(10.3)			
In Labor Force (%)	4.5	61.6	7.8	64.4		
	(2.4)	(1.8)	(10.3)			
Welfare Benefits						
Received TANF (%)	–3.5	32.9	–10.4	31.2		
	(2.3)	(1.8)	(9.7)			
Health						
BMI (Raw)	–0.9	31.0	–2.1	33.0		
	(0.4)	(0.3)	(1.5)			

Note: Neighborhood quality effects standard errors are computed using 1,000 bootstrap replications.