Evidence of Neighborhood Effects from MTO:
LATEs of Neighborhood Quality

Dionissi Aliprantis   Francisca G.-C. Richter*

Federal Reserve Bank of Cleveland

First Version: March 2012. This Version: December 2013

Abstract: This paper finds evidence of positive neighborhood effects on adult labor market outcomes using the Moving to Opportunity (MTO) housing mobility experiment. Our results stand in such sharp contrast to the current literature because our analysis focuses on outcomes of the subpopulation induced by the program to move to a higher quality neighborhood, while previous analyses have focused on outcomes of either the entire population or the subpopulation induced by the program to move. We propose and implement a new strategy for identifying heterogeneous transition-specific effects that exploits the identification of the idiosyncratic component of an ordered choice model. We estimate Local Average Treatment Effects (LATEs) of the change in quality most commonly induced by MTO vouchers, between the first and second deciles of the national distribution of neighborhood quality. Although MTO vouchers induced much larger changes in neighborhood quality than standard Section 8 vouchers, these LATEs only pertain to a subpopulation representing under 10 percent of program participants.

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

1 Introduction

"The problem of the Twentieth Century" has yet to be resolved: The distributions of blacks and whites in the United States are dramatically different for nearly every outcome of importance. One prominent theory is that these differences in outcomes can be explained by effects from living in a poor, segregated, and socially isolated neighborhood (Wilson (1987)). The large differences in the neighborhood environments of blacks and whites (Wilson (1987), Massey and Denton (1993)), as well as the recent increase in the share of Americans living in census tracts with high poverty rates (Jargowsky (1997), Kneebone et al. (2011)), have motivated a large literature to investigate neighborhood effects.

Because households endogenously sort into neighborhoods, researchers have sought to identify neighborhood effects using the exogenous variation in neighborhoods induced by housing mobility programs. One of the best known housing mobility programs is the Gautreaux program, which was designed to desegregate public housing in Chicago and relocated public housing residents in a quasi-random manner using housing vouchers. Those who moved to high-income, white-majority suburbs through Gautreaux had much better education and labor market outcomes than those who moved to segregated city neighborhoods (Rosenbaum (1995), Mendenhall et al. (2006)).

The Moving to Opportunity (MTO) housing mobility experiment sought to replicate the quasi-experimental results from Gautreaux with a randomized experiment. MTO gave households living in high-poverty neighborhoods in five US cities the ability to enter a lottery for housing vouchers to be used in low-poverty neighborhoods. In a tremendous disappointment, effects from MTO were much smaller than effects from Gautreaux.

The lack of effects from MTO has been interpreted as evidence against the theory that neighborhood characteristics influence individuals' outcomes. For example, Ludwig et al. (2013a) interpret the lack of effects from MTO as evidence "Contrary to the widespread view that living in a disadvantaged inner-city neighborhood depresses labor market outcomes" (p 288). This interpretation links effects from MTO and effects from neighborhoods under the assumption that MTO induced changes in neighborhood characteristics large enough that "If neighborhood environments affect behavior ... then these neighborhood effects ought to be reflected in ITT [Intent-to-Treat] and TOT [Treatment-on-the-Treated] impacts on behavior" (Ludwig et al. (2008), pp 181-182). Aliprantis (2013a) explicitly states the assumptions necessary to make this link between effects from MTO and effects from neighborhood environments, and presents empirical evidence that these assumptions are unlikely to hold in the case of MTO.[1]

This paper abandons the approach of learning about neighborhood effects from ITT and TOT effects of the MTO program. Rather, we identify Local Average Treatment Effects (LATEs) of neighborhood quality on adult outcomes using the random variation in quality induced by the program together with an explicit model of households' selection into neighborhood quality.[2] We

---

[1]Related analyses can be found in Clampet-Lundquist and Massey (2008), Sampson (2008), and Quigley and Raphael (2008).

[2]We leave the analysis of child and youth outcomes for future work because we believe schools are likely to be

find that moving to a higher quality neighborhood had large, positive effects on employment, labor force participation, and Body Mass Index (BMI). Although effects on household income, individual earnings, and welfare receipt are not statistically significant, these effects are also estimated to be large and positive. We find no evidence from MTO against the theory that increasing neighborhood quality improves adult outcomes.

Our evidence of neighborhood effects from MTO stands in such sharp contrast to the current literature because we focus our analysis on the subpopulation induced by MTO to move to higher quality neighborhoods. The current literature draws conclusions about neighborhood effects based on the outcomes either of the entire population or of the subpopulation induced by the program to move. For example, estimates of *neighborhood* effects on adult labor market outcomes are entirely absent from the most prominent analysis of MTO because the *program* was found to have little effect on such outcomes (Kling et al. (2007)).[3] Our focus on moves to higher quality limits the generality of our parameters, which apply to less than 10 percent of program participants. Our LATEs of neighborhood quality also only pertain to moves between low quality neighborhoods, because not only did MTO induce a relatively small subset of households to move to better neighborhoods, but improvements in neighborhood quality were small even for these movers.

Thus while our results provide support for the idea that efforts to deconcentrate poverty are well-grounded, our estimates' lack of generality is itself evidence that policies ought to be carefully designed to achieve policy-makers' objectives. Despite the fact that households were more likely to move with Section 8 vouchers than MTO vouchers, changes in neighborhood quality were much smaller for Section 8 movers than for MTO movers, and variation by site was large. Since only about a quarter of eligible households are currently able to obtain a Section 8 housing voucher (Sard and Fischer (2012)), an area for future research is understanding what types of restrictions on voucher use might optimize the extent to which households are able to realize positive neighborhood effects through the subsidy, and which of these restrictions might be feasible to implement (McClure (2010)).

A final caveat when interpreting the policy implications of our results is that our estimated LATEs are total effects, encompassing the direct effects from changes in neighborhoods and from other programs. Factors other than MTO influencing neighborhood decline or revitalization in this time period are not explicitly modeled in our analysis. Nor are other programs that might differentially impact the treatment and control groups in the program (Heckman et al. (2000)), such as job training programs, education reform, other housing programs like HOPE VI, or welfare reform. The difficulty of separately identifying the direct effect of neighborhood environments points to the methodological limitations of conducting controlled, and not just randomized, experiments in social settings.

We begin our analysis by proposing a new strategy for identifying transition-specific LATEs

---

the most important "neighborhood" for children (Oreopoulos (2003)), and there are important issues in measuring school quality in the MTO data.

[3]Specifically, Kling et al. (2007) focus their attention on "outcomes that exhibit significant treatment effects" (p. 83).

with the aid of an ordered choice model in the absence of transition-specific instruments. Alternative identification strategies yield parameters that are weighted averages of such effects (Angrist and Imbens (1995)), or else require transition-specific instruments for each margin of choice (Heckman et al. (2006)). Our identification strategy is positioned between these two extremes, in that it is both economically interesting and empirically feasible. The key insight of the identification strategy is that by observing the continuous neighborhood quality associated with each program participant, it is possible to identify the unobserved, idiosyncratic component of their latent index in an ordered choice model.[4]

Our model specification and empirical implementation is guided by the desire to build on the analysis in Kling et al. (2007) in three important ways. First and foremost, we aim to focus on outcomes of the subgroup of households that were actually induced by the program to move to higher quality neighborhoods. With this goal, the model must be able to accurately characterize selection into neighborhoods of various quality levels. An important finding from our estimated ordered choice model of selection into neighborhood quality is that although households were more likely to move with Section 8 vouchers than MTO vouchers, variation by site was large, and changes in neighborhood quality were much smaller for Section 8 movers than for MTO movers in all sites except Boston (Galiani et al. (2012)).

Second, the model is also motivated by a desire to relax the assumption that effects are homogeneous conditional on observed characteristics. Simultaneously achieving the first two objectives of our analysis requires that our model specification balance the asymmetry inherent to Instrumental Variable (IV) identification strategies, which allow for general heterogeneity in response to treatment while restricting the patterns of heterogeneity in response to the instrument from complete generality (Heckman et al. (2006), Imbens and Angrist (1994), Aliprantis (2012)). We model heterogeneity in response to the instrument using a finite mixture model satisfying the monotonicity assumption.

Third, the empirical implementation of the model allows us to relax assumptions about the precise neighborhood characteristics that influence outcomes. We create and utilize a linear index of neighborhood quality informative about several neighborhood characteristics in addition to the neighborhood poverty rate. Using this precise quality index makes assumptions about the neighborhood characteristics that affect outcomes. Unfortunately, doing so is simply unavoidable: The literature using neighborhood poverty rate as the index measuring quality also makes assumptions about the neighborhood characteristics that affect outcomes. We believe that our index of quality imposes assumptions less likely to be violated than the index of neighborhood poverty (A lengthy discussion can be found in Aliprantis (2013a).).

The paper proceeds as follows: Section 2 specifies our joint model of neighborhood choice and potential outcomes, defining the treatment effect parameters we seek to identify in terms of this model. Section 3 presents our strategy for identifying these effects, with Appendix A presenting an

---

[4]The case of a binary treatment as originally developed in Imbens and Angrist (1994), Heckman and Vytlacil (1999), Heckman and Vytlacil (2005), and Carneiro et al. (2011).

intuitive discussion for a restricted version of our model, and Appendix B presenting a comparison with alternative identification strategies in the literature. Section 4 describes the MTO housing mobility program, the data used in estimation, and descriptive statistics of those data. Section 5 presents our empirical specification and estimation results, with Sections 5.1 and 5.2 focused on the ordered choice model, and Section 5.3 focused on neighborhood effects. Section 6 discusses our interpretation of the results, and Section 7 concludes.

2   Model and Definition of Treatment Effect Parameters

We now specify a joint model of neighborhood choice and potential outcomes based on the Marginal Treatment Effect (MTE) framework developed at various stages in Heckman et al. (2006), Heckman and Vytlacil (2005) (henceforth referred to as HUV and HV, respectively), Imbens and Angrist (1994), and Björklund and Moffitt (1987). We are interested in estimating the effects that neighborhood quality had on individuals' outcomes such as employment, income, and health. We begin by modeling the way agents select into different levels of a multi-valued treatment, which in this case is neighborhood quality. More details about Moving to Opportunity (MTO) are to be found in Section 4, but the key to our analysis is that we model the Section 8 and Experimental MTO rent vouchers randomly assigned through the program as potential reductions in the cost of moving to a higher quality neighborhood.

2.1   A Joint Model of Neighborhood Choice and Potential Outcomes

2.1.1   Selection into Neighborhood Quality

Consider a model of choice in which the treatment neighborhood quality can take any of J levels. We assume that for each level $j \in \{0, 1, \ldots, J\}$ there is a latent index D_j^* representing the net marginal benefit of moving from level j to level $j + 1$. Agents select level j^* that maximizes total utility $u(D)$:

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

where

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

Note that the initial level $u(1)$ is important only as a way of indexing the model.

We assume the $J + 1$ latent indeces (D_j^*) are additively-separable functions of observed characteristics X_i and whether household i was assigned a Section 8 ($Z_i^S \in \{0, 1\}$) or experimental MTO voucher ($Z_i^M \in \{0, 1\}$).[5] We denote the unobserved factors determining selection into neighborhood quality in the absence of any program, when assigned a Section 8 voucher, and when assigned an experimental MTO voucher by $\mathbf{V}_i = (V_i, V_i^S, V_i^M)$. Here, V_i captures unobserved factors that

---

[5]Given the indicator representations, $Z_i^S = 0$ and $Z_i^M = 0$ indicates random assignment to the control group.

influence neighborhood quality choice regardless of voucher assignment. V_i^M and V_i^S capture unobserved factors that influence neighborhood choice once a voucher has been assigned and households have started searching for housing. Households in the MTO group receive a Section 8 voucher and counseling, with the restriction to move to a neighborhood with a poverty rate of less than 10%. The Section 8 group is assigned an unrestricted Section 8 voucher and standard relocation assistance from the Housing Authority.[6] With that structure in mind, we can think of V and V^M capturing factors such as family composition and child care arrangements that can influence the move and may be addressed by counseling. $\tau_i^S = 1$ indicates that household i would move when offered a Section 8 voucher, $\tau_i^S = 0$ indicates they would not move, and τ_i^M is defined analogously. This notation allows us to specify the $J + 1$ latent indeces as follows:

A1: $D_{ij}^* = \mu(X_i) + \gamma_j^S Z_i^S \tau_i^S + \gamma_j^M Z_i^M \tau_i^M - C_j - V_i$ for $j = 0, \ldots, J$

with $D_0^* > \max\{0, D_1^*\}$

A2: $C_j \qquad\qquad\qquad < C_{j+1}$ for $j = 0, \ldots, J - 1$

with $C_0 = -\infty$ and $C_J = \infty$

A3: $C_j - \gamma_j^S Z^S - \gamma_j^M Z^M < C_{j+1} - \gamma_{j+1}^S Z^S - \gamma_{j+1}^M Z^M$ for $j = 0, \ldots, J - 1$

with $C_0 - \gamma_0^S Z^S - \gamma_0^M Z^M = -\infty$

and $C_J - \gamma_J^S Z^S - \gamma_J^M Z^M = \infty$

A4: $\tau_i^S = \mathbf{1}\{\mu^S(X_i) - V_i^S \geq 0\}$

A5: $\tau_i^M = \mathbf{1}\{\mu^M(X_i) - V_i^M \geq 0\}$

A6: $\gamma_j^S \geq 0$ for all $j = 1, \ldots, J - 1$ and $\gamma_j^S > 0$ for at least one $j \in \{1, \ldots, J - 1\}$

A7: $\gamma_j^M \geq 0$ for all $j = 1, \ldots, J - 1$ and $\gamma_j^M > 0$ for at least one $j \in \{1, \ldots, J - 1\}$

Here $\mu(X_i)$ is the gain from moving up one level and is independent of the level, and C_j is a transition-specific cost that increases at each level. C_j may include increased search costs of housing in higher quality, higher priced neighborhoods due to a reduced supply of units available at the metro-level Fair Market Rent (FMR).[7] We do not explicitly include rents or house prices in our model because households pay 30 percent of their income regardless of the unit's rent, as long as it is a Section 8 eligible unit. Thus the probability of finding a Section 8 eligible unit at the metro-level FMR is the key to the household's choice, and is mainly captured in the C_j. Further discussion is presented in Appendix C.

---

[6]Once vouchers were assigned, families had four to six months to move (Kling et al. (2007)).

[7]The FMR is usually set at 40th percentile of local gross rents for typical non-substandard rental units, adjusted by household size.

Assumptions A1-A7 allow for idiosyncratic, heterogeneous response to treatment. For each of the voucher programs $\gamma_j^S Z_i^S \tau_i^S$ and $\gamma_j^M Z_i^M \tau_i^M$ can be interpreted as a reduction in the j to $j+1$ transition-specific cost brought about by offer ($Z_i = 1$) and take-up ($\tau_i = 1$) of the voucher. Households that receive a voucher ($Z_i = 1$) but do not move ($\tau_i = 0$) are those for which the voucher does not represent a cost reduction of moving. Assumption A2 ensures a cost ranking among levels justifying an ordered choice model, and A3 ensures that the voucher-induced average cost reduction of moving one quality level higher for compliers (γ_j) preserves this cost ranking.

Although we would ideally be able to allow for even more general heterogeneity in response to the instrument, the bivariate latent classes in A1-A7 relax the assumption of homogeneous responses to vouchers, and are a tractable, interpretable way of modeling this heterogeneity (Greene et al. (2008)). Furthermore, while the types of heterogeneity allowed under Assumptions A1-A7 are not completely general, comparable restrictions are necessary for Instrumental Variables (IV) to identify causal parameters of interest, as are additional modeling assumptions (Vytlacil (2002), Vytlacil (2006)). While IV identification strategies can allow for general heterogeneity of outcomes, the patterns of heterogeneity are not completely general in response to the instrument under which IV identifies causal effects (HUV, HV, Imbens and Angrist (1994), Aliprantis (2012)).

Recall that V_i represents the unobserved cost for household i of moving up one level in the absence of a voucher program, and V_i^S and V_i^M are unobserved variables influencing the decision of household i to take up a Section 8 voucher and an MTO voucher when these are offered. We allow for these variables to be correlated in an arbitrary way, possibly exhibiting patterns of correlation anywhere between being exactly identical variables to being independently distributed variables to being negatively correlated variables. However, for the sake of identification we do adopt the distributional assumption:

A8:

$$\mathbf{V}_i \equiv (V_i, V_i^S, V_i^M) \sim \mathcal{N} \left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho^S & \rho^M \\ \rho^S & 1 & \rho^{SM} \\ \rho^M & \rho^{SM} & 1 \end{bmatrix} \right).$$

We stress that the role of Assumption A8 in aiding identification is entirely through the choice model, and absolutely no distributional assumption whatsoever is placed on potential outcomes.

For the sake of exposition in coming sections we denote the CDF of V by $F_V(\cdot)$ and define

$$U_D \equiv F_V(V).$$

Figure 1 shows the ordered choice model in Assumptions A1-A6 graphically. The top panel reflects marginal utilities D_j^* for $j \in \{1, \ldots, J\}$, and the bottom panel illustrates the level of utility.

$$u(D = m) = u(1) + \sum_{j=1}^{m} D_j^*.$$

The combined benefits and costs of moving to neighborhoods of various levels of quality are

determined by social interactions, market forces, and political processes. Since the number of program movers is small relative to neighborhood size, our model assumes that movers do not influence the market price of neighborhood quality, the technology by which resources are produced, or the political process by which resources are distributed. These partial equilibrium assumptions are not only reasonable, but also greatly facilitate estimation relative to a general equilibrium model of neighborhood effects with social interactions, market forces, and political processes. Consider that we are able to estimate our joint model of both selection and outcomes, when it is difficult to estimate models of selection (Manski (1993), Brock and Durlauf (2007)) or outcomes (Manski (2013)) even when modeled separately and focused only on the mechanism of social interaction effects.[8] Appendix C discusses the interpretation of parameters in our model in greater detail, including a discussion of why rents or housing prices are not included in the model.

2.1.2 Potential Outcomes

We are interested in learning about the counterfactual joint distribution of some outcome Y_i given treatment levels $D_i \in \{1, \ldots, J\}$, invariant observed characteristics X_i, and the unobserved components of selection into treatment, $\mathbf{V}_i$. In this analysis we will focus on the conditional mean $E[Y_i(D_i)|X_i, V_i, V_i^M]$.[9] We assume potential outcomes are an additively-separable function of treatment level, invariant observed characteristics, and level-specific unobservable characteristics as follows:

$$Y_j(X, U_j) = Y(D = j, X, U_j) = \mu_j(X) + U_j \quad \text{for} \quad j = 1, \ldots, J. \tag{2}$$

We add the following assumptions to A1-A8 that are similar to those found in the ordered choice model developed in HUV:

A9 $(U_{ij}, \mathbf{V}_i) \perp\!\!\!\perp \mathbf{Z}_i$ for all $j = 1, \ldots, J$

A10 $E[|Y_j|] < \infty$ for all $j = 1, \ldots, J$

We emphasize again that no assumptions are made about potential outcomes through the U_{ij} other than the valid instrument Assumption A9. This relaxes a key assumption in the most similar analyses to ours, Kling et al. (2007) and Ludwig and Kling (2007), that $U_{ij} = U_i$ for all j, so that $U_{ij+1} - U_{ij} = 0$ is independent of V_i conditional on X_i, precluding the possibility of essential heterogeneity as defined in Heckman et al. (2006).[10]

2.2 Local Average Treatment Effects of Neighborhood Quality

Equations 1 and 2, together with assumptions A1-A10, define a joint selection and outcome model, and we might be interested in describing the Data Generating Process (DGP) represented

---

[8] Aliprantis and Carroll (2013) and Badel (2010) present empirical studies of neighborhood effects in dynamic models.

[9] See Carneiro and Lee (2009) for estimation of not only means, but also distributions of potential outcomes.

[10] See Aliprantis (2013a) for a detailed discussion.

by this model in many ways. Causal effects are thought to be more interesting than correlation coefficients for describing DGPs because they are thought to represent features of the DGP that are invariant to outside interventions (Woodward (2000), Aliprantis (2013b)). One causal effect of the DGP represented by our model is the j to $j+1$ transition-specific Local Average Treatment Effect (LATE) for the experimental MTO voucher:

$$\triangle_{j,j+1}^{LATE}\Big(\ \mu(X), V \in [\underline{v}, \overline{v}], Z^M\ \Big) \equiv E\big[\ Y_{j+1} - Y_j \mid \mu(X), V \in [\underline{v}, \overline{v}], \tau^M = 1\ \big]$$

The $\triangle_{j,j+1}^{LATE}$ parameter is defined in terms of the experimental MTO voucher Z^M.[11] Even if we indexed the parameter in terms of how individuals select into treatment in the absence of any program (ie, using (X_i, V_i)), the LATE for this subset of (X_i, V_i) would also depend on individuals' idiosyncratic response to the specific program in question. In our context, the subpopulation of MTO compliers is almost certainly different than the Section 8 compliers. As a result it is most likely the case that $\triangle_{j,j+1}^{LATE}\big(\ \mu(X), V \in [\underline{v}, \overline{v}], Z^M\ \big) \neq \triangle_{j,j+1}^{LATE}\big(\ \mu(X), V \in [\underline{v}, \overline{v}], Z^S\ \big)$, where

$$\triangle_{j,j+1}^{LATE}\Big(\ \mu(X), V \in [\underline{v}, \overline{v}], Z^S\ \Big) \equiv E\big[\ Y_{j+1} - Y_j \mid \mu(X), V \in [\underline{v}, \overline{v}], \tau^S = 1\ \big].$$

We discuss the invariance of the estimated LATE parameter in Section 5.3. One related point about invariance is especially important in our application. If the Stable Unit Treatment Value Assumption (SUTVA) fails to hold, say due to social interactions, then the joint distribution of $(V, V^S, V^M, U_1, \ldots, U_J)$, and therefore the LATE, need not be invariant even to different realizations of the same policy, even when randomized (Sobel (2006), Aliprantis (2013a)).

We interpret effects of neighborhood quality as the result of social interactions and neighborhood-level resources (by way of the technology utilizing and distributing resources across and within neighborhoods). In our model, each level of neighborhood quality is thought to represent, on average, homogeneous types of social interactions and resources available at the neighborhood level. We assume agents experience neighborhood quality and cannot determine it because we empirically define neighborhoods as census tracts, which contain about 4,000 residents on average. If we were thinking about smaller reference groups like social groups in classrooms, we would be more interested in incorporating the endogenous formation of reference groups into the model as in Badev (2012) and Goldsmith-Pinkham and Imbens (2011). Since we are dealing with large N, we believe our partial equilibrium abstraction is appropriate.

3   Identification

Defining causal parameters of the model in Section 2 is a distinct task from identifying them from an observed joint distribution (Heckman (2008), Pearl (2009a), Pearl (2009b)). Thankfully, in the

---

[11]The fact that the LATE is instrument-specific has received much discussion in the literature. See Imbens and Angrist (1994), Heckman and Urzúa (2010), HV, HUV, Heckman and Vytlacil (2001), Heckman (2010), or Imbens (2010) for some of these discussions.

case of MTO we are equipped with a randomized experiment that greatly facilitates identification of causal parameters of interest. The randomized voucher assignment in MTO, together with the shape restrictions in Assumptions A1-A8, allow us to identify the parameters of our ordered choice model. Since we also observe the precise neighborhood quality chosen by households, and their type τ_i when assigned a voucher, we are able to identify the unobserved component of the latent index households use to select their neighborhood quality.

The first stage of our identification strategy is to find households across experimental and control groups deemed by the estimated ordered choice model to be "similar" in terms of both observed and unobserved factors determining selection into neighborhood quality in the absence of a program. The second stage begins by finding "mobility" regions of the observed and unobserved characteristics for which receiving an MTO voucher would actually induce households to move to higher quality neighborhoods. This step is crucial, because the precise parameters that can be identified even with the MTO randomized experiment still depends on the joint distribution observed in the data (Aliprantis (2013a)). The second stage is completed by comparing the outcomes of "similar" households in "mobile" characteristic regions who were randomly assigned to the control and MTO voucher groups.

3.1 First Stage: Identifying Ordered Choice Model Parameters

Recall that given X_i, $\mathbf{Z}_i \equiv (Z_i^S, Z_i^M)$, and $\mathbf{V}_i$, the net marginal benefit of moving from neighborhood of quality j to a neighborhood of quality $j+1$ (D_{ij}^*) is decreasing in j. Thus the optimal quality level $j^* \in \{1, \ldots, J\}$ satisfies the standard ordered choice model condition:

$$D_i = j^* \quad \Longleftrightarrow \quad D_{ij^*}^* \leq 0 < D_{ij^*-1}^*, \tag{3}$$

which under our specification is equivalent to:

$$D_i = j^* \quad \Longleftrightarrow \quad \mu(X_i) + \gamma_{j^*}^S Z_i^S \tau_i^S + \gamma_{j^*}^M Z_i^M \tau_i^M - C_{j^*} < V_i \leq \mu(X_i) + \gamma_{j^*-1}^S Z_i^S \tau_i^S + \gamma_{j^*-1}^M Z_i^M \tau_i^M - C_{j^*-1}. \tag{4}$$

An implication of Assumption A8 is that the marginal distributions $V_i \sim \mathcal{N}(0,1)$,

$$(V_i, V_i^S) \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho_S \\ \rho_S & 1 \end{bmatrix} \right), \quad \text{and} \quad (V_i, V_i^M) \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho_M \\ \rho_M & 1 \end{bmatrix} \right). \tag{5}$$

We also know from Assumptions A4 and A5 that the probabilities of moving when offered a Section 8 and experimental MTO voucher is $P(V_i^S \leq \mu^S(X_i))$ and $P(V_i^M \leq \mu^M(X_i))$, respectively.

For the *Control* Group we do not observe whether the household would move with either type of voucher, while for both the *Section 8* and *Experimental MTO* voucher groups we observe whether the household is a "mover" or a "never-mover" with respect to the voucher they received (ie, whether they moved when offered that type of voucher). Together with actually observing the household's τ^S when $Z^S = 1$ and τ^M when $Z^M = 1$, the ordered choice condition in Equation 4

and the marginal distributions in Equation 5 allow us to express the probability of observing $D_i = j$ for households in each observed group. Where $\Phi_2(a, b; \rho)$ is the cumulative distribution function of the standardized bivariate normal distribution with correlation coefficient ρ, these probabilities are:

Control Group
$$\Pr(D_i = j | X_i, Z_i^S = 0, Z_i^M = 0) = \Phi\left(\ \mu(X_i) - C_{j-1}\ \right) - \Phi\left(\ \mu(X_i) - C_j\ \right) \qquad (6)$$

Section 8 Voucher Movers and Non-Movers
$$\Pr(D_i = j | X_i, Z_i^S = 1, Z_i^M = 0, \tau_i^S = 1) = \Phi_2\left(\ \mu(X_i) + \gamma_{j-1}^S - C_{j-1}, \quad \mu^S(X_i); \quad \rho_S\ \right)$$
$$- \Phi_2\left(\ \mu(X_i) + \gamma_j^S - C_j, \quad \mu^S(X_i); \quad \rho_S\ \right)$$

$$\Pr(D_i = j | X_i, Z_i^S = 1, Z_i^M = 0, \tau_i^S = 0) = \Phi_2\left(\ \mu(X_i) - C_{j-1}, \quad -\mu^S(X_i); \quad \rho_S\ \right)$$
$$- \Phi_2\left(\ \mu(X_i) - C_j, \quad -\mu^S(X_i); \quad \rho_S\ \right)$$

MTO Voucher Movers and Non-Movers
$$\Pr(D_i = j | X_i, Z_i^S = 0, Z_i^M = 1, \tau_i^M = 1) = \Phi_2\left(\ \mu(X_i) + \gamma_{j-1}^M - C_{j-1}, \quad \mu^M(X_i); \quad \rho_M\ \right)$$
$$- \Phi_2\left(\ \mu(X_i) + \gamma_j^M - C_j, \quad \mu^M(X_i); \quad \rho_M\ \right)$$

$$\Pr(D_i = j | X_i, Z_i^S = 0, Z_i^M = 1, \tau_i^M = 0) = \Phi_2\left(\ \mu(X_i) - C_{j-1}, \quad -\mu^M(X_i); \quad \rho_M\ \right)$$
$$- \Phi_2\left(\ \mu(X_i) - C_j, \quad -\mu^M(X_i); \quad \rho_M\ \right).$$

These probabilities allow us to identify the parameters of the model in Section 2 by expressing its log-likelihood function as:

$$\mathcal{LL}(\theta | \mathbf{X}, \mathbf{Z}, \mathbf{D}, \tau) = \sum_{i=1}^{N} \sum_{j=1}^{J} \mathbf{1}\{D_i = j\} \ln\left(\Pr\left(D_i = j | X_i, Z_i^S, Z_i^M, \tau\right)\right)$$
$$= \sum_{i=1}^{N^0} \sum_{j=1}^{J} \mathbf{1}\{D_i = j\} \ln\left(\Pr\left(D_i = j | X_i, Z_i^S = 0, Z_i^M = 0\right)\right) \qquad (7)$$
$$+ \sum_{i=1}^{N^S} \sum_{j=1}^{J} \mathbf{1}\{D_i = j\} \ln\left(\sum_{t=0}^{1} \Pr\left(D_i = j | X_i, Z_i^S = 1, Z_i^M = 0, \tau_i^S = t\right) \mathbf{1}(\tau_i^S = t)\right)$$
$$+ \sum_{i=1}^{N^M} \sum_{j=1}^{J} \mathbf{1}\{D_i = j\} \ln\left(\sum_{t=0}^{1} \Pr\left(D_i = j | X_i, Z_i^S = 0, Z_i^M = 1, \tau_i^M = t\right) \mathbf{1}(\tau_i^M = t)\right).$$

3.2 First Stage: Identifying V_i

Suppose that we actually observe a truly continuous variable that is arbitrarily discretized, such as Body Mass Index (BMI), school quality, or in this case, neighborhood quality q. In this case the discrete marginal benefit function D^*_{ij} in A1 can be interpreted as a continuous marginal benefit function

$$D^*_i(q) = \mu(X_i) + \gamma^S(q)Z^S_i\tau^S_i + \gamma^M(q)Z^M_i\tau^M_i - C(q) - V_i \tag{8}$$

evaluated at the thresholds $\overline{q}^D_j$ where treatment levels are defined such that $D_i = j$ for quality $q \in [\underline{q}^D_j, \overline{q}^D_j]$:

$$D^*_{ij} = D^*_i(\overline{q}^D_j) = \mu(X_i) + \gamma^S(\overline{q}^D_j)Z^S_i\tau^S_i + \gamma^M(\overline{q}^D_j)Z^M_i\tau^M_i - C(\overline{q}^D_j) - V_i.$$

Given estimates $\widehat{\gamma}(\overline{q}^D_j) = \widehat{\gamma}^D_j$ and $\widehat{C}(\overline{q}^D_j) = \widehat{C}^D_j$ from the ordered choice model, we can interpolate (in the simplest case, linearly) between the thresholds $\underline{q}^D_j$ and $\overline{q}^D_j$ to construct the components of the continuous marginal benefit function $D^*_i(q)$:

$$\widehat{\gamma}(q) = \widehat{\gamma}^D_j + (q - \overline{q}^D_j)(\frac{\widehat{\gamma}^D_{j+1} - \widehat{\gamma}^D_j}{\overline{q}^D_{j+1} - \overline{q}^D_j}) \quad \text{for } q \in (\overline{q}^D_j, \overline{q}^D_{j+1}) = (\underline{q}^D_{j+1}, \overline{q}^D_{j+1}) \tag{9}$$

and

$$\widehat{C}(q) = \widehat{C}^D_j + (q - \overline{q}^D_j)(\frac{\widehat{C}^D_{j+1} - \widehat{C}^D_j}{\overline{q}^D_{j+1} - \overline{q}^D_j}) \quad \text{for } q \in (\overline{q}^D_j, \overline{q}^D_{j+1}) = (\underline{q}^D_{j+1}, \overline{q}^D_{j+1}). \tag{10}$$

If we assume $D^*(q)$ is strictly decreasing, $D^*(0) > \max\{0, D^*(\epsilon)\}$ for any $\epsilon \in (0, 1]$, and that households maximize their utility at q^*, then q^* satisfies the following first order condition equivalent to conditions in Equations 3 and 4 in the multilevel case:

$$q_i = q^* \iff \mu(X_i) + \gamma^S(q^*)Z^S_i\tau^S_i + \gamma^M(q^*)Z^M_i\tau^M_i - C(q^*) - V_i = 0. \tag{11}$$

Equations 8 and 11 imply that

$$V_i = \mu(X_i) + \gamma^S(q_i)Z^S_i\tau^S_i + \gamma^M(q_i)Z^M_i\tau^M_i - C(q_i). \tag{12}$$

Since we actually observe the chosen value of the continuous neighborhood quality treatment chosen by households ($q_i = q^*$), and not some arbitrary discretization of that choice, we can use the estimates from the ordered choice model to construct estimates of $\widehat{\mu}(X_i)$, $\widehat{\gamma}^S(q_i)$, $\widehat{\gamma}^M(q_i)$, and $\widehat{C}(q_i)$ for all i. And since we observe Z^S_i, Z^M_i for all individuals, as well as τ^S_i when $Z^S_i = 1$ and τ^M_i when $Z^M_i = 1$, we can evaluate the right hand side of Equation 12 to identify

$$\widehat{V}_i = \widehat{\mu}(X_i) + \widehat{\gamma}^S(q_i)Z^S_i\tau^S_i + \widehat{\gamma}^M(q_i)Z^M_i\tau^M_i - \widehat{C}(q_i). \tag{13}$$

Identifying V_i is the crucial step in our identification strategy, and is what differentiates it from other approaches in the literature. Appendix B compares our identification strategy with

others in the literature, but a brief comparison is as follows: A key feature of our strategy is that it allows for heterogeneous effects even conditional on observed characteristics, relaxing assumptions of the generalized propensity score framework as developed in Imai and van Dyk (2004) or Hirano and Imbens (2005). By identifying V_i, we are able to identify j to $j + 1$ transition-specific effects, an improvement over the average of such effects represented by the Average Causal Response (ACR) of Angrist and Imbens (1995). And finally, our identification strategy only requires the existence of a single instrument. This single instrument requirement relaxes the assumptions of the alternative strategy in the literature for identifying heterogeneous transition-specific effects, which requires the existence of transition-specific instruments (HUV).

Although we could have specified functional forms for $C(q)$, $\gamma^S(q)$, and $\gamma^M(q)$ and directly estimated a continuous version of the model, we estimated an ordered choice model because doing so allows for flexible specifications of these functions. Identification in our model comes from the increasing cost of moving to a higher quality neighborhood, but the ordered choice model allows for this increase to take extremely general forms. Although this model might be classified as a Tobit-type model, but it is important to remember that normality assumptions imposed for the sake of identification are only placed on the ordered choice model, and not on the potential outcomes.[12]

Additionally, even when we are interested in the effects on outcomes of moving between discrete treatment levels defined by the intervals of quality $\{\underline{q}_j^Y, \overline{q}_j^Y\}$, we can improve precision by estimating the ordered choice model using treatment levels defined by intervals $\{\underline{q}_j^D, \overline{q}_j^D\}$ determined by the mass of the population observed in the data to have chosen various quality levels. Interpolating as in Equations 9 and 10, we can use the estimates of the ordered choice model obtained from the intervals $\{\underline{q}_j^D, \overline{q}_j^D\}$ to generate estimates over the intervals $\{\underline{q}_j^Y, \overline{q}_j^Y\}$ of interest for effects on outcomes.

3.3 Second Stage: Identifying LATEs of Neighborhood Quality

We use the potential outcome notation $D_i(Z^M = z^m, Z^S = z^s)$ to denote the counterfactual value that D_i would obtain if Z_i^M and Z_i^S were set to z^m and z^s, respectively. We denote the support of $\mu(X_i)$ by $\mathcal{M}$ and remind the reader that $V_i \in \mathbb{R}$ captures unobserved components of the cost of moving without a voucher. Finally, we also adopt the potential outcome notation Y_j to denote the value that Y would obtain if D were externally set to j.

We are interested in the regions of observed and unobserved characteristics, $(\mu(X_i), V_i)$, where there are households that would select into treatment level j in the absence of a voucher, but that would select into treatment level $j + 1$ if induced to move with the MTO voucher. We do not restrict this set by households' response to a Section 8 voucher. Our notation allows us to define

---

[12]See Amemiya (1985) for a thorough classification of Tobit models. Related applications can be found in Ross and Tootell (2004) and Munnell et al. (1996).

these sets as:

$$\mathcal{S}_{j,j+1}^M \equiv \left\{ \begin{array}{ll} \big(\mu(X_i), V_i\big) & \mid \quad D_i(Z^M = 0, Z^S = 0) = j \qquad \forall \quad i, \\[2mm] & \quad D_i(Z^M = 1, Z^S = 0) = j \qquad \text{if } \tau_i^M = 0, \\[2mm] & \quad D_i(Z^M = 1, Z^S = 0) \in \{j, j+1\} \text{ if } \tau_i^M = 1 \end{array} \right\} \subset \quad \mathcal{M} \times \mathbb{R}$$

A similar set can be defined for the Section 8 group. Since it is possible that a household offered a voucher will move, but not to a higher quality neighborhood (ie, $Z_i^M = 1$, $\tau_i^M = 1$, and $D_i = j$) , we also define the set of j to $j+1$ compliers as

$$\mathcal{C}_{j,j+1}^M \equiv \left\{ \begin{array}{ll} \big(\mu(X_i), V_i\big) & \mid \quad D_i(Z^M = 0, Z^S = 0) = j \qquad \forall \quad i, \\[2mm] & \quad D_i(Z^M = 1, Z^S = 0) = j \qquad \text{if } \tau_i^M = 0, \\[2mm] & \quad D_i(Z^M = 1, Z^S = 0) = j+1 \text{ if } \tau_i^M = 1 \end{array} \right\} \subseteq \quad \mathcal{S}_{j,j+1}^M \subset \quad \mathcal{M} \times \mathbb{R}.$$

Focusing on the group of individuals for which the program would randomly induce "mover" households to move from a neighborhood of quality j to a neighborhood of quality $j+1$, define

$$\pi(\mathcal{C}_{j,j+1}^M) \equiv \Pr\bigg(D = j+1 \quad \mid \quad Z^M = 1, \ (\mu(X), V) \in \mathcal{S}_{j,j+1}^M \ \bigg)$$

and note that

$$\pi(\mathcal{C}_{j,j+1}^M) \leq \Pr\big(\tau^M = 1 \quad \mid \quad (\mu(X), V) \in \mathcal{S}_{j,j+1}^M \ \big).$$

The Wald estimator applied to the subsample of experimental and control households in $\mathcal{S}_{j,j+1}^M$ identifies the j to $j+1$ transition-specific LATE for compliers:

$$
\begin{aligned}
&\frac{\Big\{ E\big[Y \mid Z^M = 1, (\mu(X), V) \in \mathcal{S}_{j,j+1}^M \big] \Big\} - \Big\{ E\big[Y \mid Z^M = 0, (\mu(X), V) \in \mathcal{S}_{j,j+1}^M \big] \Big\}}{\Big\{ E\big[D \mid Z^M = 1, (\mu(X), V) \in \mathcal{S}_{j,j+1}^M \big] \Big\} - \Big\{ E\big[D \mid Z^M = 0, (\mu(X), V) \in \mathcal{S}_{j,j+1}^M \big] \Big\}} \\[3mm]
&= \frac{E\bigg(Y_{j+1} - Y_j \mid (\mu(X), V) \in \mathcal{C}_{j,j+1}^M \bigg) \pi(\mathcal{C}_{j,j+1}^M)}{\pi(\mathcal{C}_{j,j+1}^M)} \qquad\qquad (14) \\[3mm]
&= E\big[Y_{j+1} - Y_j \mid (\mu(X), V) \in \mathcal{C}_{j,j+1}^M, \tau^M = 1 \big] \\[2mm]
&\equiv \triangle_{j,j+1}^{LATE}\bigg((\mu(X), V) \in \mathcal{C}_{j,j+1}^M, \ \tau^M = 1 \bigg),
\end{aligned}
$$

14

where line 14 follows from the facts that

$$\left\{ E\left[\, Y \mid Z^M = 1, (\mu(X), V) \in \mathcal{S}^M_{j,j+1} \,\right] \right\} - \left\{ E\left[\, Y \mid Z^M = 0, (\mu(X), V) \in \mathcal{S}^M_{j,j+1} \,\right] \right\}$$

$$= \left\{ E\left[\, Y_{j+1} \mid (\mu(X), V) \in \mathcal{C}_{j,j+1} \,\right] \pi(\mathcal{C}_{j,j+1}) \quad + \quad E\left[\, Y_j \mid (\mu(X), V) \in \mathcal{C}^M_{j,j+1} \,\right] (1 - \pi(\mathcal{C}^M_{j,j+1})) \right\}$$

$$- \left\{ E\left[\, Y_j \mid (\mu(X), V) \in \mathcal{C}^M_{j,j+1} \,\right] \pi(\mathcal{C}^M_{j,j+1}) \quad + \quad E\left[\, Y_j \mid (\mu(X), V) \in \mathcal{C}^M_{j,j+1} \,\right] (1 - \pi(\mathcal{C}^M_{j,j+1})) \right\}$$

$$= \left(E\left[\, Y_{j+1} \mid (\mu(X), V) \in \mathcal{C}^M_{j,j+1} \,\right] \quad - \quad E\left[\, Y_j \mid (\mu(X), V) \in \mathcal{C}^M_{j,j+1} \,\right] \right) \pi(\mathcal{C}^M_{j,j+1})$$

$$+ \left(E\left[\, Y_j \mid (\mu(X), V) \in \mathcal{C}^M_{j,j+1} \,\right] \quad - \quad E\left[\, Y_j \mid (\mu(X), V) \in \mathcal{C}^M_{j,j+1} \,\right] \right) (1 - \pi(\mathcal{C}^M_{j,j+1}))$$

$$= \left(E\left[\, Y_{j+1} \mid (\mu(X), V) \in \mathcal{C}^M_{j,j+1} \,\right] \quad - \quad E\left[\, Y_j \mid (\mu(X), V) \in \mathcal{C}^M_{j,j+1} \,\right] \right) \pi(\mathcal{C}^M_{j,j+1})$$

and

$$\left\{ E\left[\, D \mid Z^M = 1, (\mu(X), V) \in \mathcal{S}^M_{j,j+1} \,\right] \right\} - \left\{ E\left[\, D \mid Z^M = 0, (\mu(X), V) \in \mathcal{S}_{j,j+1} \,\right] \right\}$$

$$= \left\{ (j+1)\pi(\mathcal{C}^M_{j,j+1}) + (j)(1 - \pi(\mathcal{C}^M_{j,j+1})) \right\} - \left\{ (j)\pi(\mathcal{C}^M_{j,j+1}) + (j)(1 - \pi(\mathcal{C}^M_{j,j+1})) \right\}$$

$$= \left[(j+1) - j \right] \pi(\mathcal{C}^M_{j,j+1}) + \left[j - j \right] (1 - \pi(\mathcal{C}^M_{j,j+1}))$$

$$= \pi(\mathcal{C}^M_{j,j+1}).$$

3.3.1 Finding $\mathcal{S}^M_{j,j+1}$ and Standard Errors Algorithm

Only now dropping the Section 8 group from the analysis of outcomes, we estimate a j to $j+1$ transition-specific LATE as follows:

Step 1 Estimate the ordered choice model to obtain estimates of $\widehat{\mu}(X_i)$ and $\widehat{V}_i$ for all i.

Step 2 Find an area $\mathcal{A}_j \subset \mathcal{M} \times \mathbb{R}$ such that individuals with $(\widehat{\mu}(X_i), \widehat{V}_i) \in A_j$ select treatment level $D_i = j$ when in the control group

Step 3 Find an area $\mathcal{A}^M_{j,j+1} \subset \mathcal{M} \times \mathbb{R}$ such that movers (ie, households with $\tau^M_i = 1$) with $(\widehat{\mu}(X_i), \widehat{V}_i) \in \mathcal{A}^M_{j,j+1}$ will possibly select treatment level $D_i = j + 1$ when randomly assigned the MTO voucher

Step 4 Define the identification support set $\mathcal{S}^M_{j,j+1} \equiv \mathcal{A}_j \cap \mathcal{A}^M_{j,j+1}$

Step 5 Estimate the j to $j+1$ transition-specific LATE over $\mathcal{C}^M_{j,j+1}$ using the Wald estimator from Equation 14 applied to $\mathcal{S}^M_{j,j+1}$

Step 6 Repeat the following steps T times:

Step 6a Sample with replacement

Step 6b Repeat Step 1: Estimate the ordered choice model on the new sample

Step 6c Repeat Step 5: Calculate $\widehat{\triangle}_{j,j+1}^{LATE}\big((\mu(X),V) \in \mathcal{C}_{j,j+1}^{M}, \tau^{M} = 1\big)$ on the new sample where the set $\mathcal{S}_{j,j+1}^{M}$ maintains the definition determined in Step 4 for the original sample

Construct standard errors using the T parameter estimates.

4 Moving to Opportunity (MTO): Program Description and Data

4.1 MTO as a Legacy of the Civil Rights Movement

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)).[13]

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

---

[13]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)).

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.[14]

4.2 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.

4.2.1 Variables

We create a variable measuring neighborhood quality using 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 accounts for the most variation in 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.[15]

The resulting univariate index explains 69 percent of the variance of these neighborhood characteristics, and Table 1 reports 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, the magnitudes of which are similar to the magnitude of the coefficient for poverty for most variables. Finally, while poverty is negatively correlated with quality as expected, Figure 2 shows the existence near the 10 percent poverty cutoff of eligibility for using an MTO voucher, which is approximately the median of the national distribution, of neighborhoods of both very high and very low quality. This issue will be revisited when examining sorting patterns in Section 5.2.

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 or expected quality improvements, like new magnet schools in the baseline neighborhood, will reduce the benefit

---

[14]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.

[15]Given the importance of neighborhood violence (Anderson (1999), Aliprantis (2013c)), especially as documented in the context of MTO (Kling et al. (2005)), we would like to include measures of neighborhood violence and the rule of law. We do not include such measures in our index of quality because to the authors' knowledge variables comparable to those used in this analysis are not available.

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_i.

In addition to baseline neighborhood quality, other baseline characteristics of the MTO households used in this model are whether the respondent had family living in their 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. Site of residence is the only other observed characteristics included in X_i; when models were estimated with additional variables like number of children or residence in an early HOPE VI building, the coefficients on these other observables were all statistically insignificant.

Outcome variables for adults from the MTO Interim Evaluation include the labor market status of the adult at the time of the interim survey (ie, Two binary variables, one indicating labor force participation, the other indicating whether the adult was employed.), the self-reported total household income (all sources summed), the individual earnings in 2001 of the sample adult, receipt of Temporary Assistance for Needy Families (TANF) benefits, and the respondent's body mass index (BMI). Weights are used in constructing all estimates.[16]

4.2.2 Sample and Descriptive Statistics

The focus of this analysis is adults in the MTO Interim Evaluation sample. Figures 3 and 4 reproduce the broad result from Aliprantis (2013a) that MTO induced small shares of adult participants into high quality neighborhoods. These results are similar for another measure of neighborhood quality, school quality measured as the school's percentile ranking on state-level standardized tests, and are shown in Figure 5. We choose not to include children in our analysis for two reasons. First, careful analysis of program effects from MTO on children's outcomes has already been conducted (Sanbonmatsu et al. (2006), Kling et al. (2007)). Second, as shown in Figure 5 and investigated in depth in DeLuca and Rosenblatt (2010), MTO did not induce large changes in school quality.[17] Both intuition and previous findings in the literature suggest schools may be the most relevant "neighborhood" for children (Oreopoulos (2003)).

The ensuing analysis is focused on a sample that is restricted in two ways. The first restriction ensures that we are focusing on a relatively homogeneous population. To satisfy this restriction we drop all households living at baseline in a neighborhood above the tenth percentile of the national distribution of quality. Looking at Figures 4a and 6a, 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 and Boston, however, at

---

[16]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.

[17]The advent of school choice may be important for these results: Thirty percent of MTO control group children in Chicago and Los Angeles were attending magnet schools (Sanbonmatsu et al. (2006), p 684).

baseline a non-trivial share of program 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 represent a little under 15 percent of the interim evaluation sample, and are dropped from our estimation sample.

The second sample restriction facilitates the estimation of the ordered choice model. To satisfy this restriction we top-code neighborhood quality at the the median of the national distribution of quality in 2000. Figure 6c shows the final results of these restrictions, with Figures 6a and 6b showing the original sample and the one resulting from the first restriction alone, respectively.

The final estimation sample used in our analysis has a little under 3,100 adults (a little over 85 percent of the interim sample and a little under 75 percent of the original adult sample). Our sample represents "the other one percent:" At baseline, 67 percent of the estimation sample lived in a neighborhood whose quality was below the 1st percentile of the national distribution of neighborhood quality. There was enough mobility in the control group so that by 2002 this had decreased to 39 percent living in first percentile neighborhoods. On the other hand, though, the mobility of the control group was not extraordinary. By 2002, 81 percent of the sample in the control group lived in a neighborhood whose quality was less than the 10th percentile of the national distribution of neighborhood quality.

Although we include Boston in the analysis, it is important to note that it is clearly an outlier relative to the other MTO sites. In Figure 4a 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 at the time of the interim evaluation. We can see that the control group in Boston looks more like the experimental group in every other site. Boston comprised approximately 15 percent of the weighted observations in the estimation sample.

5 Empirical Specification and Estimation Results

5.1 Ordered Choice Model Specification

The marginal benefit of choosing to move from treatment level j to $j+1$ in the ordered choice model from Section 2.1.1 is specified to be

$$D_{ij}^* = \mu(X_i) + \gamma_j^S Z_i^S \tau_i^S + \gamma_j^M Z_i^M \tau_i^M - C_j - V_i.$$

We specify the components of D_{ij}^* to be

$$\mu(X_i) = \beta_1 X_1 + \cdots + \beta_8 X_8$$
$$\gamma_j^S = \Gamma_0^S + \Gamma_j^S$$
$$\gamma_j^M = \Gamma_0^M + \Gamma_j^M$$
$$C_j = \delta_0 + \delta_j,$$

with households' decisions to move with a voucher determined by the similarly-specified latent index models:

$$\tau_i^S = \mathbf{1}\{\ \mu^S(X_i) - V_i^S\ \geq\ 0\ \} \quad = \mathbf{1}\{\ \beta_1^S X_1 + \cdots + \beta_8^S X_8 - V_i^S\ \geq\ 0\ \},$$
$$\tau_i^M = \mathbf{1}\{\ \mu^M(X_i) - V_i^M\ \geq\ 0\ \} \quad = \mathbf{1}\{\ \beta_1^M X_1 + \cdots + \beta_8^M X_8 - V_i^M\ \geq\ 0\ \}.$$

Recall that the first four variables in X_i are baseline neighborhood quality, 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 at baseline. The final four variables in X_i are site indicators. Γ_0^S and Γ_0^M are themselves site-specific fixed effects that capture differences in factors like local labor and housing markets across sites. We do not attempt to explicitly model housing market prices since these are largely offset by the nature of the payment structure of the rental vouchers and project-based programs. Individuals pay 30 percent of their income towards rent in project-based units, and pay this same rent when using vouchers as long as the price of rent is not above the FMR. Further discussion of the interpretation of model parameters is provided in Appendix C.

Like Galiani et al. (2012), we interpret Z^S and Z^M as the random assignment of a potential reduction in the cost to accessing a higher quality neighborhood relative to staying in the baseline neighborhood. It is important to note that secular trends outside the control of the program might swamp this cost reduction. One can imagine changing costs to accessing higher quality neighborhoods due to changes in the local labor or housing markets, changes in school quality due to the provision of magnet schools, or simply an improvement in the quality of the baseline neighborhood.

We estimate the parameters of this ordered choice model via Maximum Likelihood using the log-likelihood function in Equation 7. We estimate $\widehat{V}_i$ as in Equation 13 after linearly interpolating between the $\widehat{\gamma}_j$ and $\widehat{C}_j$ as in Equations 9 and 10.

5.1.1 Definition of Treatment

When estimating the parameters of the ordered choice model, we define the discrete treatment levels by the intervals $\{\underline{q}_j^D, \overline{q}_j^D\}$ dividing the observed sample into its deciles as the time of the interim survey. When estimating treatment effect parameters we define the discrete treatment levels in terms of deciles of the national distribution:

$$D_i = \begin{cases} 1 & \text{if } q_i \in [\underline{q}_1^Y, \overline{q}_1^Y) = [0, 10); \\ \vdots & \qquad \vdots \\ 5 & \text{if } q_i \in [\underline{q}_5^Y, \overline{q}_5^Y) = [40, 50). \end{cases}$$

Section 3.2 discusses how we can use the interpolated $C(q)$, $\gamma^S(q)$, and $\gamma^M(q)$ together with the estimated parameters of the ordered choice model and values of V_i to go between these two defini-

tions.

We choose deciles to discretize neighborhood quality when investigating treatment effects not because we believe treatment should have an effect when crossing the particular thresholds of neighborhood quality used in this definition, 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.[18] 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 neighborhood effects using the MTO interim survey data. As the next Section shows, even this discretization leaves us with undesirably small sample sizes of compliers. As a result, the only LATEs we attempt to estimate are of moves between $j = 1$ and $j = 2$.

5.2 Empirical Evidence on Residential Sorting: Ordered Choice Model Estimation Results

Figure 7 shows the model fit as characterized by the estimated distributions of observed and unobserved variables. It is important to note that of the distributions shown in Figure 7, only Figure 7c is weighted. The estimated cost function $\widehat{C}(q)$ is shown in Figure 8 together with the cutoffs for treatment in the ordered choice model ($\{\underline{q}_j^D, \overline{q}_j^D\}$) and for effects on outcomes ($\{\underline{q}_j^Y, \overline{q}_j^Y\}$). The cost function is estimated to take the expected shape.

Several stylized facts emerge from the estimated cost reductions $\widehat{\gamma}^S(q)$ and $\widehat{\gamma}^M(q)$ displayed in Figure 9, allowing us to characterize the Section 8 and experimental MTO voucher programs. First, the MTO voucher was much more effective than the standard Section 8 voucher in getting complier households to access higher quality neighborhoods. Second, the effectiveness of both types of vouchers varied considerably by site. Vouchers represented the largest cost-reductions in Los Angeles (LA) and New York City (NYC), and represented the smallest cost-reductions in Baltimore and Boston. Chicago displays the largest gap in cost reduction between programs.

Figure 11 compares the neighborhood quality of mover and non-mover households after receiving vouchers. For Section 8 voucher holders in Chicago and Baltimore, we see that the neighborhood quality of movers is almost indistinguishable from that of non-movers, while for NYC and LA improvements in neighborhood quality are small.[19] Boston is the only site in which moving with a Section 8 voucher represented a significant improvement in neighborhood quality. Across all sites,

---

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

[19]See Ludwig et al. (2005) for a study of MTO's program effects in Baltimore.

MTO compliers are much more likely than Section 8 compliers to access neighborhoods with a quality index above the first decile.

The differences between the Section 8 and MTO voucher programs are evident when examining the actual mobility of program participants as shown in Figure 12. To begin, Figure 12a shows all program participants, color-coded by whether they lived in a neighborhood at the interim evaluation ranked in the first, second, third, or fourth decile of the national distribution of neighborhood quality. On the x-axis is the $\widehat{\mu}(X_i)$ of each household, and on the y-axis is the percentile of the household's unobserved determinant of selection in the absence of a program, $\widehat{u}_{Di} \equiv \Phi(\widehat{V}_i)$.

Since vouchers were randomly assigned in MTO, Figures 12b-12d illustrate counterfactual distributions of neighborhood quality under external manipulations to voucher type. For each household, given observed variables summarized by $\mu(X_i)$ and unobserved variables u_{Di}, these figures show the neighborhood quality households would select into under each setting of the vouchers.

Proceeding to Figure 12b, we can see that almost all of the control group remained in low-quality neighborhoods, most remaining in the first decile of neighborhood quality. Only households with very high observed factors $\widehat{\mu}(X_i)$ and very low unobserved cost factors $\widehat{u}_D$ managed to move to higher quality neighborhoods, even when defined as moving only to the second, third, or fourth deciles of the national distribution. In the bottom left corner of Figure 12c some dots have turned to blue, indicating that for low $\widehat{u}_D$ some households would be induced by the Section 8 voucher to move to a higher quality neighborhood. However, the similarity of Figures 12b and 12c is quite remarkable, suggesting that Section 8 vouchers were not very effective in getting households to move to higher quality neighborhoods.

Figure 12d shows that the MTO voucher was far and away more effective in getting complier households to move to higher quality neighborhoods than the ordinary Section 8 voucher. Unfortunately, we see that most of this mobility is from the first to the second decile of the national distribution of quality. Although it was still relatively rare, the MTO voucher did manage to induce some households to move into the third and fourth deciles of neighborhood quality.

The differences between the effects of Section 8 and MTO vouchers on selection into neighborhood quality are even more interesting when one considers that although 59 percent of our sample moved with the voucher when offered a Section 8 voucher, only 43 percent of MTO voucher recipients moved with their voucher. This suggests that simply asking whether recipients take-up a voucher and move when it is offered need not be the best way to judge the effectiveness of housing mobility vouchers. These results also reiterate that selection into neighborhoods of various quality levels is an inherently interesting phenomenon to study (Clampet-Lundquist and Massey (2008), Sampson (2008), Sampson (2012)).

Table 5 reports the estimated coefficients in $\widehat{\mu}(X)$ and the site fixed effects in $\widehat{\delta}_0$. There are large differences between sites: In the absence of any voucher program, program participants in Boston tended to live in much higher quality neighborhoods than their counterparts in the other MTO sites at the time of the interim evaluation. Baltimore and Chicago tended to be similar, while LA and NYC were the worst sites by far.

At the household level, we see that having teens in the household reduces the likelihood of moving with and without a voucher. It is possible that room occupancy restrictions according to age and gender of children may have made it harder for families with older children to find housing. Having no family in the baseline neighborhood makes households more likely to move. Living in a higher quality neighborhood at baseline increases the likelihood of moving without a voucher, but decreases moving prospects post MTO voucher assignment and counseling. The fact that the correlation between V and V^M is negative suggests that MTO was able to increase the likelihood of moving for households that had intrinsically larger unobserved costs to move.

5.3 Empirical Evidence on Neighborhood Effects: LATEs of Neighborhood Quality Estimation Results

5.3.1 What Effects are Identified?

Recalling the counterfactual distributions displayed in Figure 12, there is a range of values of $(\mu(X_i), U_{Di})$ in $\mathcal{M} \times [0, 1]$ for which households would be induced by receiving an MTO voucher to move from a $D = 1$ quality neighborhood to quality $D = 2$.[20] There is another range for which households would not move from $D = 1$ (those with high U_{Di}), and there are also ranges for which households would be induced to make other moves, such as from $D = 2$ to $D = 3$.

Due to the observed patterns of neighborhood selection displayed in Figure 12, we focus on identifying effects of moving from the first to the second decile of neighborhood quality.[21] Table 6 characterizes some of the changes in neighborhood characteristics that would typically accompany a move from $D = 1$ to $D = 2$. On average, the poverty rate would decline from 33 to 22 percent, BA attainment would go from 7 to 11 percent, the share of single-headed households would drop from 52 to 38 percent, and the female unemployment rate would drop from 16 to 10 percent. While these changes in neighborhood characteristics are non-trivial, it is worth pointing out that they are still far worse than the unconditional median neighborhood in the US in 2000, and changes of these magnitudes would have to occur several times to achieve the characteristics of the highest quality neighborhoods. As discussed in Aliprantis (2013a) and elsewhere, these are moves from the most extreme areas of the left tail of the distribution of quality to neighborhoods that are still within the left tail of quality.

This characterization of the neighborhood mobility induced by the MTO voucher might be surprising if we were to define neighborhood quality in terms of poverty alone. Recall Figure 2 and the discussion in Section 4.2: There exist low quality neighborhoods that are also low-poverty. Table 7 quantifies the prevalence of these neighborhoods in MTO states in 2000. The Table shows that while moving with an MTO voucher essentially ruled out neighborhoods in the lowest decile of quality, MTO voucher holders had many options for using their vouchers in low-

---

[20]We refer interchangeably to parameters and sets defined in terms of V_i and U_{Di}, where $U_{Di} = F_V(V_i)$.

[21]We have also estimated Average Causal Responses (ACRs) from Angrist and Imbens (1995) for subsets in which many possible moves are induced. The results are broadly consistent with our LATE estimates, but we omit the results for the sake of exposition.

quality neighborhoods that met the MTO voucher poverty restriction.

5.3.2 For Whom are Effects Identified?

We proceed graphically using Figure 13 to empirically implement the procedure from Section 3.3.1 to determine the support of $(\mu(X_i), U_{Di})$ for which LATEs of moving from $D = 1$ to $D = 2$ are identified. We can define the area $\mathcal{A}_1$ for which households would select into neighborhood quality $D = 1$ without any voucher (Figure 13h). We can also define the subset $\mathcal{A}_{1,2}^M \subset \mathcal{A}_1$ for which households would select into neighborhood quality $D = 2$ with an MTO voucher (Figure 13i). Defining $\mathcal{S}_{1,2}^M \equiv \mathcal{A}_1 \cap \mathcal{A}_{1,2}^M$, the identification support set is

$$\mathcal{S}_{1,2}^M \equiv \left\{ \quad \big(\mu(X_i), U_{Di}\big) \quad | \quad \mu(X_i) \in \big[-0.6, 0.4\big], \quad U_{Di} \in \big[0.43+0.30\mu(X_i), \ 0.68+0.15\mu(X_i)\big] \quad \right\},$$

which can be seen in Figures 13d-13f.

5.3.3 LATEs of Neighborhood Quality Estimation Results

Estimates of LATEs of neighborhood quality are reported for the subpopulation of compliers in $\mathcal{C}_{1,2}^M \subseteq \mathcal{S}_{1,2}^M$ in Table 8. All of these effects conform with the theory that living in higher quality neighborhoods improves adult labor market and health outcomes while decreasing receipt of welfare benefits. All of these point estimates are large, and even moving from the first to second levels of neighborhood quality alone is estimated to have statistically significant effects at the 10 percent level on adult labor market outcomes like labor force participation, employment, and health outcomes like Body Mass Index (BMI). The decrease in welfare receipt is statistically significant at the 14 percent level.

While some of the LATEs in Table 8 are imprecisely estimated, it is difficult to interpret the estimated effects as evidence against the theory that living in higher quality neighborhoods improves adult labor market and health outcomes.

5.3.4 Falsification Test: Outcomes for Non-Complier Households

To highlight the difference between our analysis and the program effect approach adopted in most of the literature on MTO (Ludwig et al. (2008), Ludwig et al. (2013b)), we now use Figure 13 to define a falsification set $\mathcal{F}_{1,2}^M$ for which households would remain in neighborhoods of quality $D = 1$ even if they were assigned an MTO voucher:

$$\mathcal{F}_{1,2}^M \equiv \left\{ \quad \big(\mu(X_i), U_{Di}\big) \quad | \quad U_{Di} \in \big[0.70, \ 1.0\big] \quad \right\}.$$

Households in the falsification set are determined not only by their values of $\big(\mu(X_i), U_{Di}\big)$, but also by the general cost function $C(q)$ *and* the value of the cost reduction function $\gamma^M(q)$ at various levels of quality q.

Neighborhood selection for the subpopulations in $\mathcal{S}_{1,2}^M$ and $\mathcal{F}_{1,2}^M$ is shown in the CDFs in Figure 14. Figure 14a shows that there is considerable variation in the neighborhood quality selected by the control and MTO voucher holders in the identification support set $\mathcal{S}_{1,2}^M$: No households in the control group selected into $D = 2$, while 37 percent of MTO voucher holders did. Neighborhood selection for households in the falsification set $\mathcal{F}_{1,2}^M$ was quite different: No households in the control group selected into $D = 2$, and only 2 percent of MTO voucher holders selected into $D = 2$.

The effects of the MTO program are then compared in Table 9 for households in the LATE identification support set $\mathcal{S}_{1,2}^M$ and for households in the falsification set $\mathcal{F}_{1,2}^M$. While receiving an MTO voucher resulted in large improvements to labor force participation rates for households with characteristics making them likely to move to a higher quality neighborhood when offered an MTO voucher (ie, those in $\mathcal{S}_{1,2}^M$), there was no effect on labor force participation for households receiving the MTO voucher who did not move to a higher quality neighborhood. Employment actually went down for those who did not move to higher quality neighborhoods, perhaps due to the disruptiveness of moving without the benefits of moving closer to jobs (Weinberg (2000)). And while welfare (TANF) receipt and BMI decreased for voucher recipients who did not move to higher quality neighborhoods, this effect was between multiple times and an order of magnitude larger for those who did move to a higher quality neighborhood.

This falsification test helps to illustrate that the effects of the MTO program are not interchangeable with effects from neighborhood quality. A list of assumptions must be made before translating effects of variation in MTO voucher assignment into effects of variation in neighborhood quality. Our assumptions have been stated explicitly in Sections 2-5; most assumptions in the MTO literature have been made implicitly (Aliprantis (2013a)).

6   Discussion

Aliprantis (2013a) clarifies the inability to learn about neighborhood effects from program effect estimates, given that compliance with random assignment often led to little or no improvements in neighborhood quality. Drawing on the work of HV and HUV, we specified a joint model of neighborhood quality choice and outcomes, allowing unobserved heterogeneity to influence moving decisions with and without voucher assignment. Since we observe the precise level of neighborhood quality selected by households, the structure of the ordered choice model allows us to estimate observed and unobserved components of choice in the absence of a voucher for all program participants. With households thus classified, we are able to obtain LATE estimates for the most common neighborhood quality transition induced by MTO: a move from the first to the second decile of neighborhood quality.

The fact that LATE estimates are large for health and labor market outcomes, despite the relatively small improvement in neighborhood quality, strongly suggests that MTO does provide evidence of positive neighborhood effects. However, these benefits are realized by a small percentage of voucher holders. Not only was the MTO voucher take-up rate low at 43 percent, but an upward

25

move in quality was even less likely. The LATEs we estimated pertain to less than 10 percent of the estimation subsample, which is itself a highly select group of MTO volunteers.

We interpret our estimates' lack of generality as evidence that policies ought to be carefully designed to achieve policy-makers' objectives. Despite the fact that households were more likely to move with Section 8 vouchers than MTO vouchers, this does not imply that Section 8 vouchers are preferable to MTO vouchers. Changes in neighborhood quality were much smaller for Section 8 movers than for MTO movers, and variation by site was large. Since only about a quarter of eligible households are currently able to obtain a Section 8 housing voucher (Sard and Fischer (2012)), an area for future research is understanding what types of restrictions on voucher use might optimize the extent to which households are able to realize positive neighborhood effects through the subsidy, and which of these restrictions might be feasible to implement (McClure (2010)).

A final consideration when interpreting our results is that programs and neighborhood changes differentially impacting the treatment and control groups will interfere with the ability to identify neighborhood effects using voucher assignment as an instrument. Ideally, the voucher assignment would induce a change in the cost of moving holding all else equal. However, households may have responded to their group assignment, and baseline neighborhood conditions may well have changed during the multiple-year period between the decision to move and the time of the interim evaluation when outcomes were measured. For example, households assigned to the control group might have responded by applying for Section 8 vouchers on their own outside of the MTO program (Orr et al. (2003)). And according to de Souza Briggs et al. (2010), during the implementation of MTO Jobs-Plus saturated public housing developments with state-of-the-art employment, training, and child care services, while providing rent incentives to encourage employment. The US also enacted major welfare reform legislation in August 1996, precisely while MTO vouchers were being assigned (Blank (2002)). That these and other factors are not explicitly modeled in our analysis points to the methodological limitations of conducting controlled, and not just randomized, experiments in social settings (Deaton (2009), Aliprantis (2013b)).

7   Conclusion

Because households endogenously sort into neighborhoods, identifying causal effects of neighborhood environments has proven to be a substantial challenge. Researchers have sought to identify neighborhood effects using the exogenous variation in neighborhoods induced by housing mobility programs. The Moving to Opportunity (MTO) housing mobility experiment gave households living in high-poverty neighborhoods in five US cities the ability to enter a lottery for housing vouchers to be used in low-poverty neighborhoods. In a tremendous disappointment, effects from MTO were very small, and this lack of effects has been interpreted as evidence against the theory that neighborhood characteristics influence individuals' outcomes.

Aliprantis (2013a) clarifies that program effect estimates are less informative about neighborhood effects than currently appreciated in the literature, given that compliance with random as-

signment often led to little or no improvements in neighborhood quality. This paper abandoned the approach of learning about neighborhood effects from ITT and TOT effects of the MTO program. Rather, we proposed and implemented a new strategy for identifying LATEs of neighborhood quality using the random variation in quality induced by the program together with an ordered choice model of households' selection into neighborhood quality.

We found that moving to a higher quality neighborhood had large, positive effects on employment, labor force participation, and BMI. Although effects on household income, individual earnings, and welfare receipt were not statistically significant, these effects were also estimated to be large and positive. Due to the limited changes in neighborhood quality induced by MTO, these LATE estimates pertain to less than 10 percent of our estimation sample. We found no evidence from MTO against the theory that increasing neighborhood quality improves adult outcomes.

References

Aliprantis, D. (2012). Redshirting, compulsory schooling laws, and educational attainment. *Journal of Educational and Behavioral Statistics 37*(2), 316–338.

Aliprantis, D. (2013a). Assessing the evidence on neighborhood effects from Moving to Opportunity. *Federal Reserve Bank of Cleveland Working Paper 12-33r*.

Aliprantis, D. (2013b). Covariates and causal effects: The problem of context. *Federal Reserve Bank of Cleveland Working Paper 13-10*.

Aliprantis, D. (2013c). Human capital in the inner city. *Federal Reserve Bank of Cleveland Working Paper 13-02*.

Aliprantis, D. and D. Carroll (2013). Neighborhood dynamics and the distribution of opportunity. *Federal Reserve Bank of Cleveland Working Paper 12-12r*.

Amemiya, T. (1985). *Advanced Econometrics*. Harvard University Press.

Anderson, E. (1999). *Code of the Street: Decency, Violence, and the Moral Life of the Inner City*. W. W. Norton and Company.

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.

Badel, A. (2010). Understanding permanent black-white earnings inequality. *Federal Reserve Bank of St. Louis Working Paper 2010-047B*.

Badev, A. (2012). Discrete games in endogenous networks: Theory and policy. *Mimeo., University of Pennsylvania*.

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.

Blank, R. M. (2002). Evaluating welfare reform in the United States. *Journal of Economic Literature 40*(4), 1105–1166.

Brock, W. and S. Durlauf (2007). Identification of binary choice models with social interactions. *Journal of Econometrics 140*(1), 52–75.

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

Carneiro, P. and S. Lee (2009). Estimating distributions of potential outcomes using local instrumental variables with an application to changes in college enrollment and wage inequality. *Journal of Econometrics 149*(2), 191–208.

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.

de Souza Briggs, X., S. J. Popkin, and J. Goering (2010). *Moving to Opportunity: The Story of an American Experiment to Fight Ghetto Poverty*. Oxford University Press.

Deaton, A. (2009). Instruments of development: Randomization in the tropics, and the search for the elusive keys to economic development. *Mimeo, Princeton University*.

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.

DeLuca, S. and P. Rosenblatt (2010). Does moving to better neighborhoods lead to better schooling opportunities? Parental school choice in an experimental housing voucher program. *Teachers College Record 112*(5), 1443–1491.

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).

Goldsmith-Pinkham, P. and G. Imbens (2011). Social networks and the identification of peer effects. *Mimeo., Harvard University*.

Greene, W., M. N. Harris, B. Hollingsworth, and P. Maitra (2008). A bivariate latent class correlated generalized ordered probit model with an application to modeling observed obesity levels. *Mimeo., New York University*.

Heckman, J. J. (2008). Econometric Causality. *International Statistical Review 76*(1), 1–27.

Heckman, J. J. (2010). Building bridges between structural and program evaluation approaches to evaluating policy. *Journal of the Economic Literature 48*(2), 356–398.

Heckman, J. J., N. Hohmann, J. Smith, and M. Khoo (2000). Substitution and dropout bias in social experiments: A study of an influential social experiment. *The Quarterly Journal of Economics 115*(2), 651–694.

Heckman, J. J. and S. Urzúa (2010). Comparing IV with structural models: What simple IV can and cannot identify. *Journal of Econometrics 156*(1), 27–37.

Heckman, J. J., 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 (1999). Local instrumental variables and latent variable models for identifying and bounding treatment effects. *Proceedings of the National Academy of Sciences 96*(8), 4730–34.

Heckman, J. J. and E. Vytlacil (2001). Policy-relevant treatment effects. *The American Economic Review 91*(2), 107–111.

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. (2010). Better LATE than nothing: Some comments on Deaton (2009) and Heckman and Urzua (2009). *Journal of the Economic Literature 48*(2), 399–423.

Imbens, G. W. and J. D. Angrist (1994). Identification and estimation of Local Average Treatment Effects. *Econometrica 62*(2), 467–475.

Jargowsky, P. A. (1997). *Poverty and Place: Ghettos, Barrios, and the American City.* New York: Russell Sage Foundation.

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 (2005). Bullets don't got no name: Consequences of fear in the ghetto. In T. S. Weisner (Ed.), *Discovering Successful Pathways in Children's Development: Mixed Methods in the Study of Childhood and Family Life*, pp. 243–281. University of Chicago Press.

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

Kneebone, E., C. Nadeau, and A. Berube (2011, November). *The Re-Emergence of Concentrated Poverty: Metropolitan Trends in the 2000s.* Washington, DC: The Brookings Institution.

Ludwig, J., G. J. Duncan, L. A. Gennetian, L. F. Katz, R. C. Kessler, J. R. Kling, and L. Sanbonmatsu (2013a). Long-term nieghborhood effects on low-income families: Evidence from Moving to Opportunity. *American Economic Review 103*(3), 226–231.

Ludwig, J., G. J. Duncan, L. A. Gennetian, L. F. Katz, R. C. Kessler, J. R. Kling, and L. Sanbonmatsu (2013b). Long-term nieghborhood effects on low-income families: Evidence from Moving to Opportunity. *NBER Working Paper 18772*.

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

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.

Manski, C. F. (1993). Identification of endogenous social effects: The reflection problem. *The Review of Economic Studies 60*(3), pp. 531–542.

Manski, C. F. (2013). Identification of treatment response with social interactions. *The Econometrics Journal 16*(1), S1–S23.

Massey, D. and N. Denton (1993). *American Apartheid: Segregation and the Making of the Underclass.* Cambridge: Harvard University Press.

McClure, K. (2010). The prospects for guiding housing choice voucher households to high-opportunity neighborhoods. *Cityscape 12*(3), 101–122.

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.

Munnell, A. H., G. M. B. Tootell, L. E. Browne, and J. McEneaney (1996). Mortgage lending in Boston: Interpreting HMDA data. *The American Economic Review 86*(1), 25–53.

Oreopoulos, P. (2003). The long-run consequences of living in a poor neighborhood. *The Quarterly Journal of Economics 118*(4), pp. 1533–1575.

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.

Pearl, J. (2009a). Causal inference in statistics: An overview. *Statistical Surveys 3*, 96–146.

Pearl, J. (2009b). *Causality: Models, Reasoning and Inference* (2nd ed.). Cambridge University Press.

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

Quigley, J. M. and S. Raphael (2008). Neighborhoods, economic self-sufficiency, and the MTO program. *Brookings-Wharton Papers on Urban Affairs 8*(1), 1–46.

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.

Ross, S. L. and G. Tootell (2004). Redlining, the Community Reinvestment Act, and private mortgage insurance. *Journal of Urban Economics 55*(2), 278–297.

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

Sampson, R. J. (2012). *Great American City: Chicago and the Enduring Neighborhood Effect.* University of Chicago Press.

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.

Sard, B. and W. Fischer (2012, July). Renter's tax credit would promote equity and advance balanced housing policy. *Center on Budget and Policy Priorities Report*.

Sobel, M. E. (2006). What do randomized studies of housing mobility demonstrate?: Causal inference in the face of interference. *Journal of the American Statistical Association 101*(476), 1398–1407.

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. (2002). Independence, monotonicity, and latent index models: An equivalence result. *Econometrica 70*(1), 331–341.

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.

Weinberg, B. A. (2000). Black residential centralization and the spatial mismatch hypothesis. *Journal of Urban Economics 48*(1), 110–134.

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

Woodward, J. (2000). Explanation and invariance in the special sciences. *The British Journal for the Philosophy of Science*.

Figures
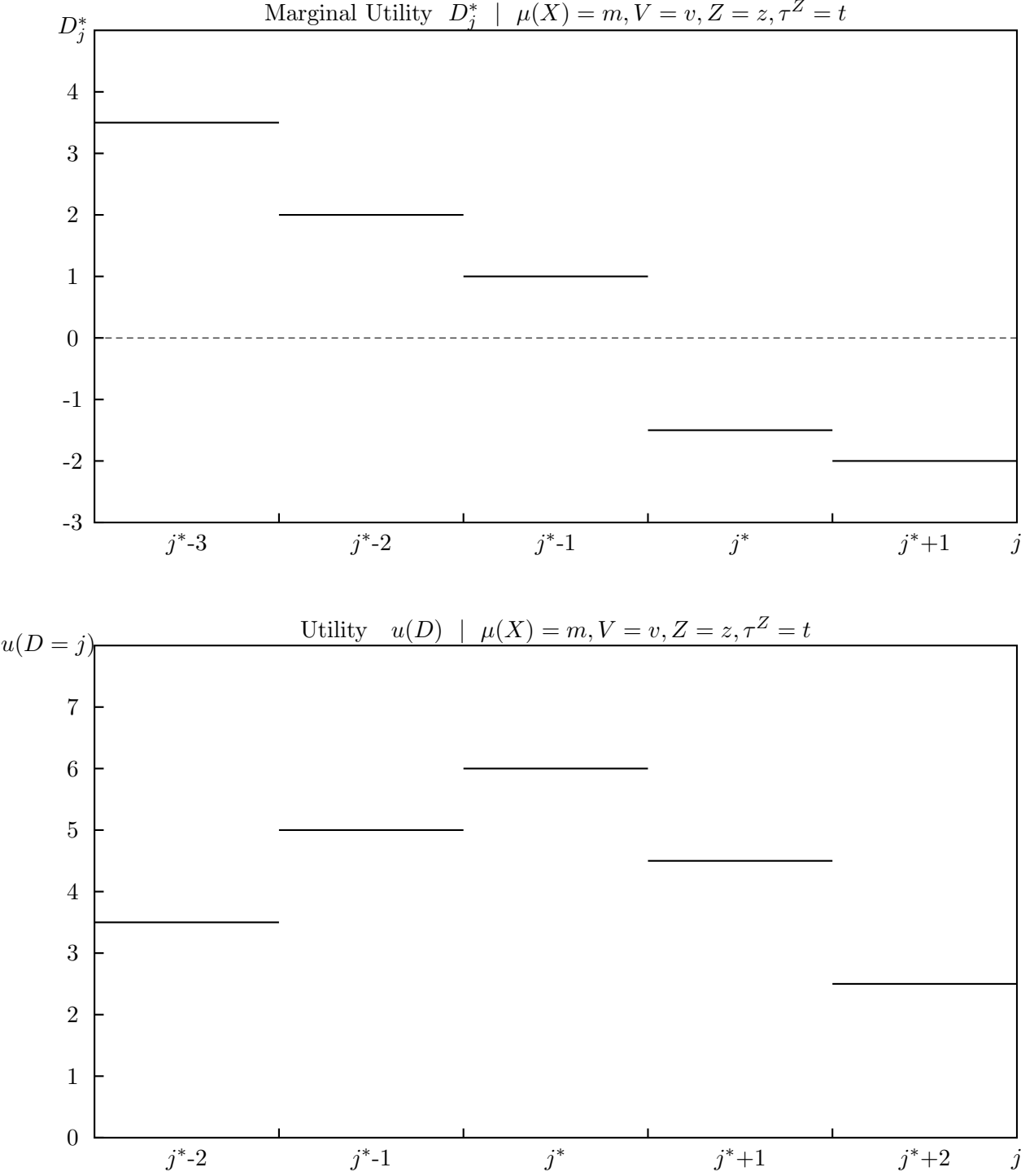
Figure 1: The First Stage Ordered Choice Model
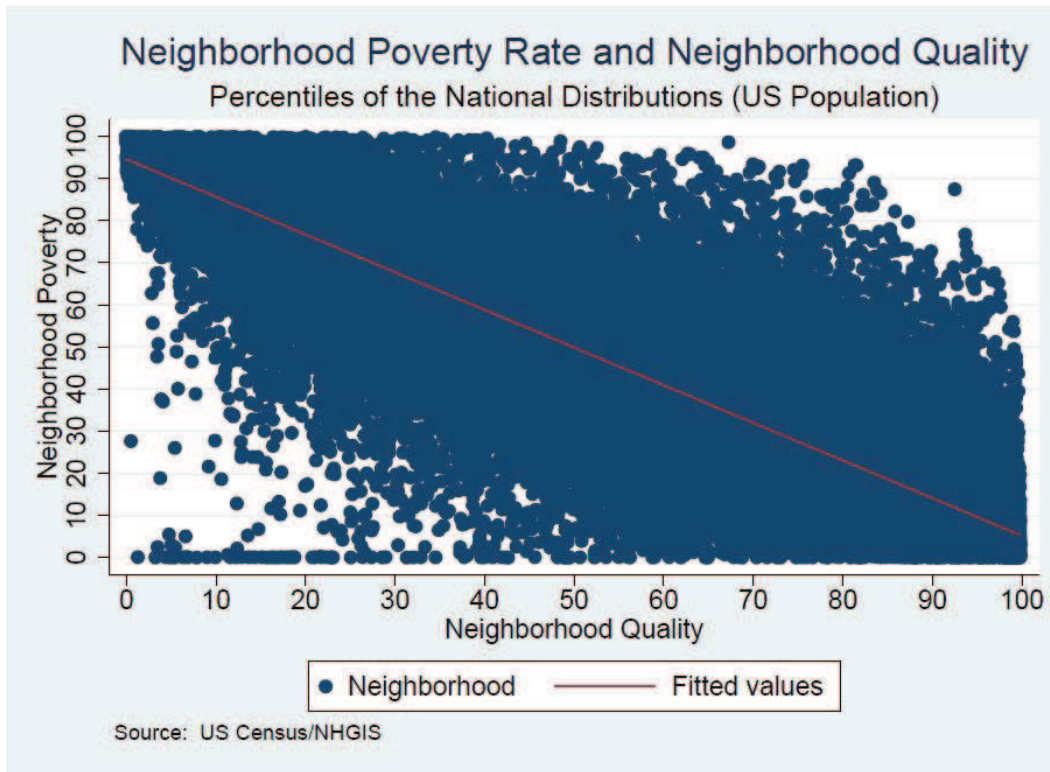
Figure 2: Neighborhood Poverty and Neighborhood Quality
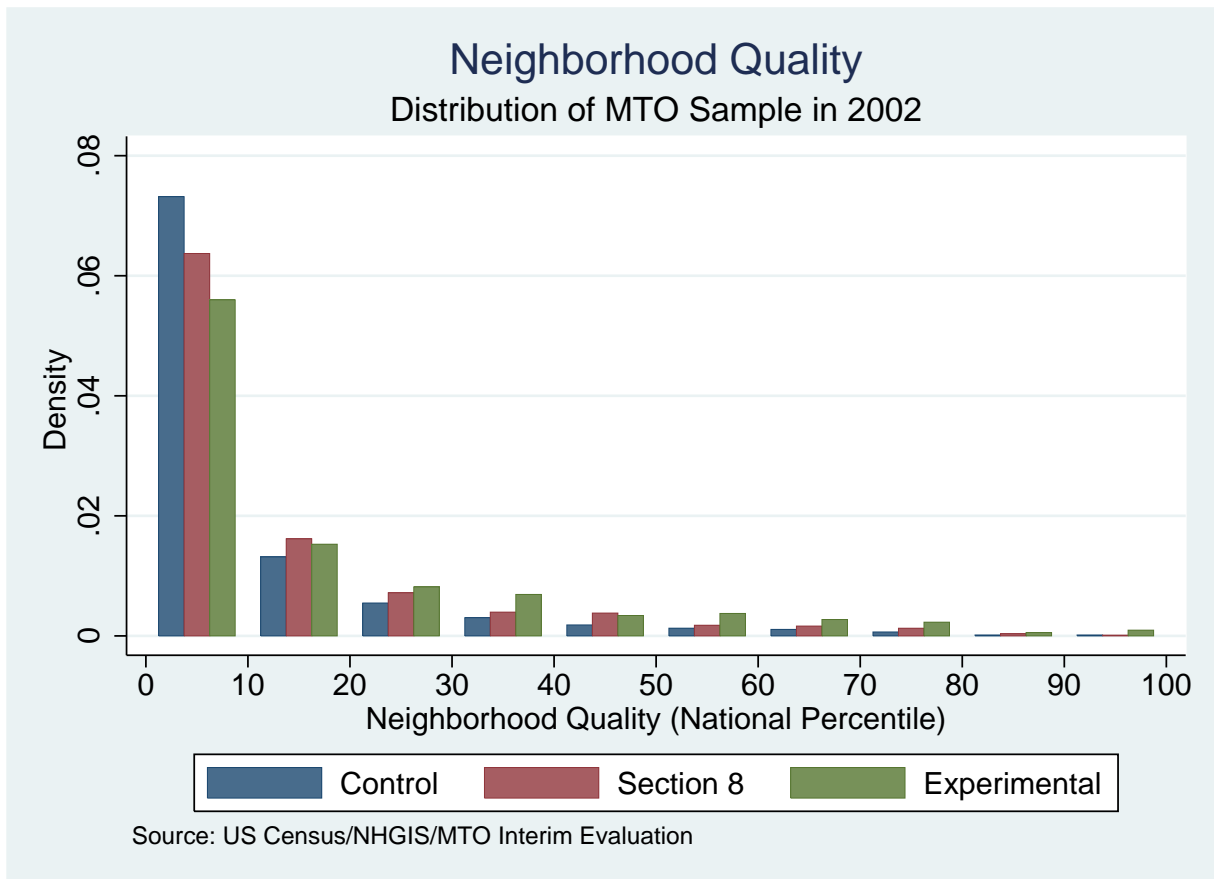
Figure 3: Neighborhood Quality of MTO Adults
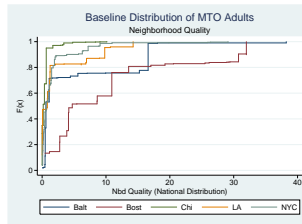
(a) Baseline Quality
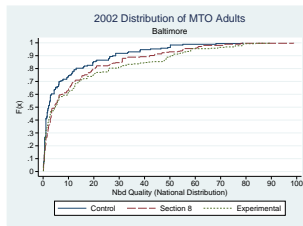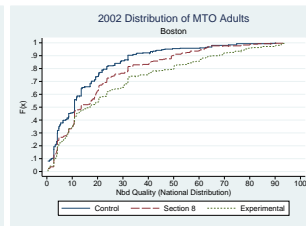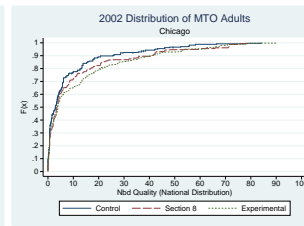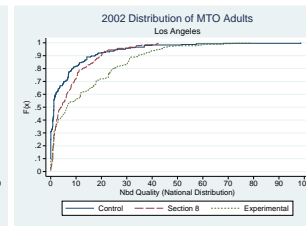
(b) Baltimore (c) Boston (d) Chicago (e) LA (f) NYC
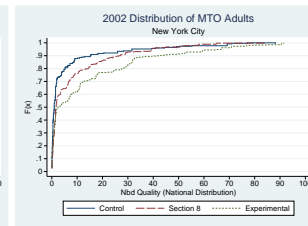
Figure 4: Neighborhood Quality by Site
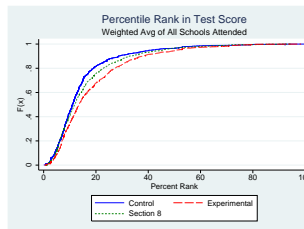
(a) All Sites Pooled
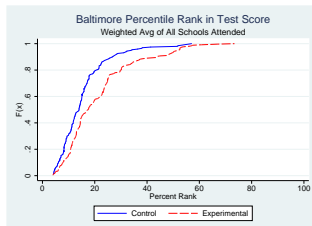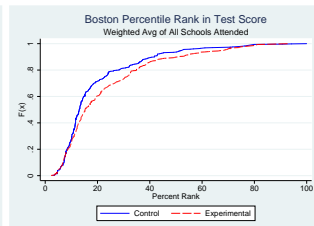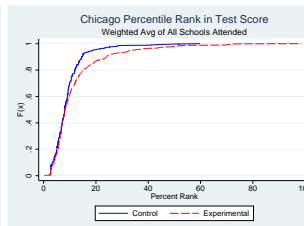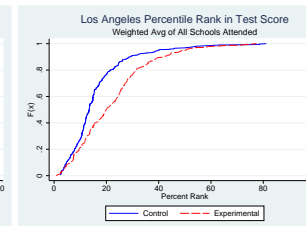
(b) Baltimore (c) Boston (d) Chicago (e) Los Angeles (f) New York City
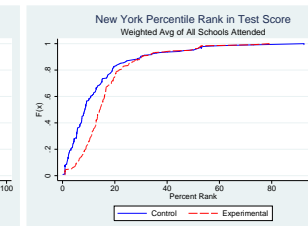
Figure 5: School Ranking on State Tests, Weighted Average Percentile over all Schools Attended (by Site)
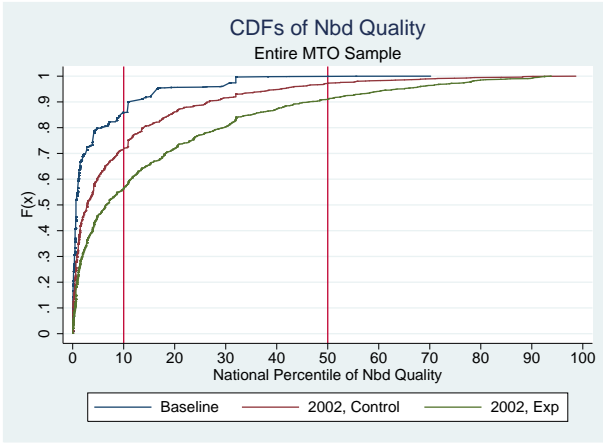
(a) MTO Distributions
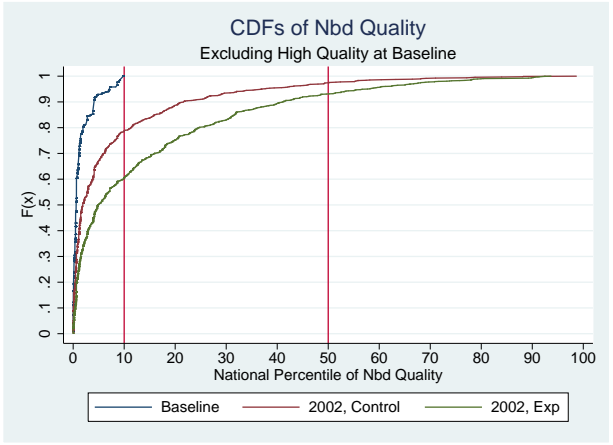
(b) Excluding Households Living in "High-Quality" Neighborhoods at Baseline
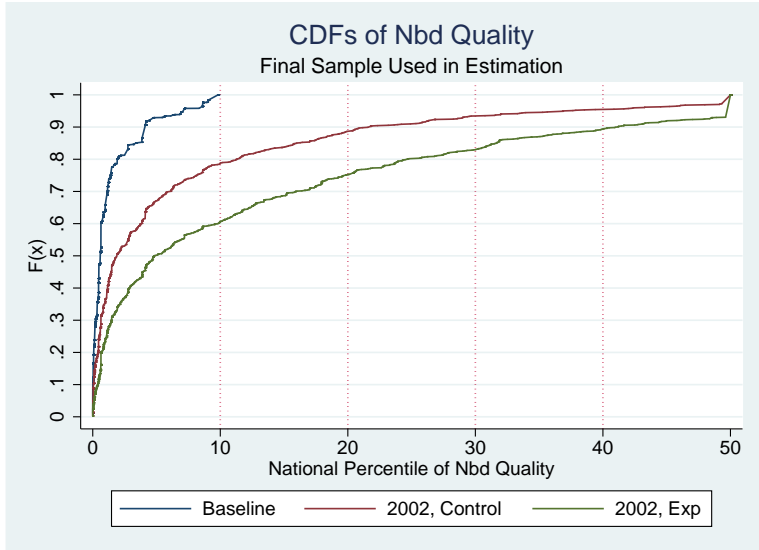
(c) Final Sample Used in Estimation

Figure 6: Neighborhood Quality
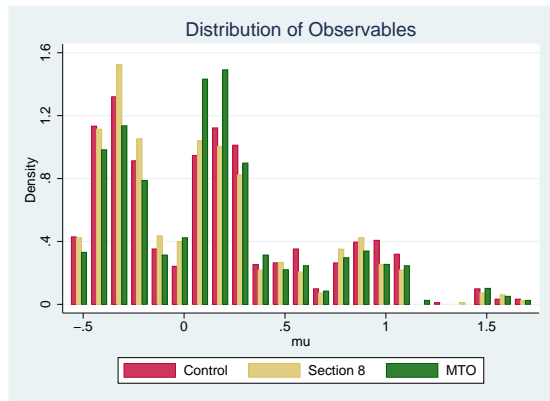
(a) Observables
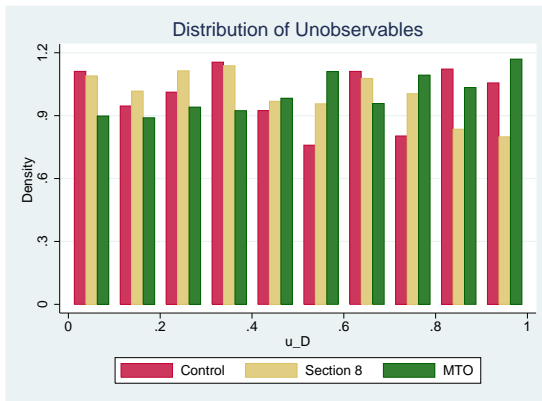
(b) $\widehat{U}_D | Z$
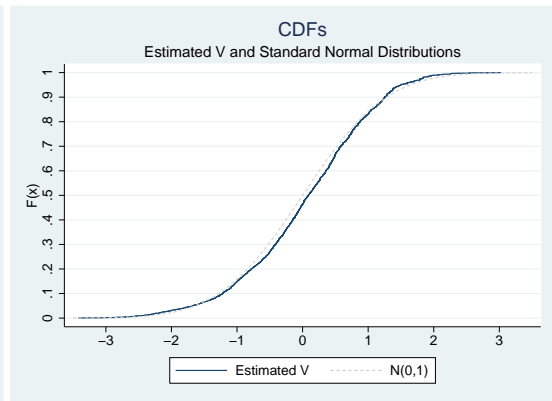
(c) $\widehat{F}_V$ and Φ
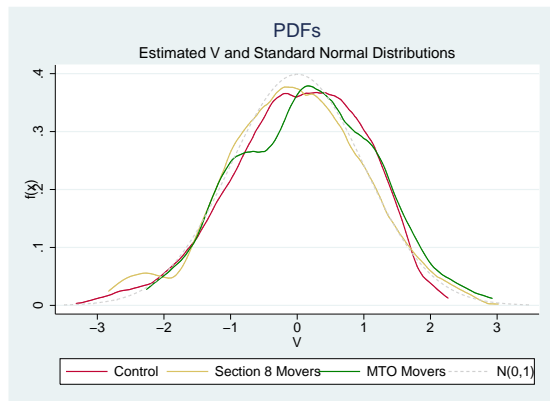
(d) V_i For Control Group and Voucher Movers
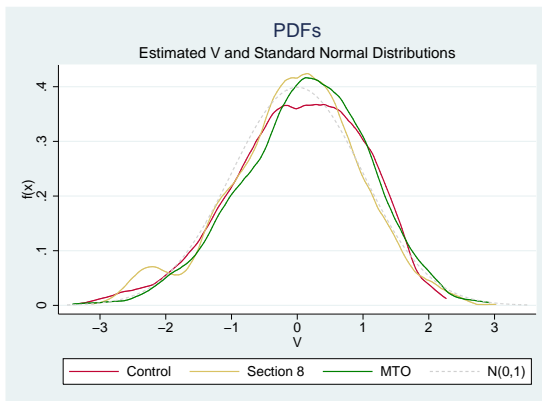
(e) V_i For Control Group and Voucher Holders
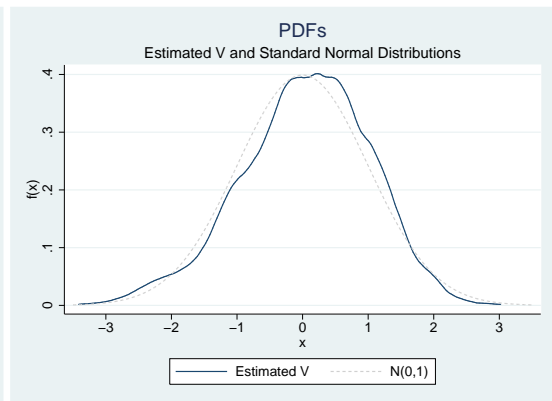
(f) V_i For Sample

Figure 7: Model Fit: The Distributions of Observables and Unobservables
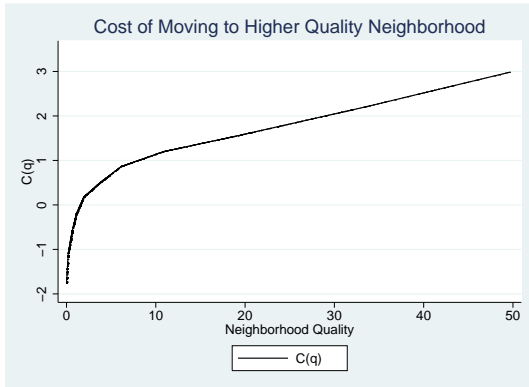
(a) C(q)
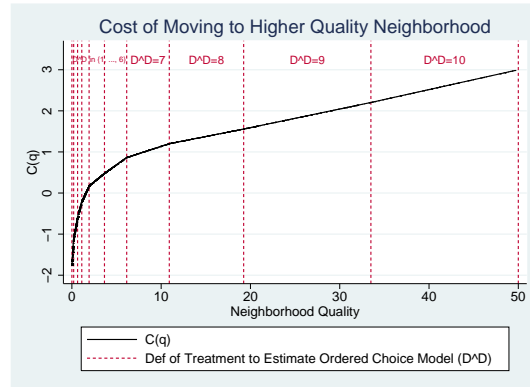
(b) C(q) with $\{\underline{q}_j^D, \overline{q}_j^D\}$
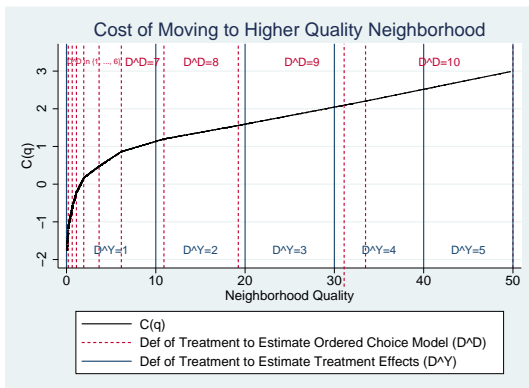
(c) C(q) with $\{\underline{q}_j^D, \overline{q}_j^D\}$ and $\{\underline{q}_j^Y, \overline{q}_j^Y\}$
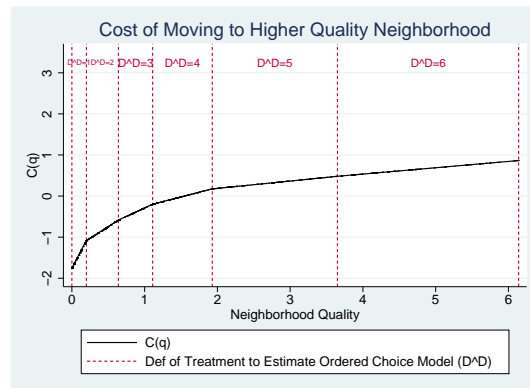
(d) C(q) with $\{\underline{q}_j^D, \overline{q}_j^D\}$

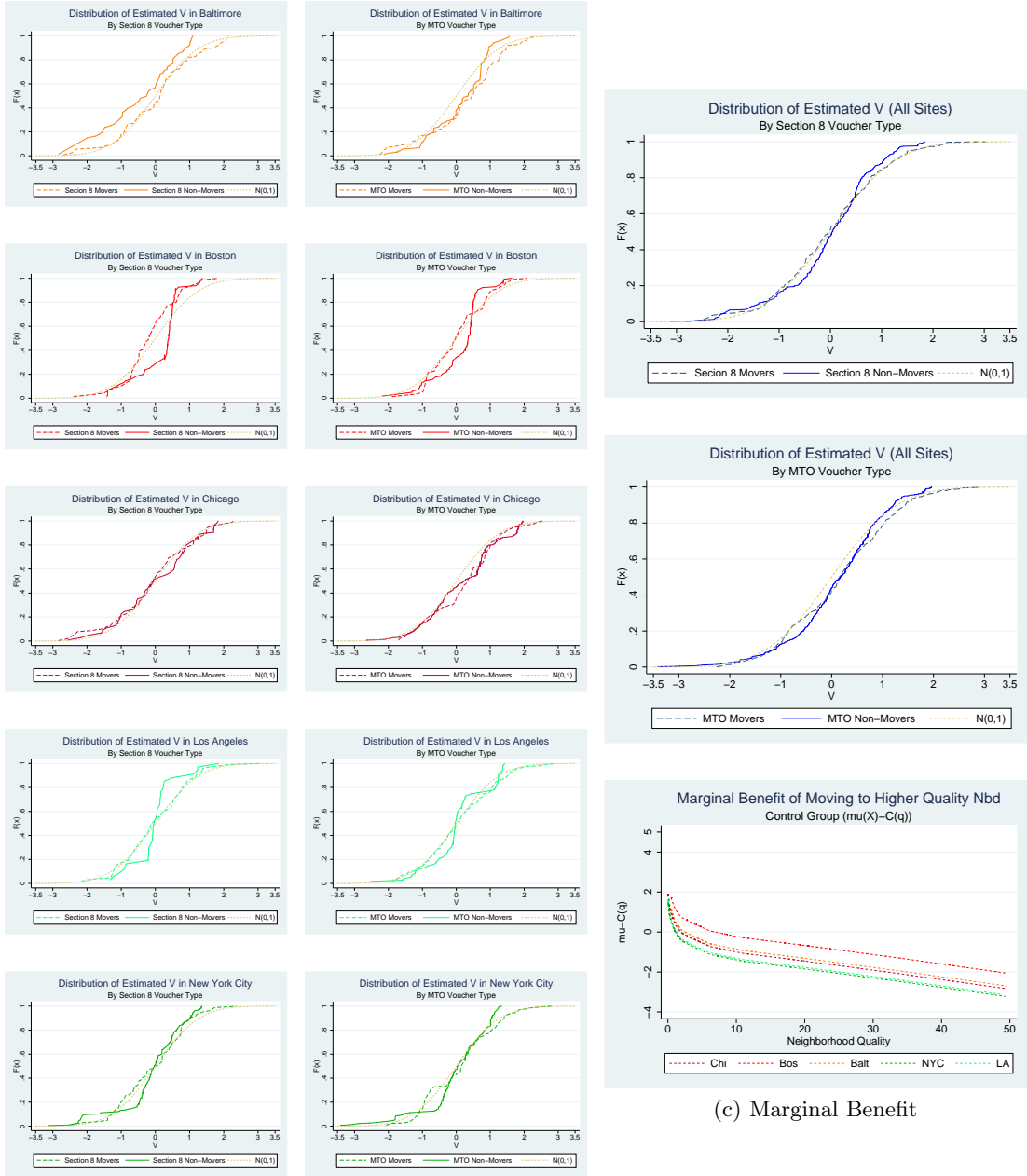Figure 8: Cost Function

(a) By Site

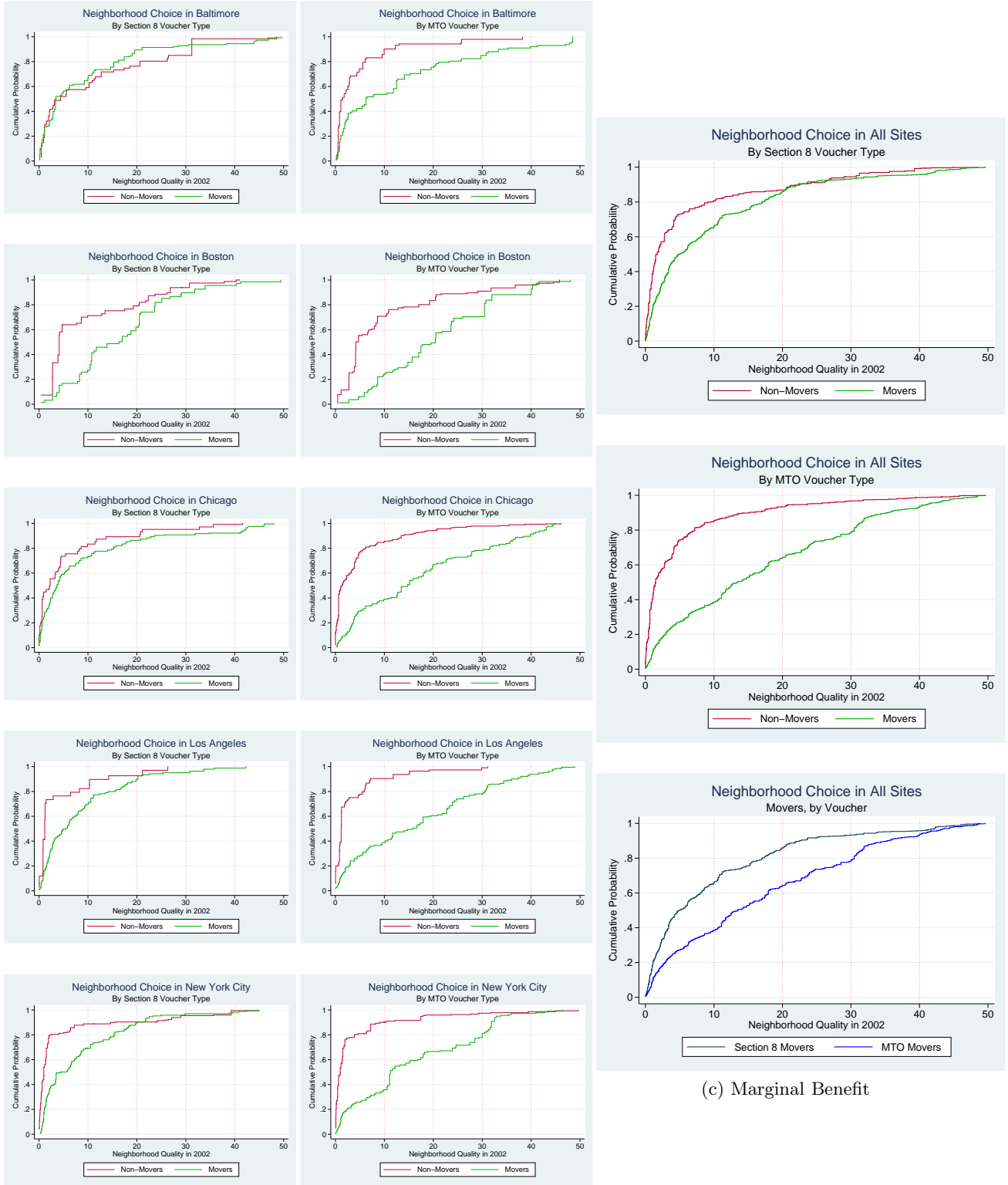(b) By Voucher Type
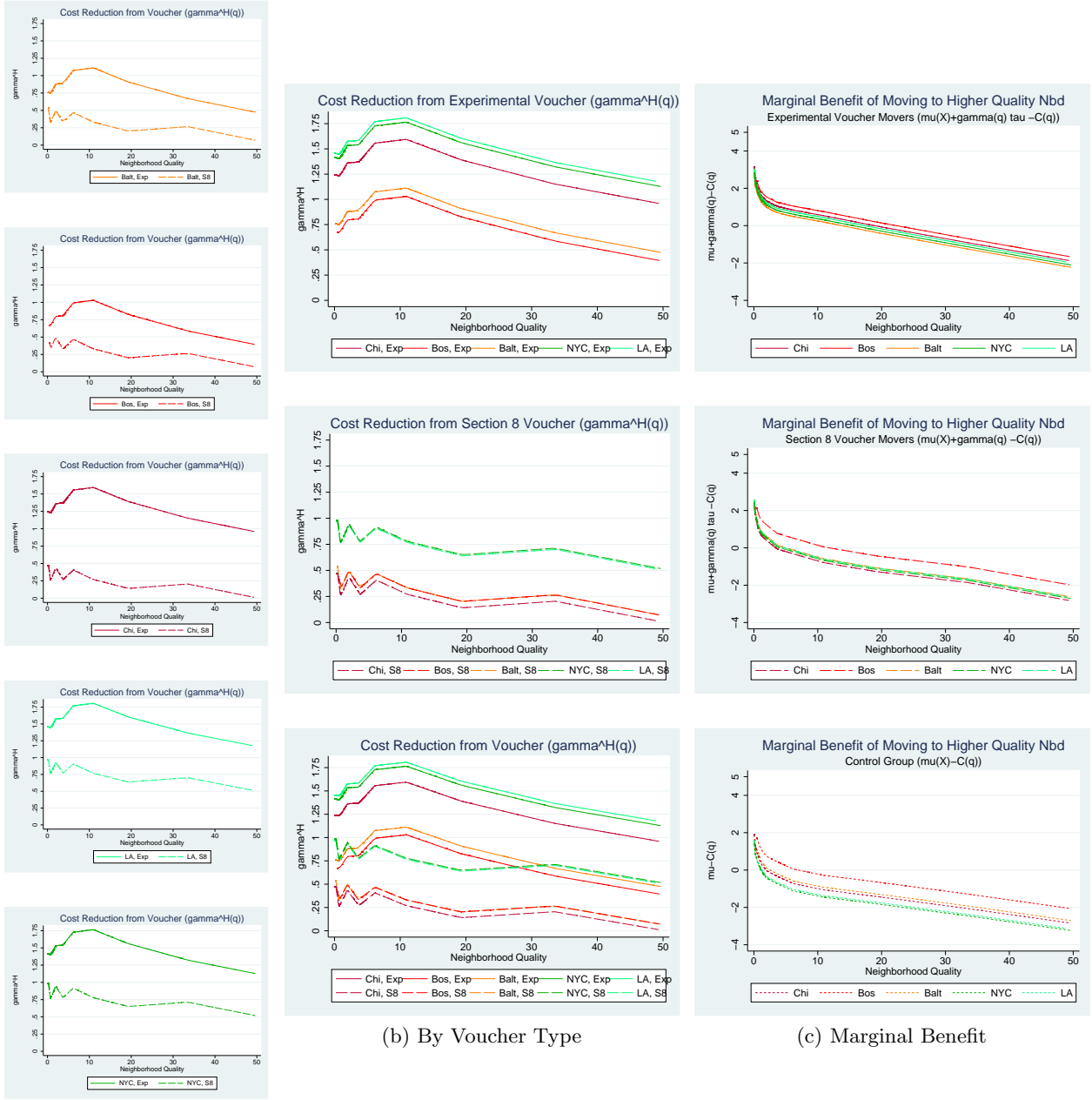
(c) Marginal Benefit

Figure 9: Cost Reduction from Vouchers and Marginal Benefit Functions at Average $\widehat{\mu}(X)$

(a) By Site (b) By Voucher Type (c) Marginal Benefit

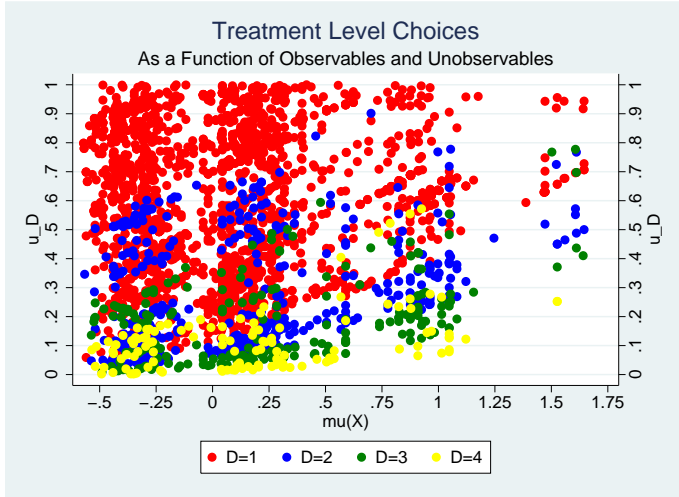Figure 10: Cost Reduction from Vouchers and Marginal Benefit Functions at Average μ
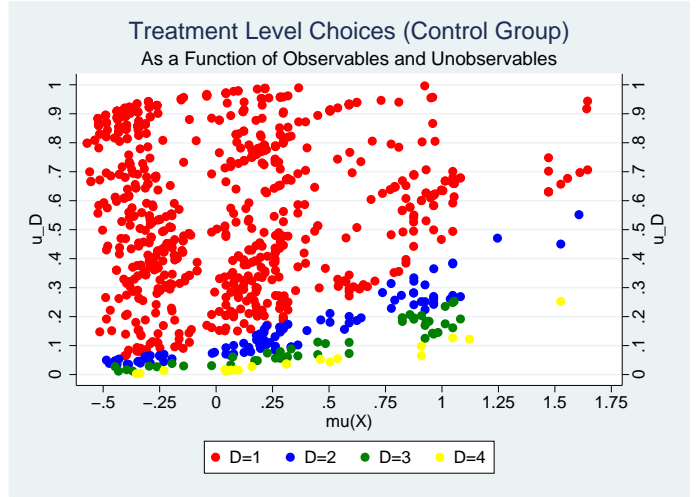
(a) By Site

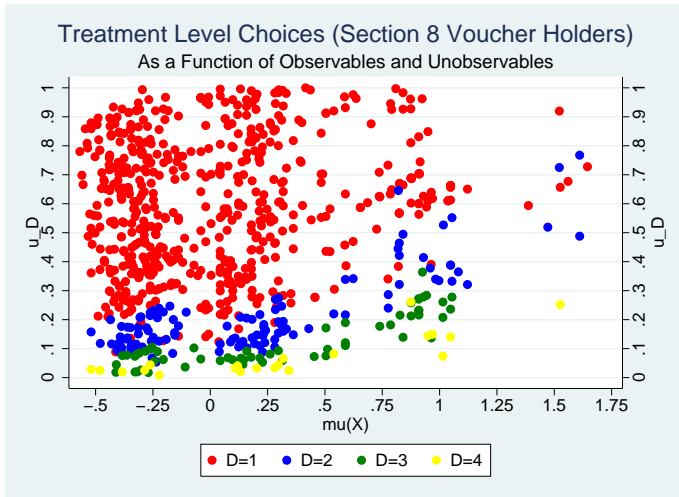(b) By Voucher Type

(c) Marginal Benefit

Figure 11: Selection into Neighborhood Quality by Voucher and Type
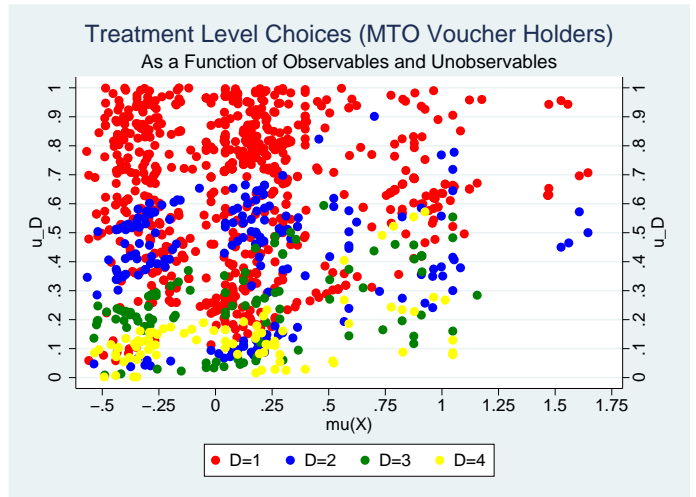
41

(a) All Groups

(b) Control Group

(c) Section 8 Voucher Holders

(d) MTO Voucher Holders

Figure 12: Selection into Treatment

(a) All Groups   (b) Control Group and $\mathcal{A}_1$ (c) MTO Voucher Holders and $\mathcal{A}_{1,2}^M$

(d) All Groups and $\mathcal{S}_{1,2}^M$ (e) Control Group and $\mathcal{S}_{1,2}^M$ (f) MTO Voucher Holders and $\mathcal{S}_{1,2}^M$

(g) All Groups   (h) Control Group and $\mathcal{F}_{1,2}^M$ (i) MTO Voucher Holders and $\mathcal{F}_{1,2}^M$
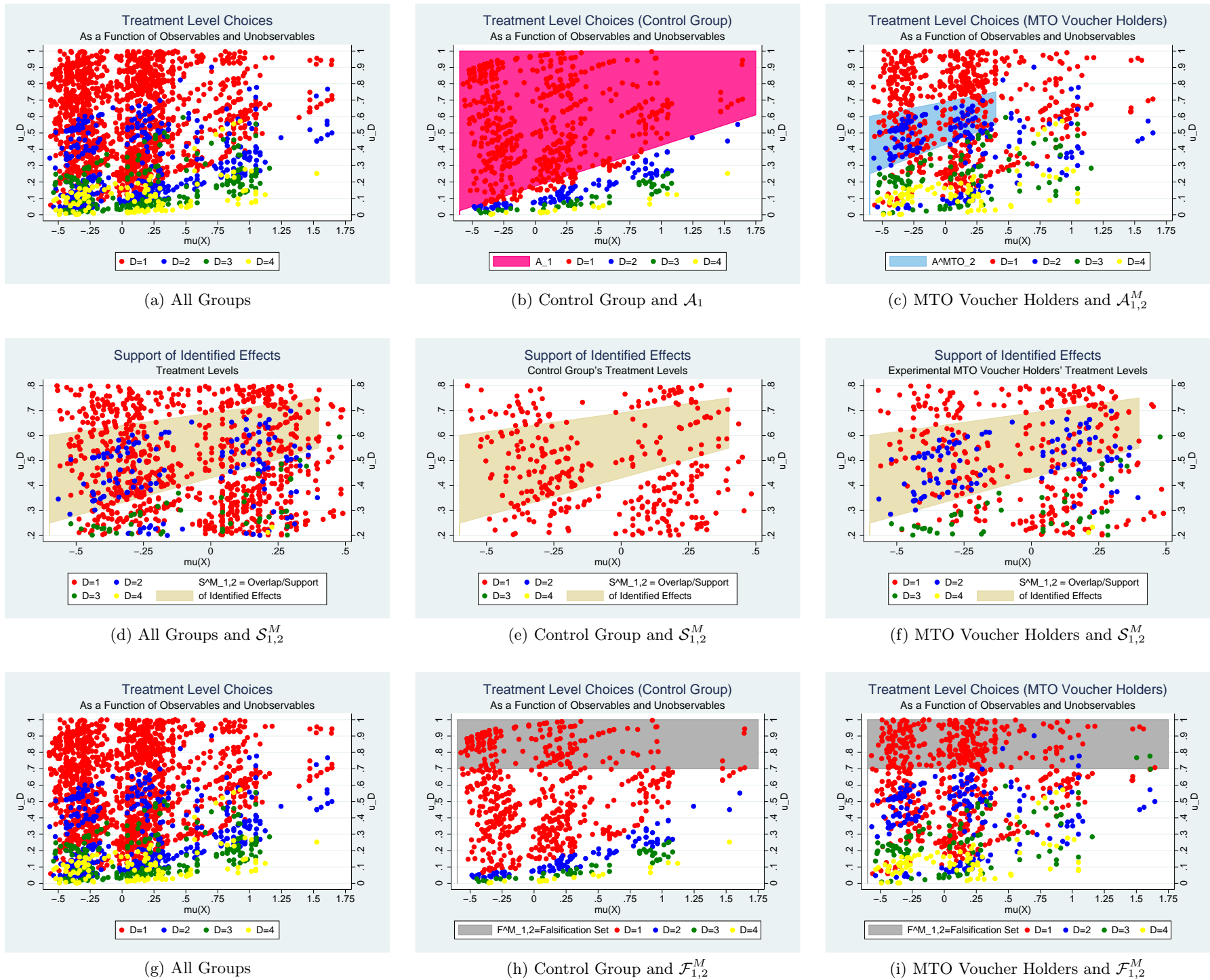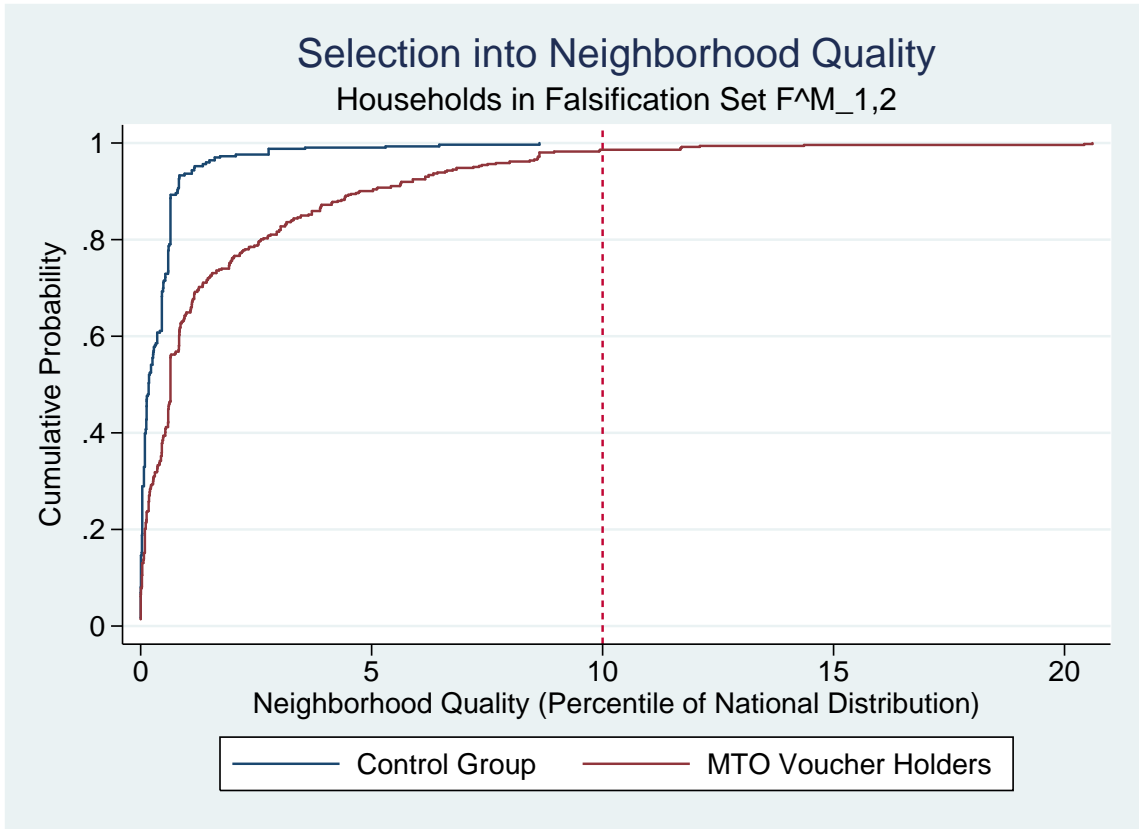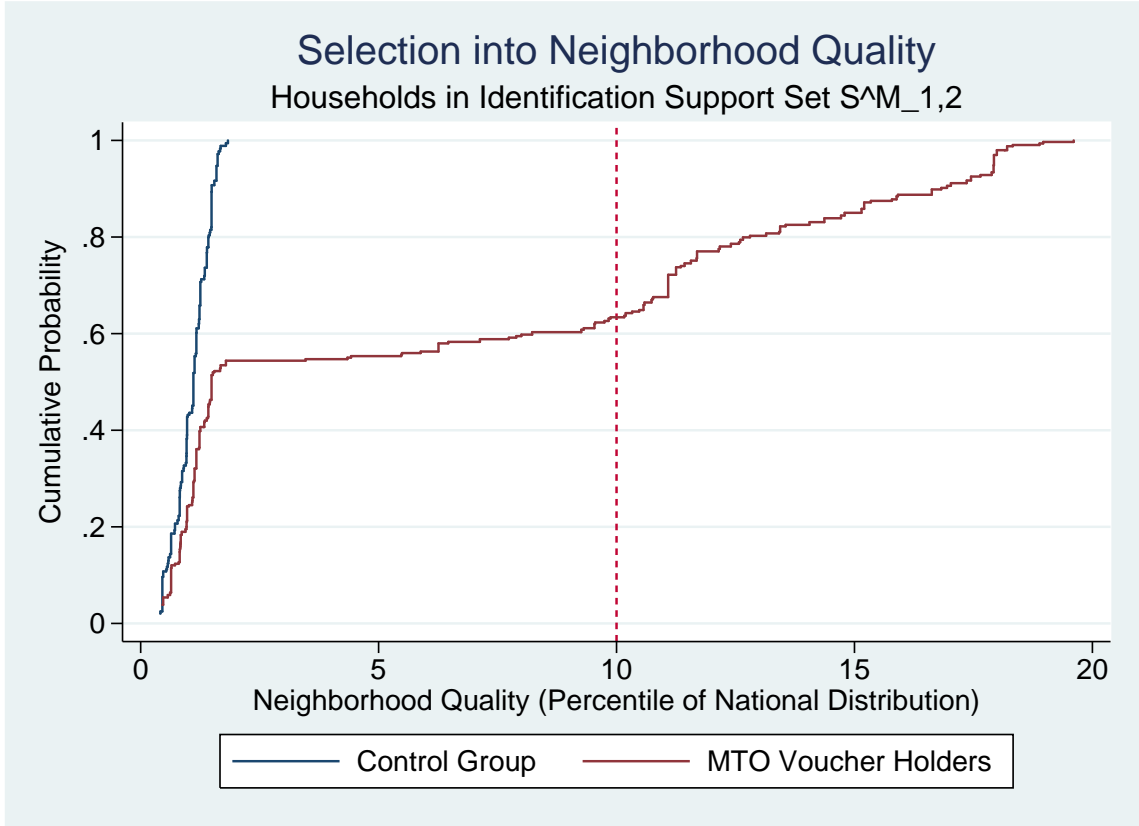
Figure 13: Selection into Treatment, , Counterfactual Areas $\mathcal{A}_1$ and $\mathcal{A}_{1,2}^M$, and Identification Support Set $\mathcal{S}_{1,2}^M$

(a) LATE Identification Support Set $\mathcal{S}_{1,2}^M$

(b) Falsification Set $\mathcal{F}_{1,2}^M$

Figure 14: Selection into Neighborhood Quality for Various Subpopulations

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

Site	Mean			Median		
	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: Share of Subsample at Each Site (%)

	Estimation Sample	$\mathcal{S}_{1,2}^{M}$	$\mathcal{F}_{1,2}^{M}$	$\mathcal{ACR}_{1,2}^{M}$
Share of Subsample (%)				
Baltimore	13.0	11.8	14.2	19.1
Boston	14.6	0.3	9.3	0.1
Chicago	26.4	15.6	33.3	40.0
Los Angeles	18.2	26.1	18.5	16.3
New York City	27.8	46.2	24.7	24.5

Table 5: Ordered Choice Model Parameter Estimates

X_k and $\mathbf{V}$	$\widehat{\beta}_k$	$\widehat{\beta}_k^S$	$\widehat{\beta}_k^M$
Baseline Characteristics			
Teens in HH	−0.08	−0.48	−0.39
	(0.05)	(0.10)	(0.09)
Family in Nbd	−0.14	−0.16	0.00
	(0.05)	(0.12)	(0.06)
HH Member Victim	0.03	0.10	0.10
	(0.05)	(0.10)	(0.10)
Baseline Nbd Quality	0.13	−0.02	−0.10
	(0.01)	(0.03)	(0.02)
Site Fixed Effects/Constant			
Baltimore	0	0.03	−0.25
	−	(0.13)	(0.13)
Boston	0.31	−0.49	0.02
	(0.10)	(0.21)	(0.02)
Chicago	−0.04	0.04	−0.50
	(0.09)	(0.13)	(0.13)
Los Angeles	−0.52	0.39	0.47
	(0.10)	(0.17)	(0.13)
New York City	−0.58	0.60	−0.13
	(0.09)	(0.15)	(0.15)
Unobserved Factors			
ρ^S and ρ^M	−	0.07	−0.17
	−	(0.10)	(0.11)

Table 6: Average Neighborhood Characteristics in 2000 Conditional on Neighborhood Quality

Nbd Characteristic	Mean $|D = 1$	Mean $|D = 2$	Unconditional Median	Mean $|D = 10$
Poverty Rate (%)	33	22	9	3
HS Diploma (%)	55	65	83	95
BA (%)	7	11	19	52
Single-Headed HHs (%)	52	38	24	11
Female Unemployment Rate (%)	16	10	5	2
Male EPR (×100)	55	65	79	89

Table 7: Low-Poverty ($\leq 10\%$), Low-Quality ($D \leq 3$) Neighborhoods in MTO States in 2000

Nbd Quality	Number of Residents
$D = 1$	6,362
$D = 2$	93,385
$D = 3$	751,738

Table 8: Adult LATE Estimates

Outcome	$\widehat{\triangle}_{1,2}^{LATE}\left((\mu(X_i), V_i) \in \mathcal{C}_{1,2}^M, \tau^M = 1\right)$	Control Mean in $\mathcal{S}_{1,2}^M$
Labor Market		
In Labor Force (%)	25.8*	53.2
	(20.2)	
Employed (%)	31.2*	41.7
	(21.7)	
Household Income ($)	5,616	13,506
	(4,990)	
Earnings ($)	1,970	7,642
	(4,672)	
Welfare Benefits		
Received TANF (%)	−40.0	39.9
	(32.0)	
Health		
BMI (Raw)	−3.1*	30.9
	(8.1)	

Note: $\widehat{\triangle}_{1,2}^{LATE}\big((\mu(X_i), U_{Di}) \in \mathcal{C}_{1,2}^M, \tau_i^M = 1\big)$ estimates are for individuals with observed and unobserved choice model components in $\mathcal{S}_{1,2}^M \equiv \{(\mu(X_i), V_i) | \mu(X_i) \in [-0.6, 0.4], u_{Di} \in [0.43 + 0.30\mu(X_i), 0.68 + 0.15\mu(X_i)]\}$. Control means are also computed for the subsample in this region and outside of this region (both conditional on $D_i = 1$). Standard errors are computed using 100 bootstrap replications, with * denoting statistical significance at the 10% level determined using either the 10th or 90th percentile of the 100 bootstrapped replications, and * denoting statistical significance at the 10% level determined by $\widehat{\triangle}_{1,2}^{LATE} \pm 1.2816\widehat{\sigma}$.

Table 9: Adult Program Effects Estimates by Neighborhood Selection Groups

	Falsification Set: $(\mu(X_i), V_i) \in \mathcal{F}_{1,2}^M$			Identification Support Set: $(\mu(X_i), V_i) \in \mathcal{S}_{1,2}^M$		
	$E[Y|Z^M=1]$	$E[Y|Z^M=0]$	$E[Y|Z^M=1]-E[Y|Z^M=0]$	$E[Y|Z^M=1]$	$E[Y|Z^M=0]$	$E[Y|Z^M=1]-E[Y|Z^M=0]$
Neighborhood Selection						
Nbd Quality ($D \in \{1,2,3,4,5\}$)	1.02	1.00	0.02	1.37	1.00	0.37
Nbd Quality ($q \in [0,50]$)	1.7	0.4	1.2	6.4	1.1	5.3
Labor Market						
In Labor Force (%)	63.6	63.6	0.0	63.0	53.2	9.8
Employed (%)	47.1	53.6	−6.5	53.5	41.7	11.8
Household Income ($)	14,252	14,134	119	15,629	13,506	2,123
Earnings ($)	7,583	8,554	−971	8,364	7,642	722
Welfare Benefits						
Received TANF (%)	32.2	33.7	−1.5	24.9	39.9	−15.0
Health						
BMI (Raw)	30.0	30.4	−0.3	29.7	30.9	−1.2

8 Appendix A: Intuition for the Identification Support Set $\mathcal{S}^M_{j,j+1}$

We will interchangeably define parameters in terms of V or U_D based on which best facilitates exposition, recalling that

$$U_D \equiv F_V(V).$$

For the sake of intuition, the following discussion assumes a restricted version of the choice model from the text in which $\tau^M_i = 1$ for all i. We may refer to Z^M as Z, and for the sake of exposition here we assume away the Section 8 voucher group. The following intuition would also apply to the scenario in which $V^M \perp\!\!\!\perp V_i$, so that $\triangle^{LATE}_{j,j+1}\big((\mu(X), U_D) \in \mathcal{C}^M_{j,j+1}, \ \tau^M = 1 \big) = \triangle^{LATE}_{j,j+1}\big((\mu(X), U_D) \in \mathcal{C}^M_{j,j+1} \big)$.

In the general model in the paper, compliance (ie, response to the instrument) is determined by household i's $\mu(X_i)$, V_i, *and* their $\{\gamma^M_{ji}\}$ (ie, τ^M_i through their $\mu^M(X_i)$, V^M_i). In the restricted version of the model considered in this Appendix for the sake of providing intuition, $\gamma^M_{ji} = \gamma^M_j$ for all households i for each transtion j, so that all heterogeneity in household i's response to the instrument is entirely determined by their $\mu(X_i)$ and V_i. Although compliance is still instrument-specific, driven by the homogeneous effects each specific instrument Z has on selection into treatment $\{\gamma^M_j\}$, the treatment effects no longer need to be defined in terms of compliers for a specific instrument. For example, we can define the Marginal Treatment Effect (MTE) as

$$\triangle^{MTE}_{j,j+1}\big(\mu(x), \ u_D \big) \ \equiv \ E\big[Y(j+1) \, \big| \, \mu(x), \ u_D \big] \ - \ E\big[Y(j) \, \big| \, \mu(x), \ u_D \big]$$

and the Local Average Treatment Effect (LATE) from the text now becomes:

$$\triangle^{LATE}_{j,j+1}\big(\mu(x), \ \underline{u}, \ \overline{u} \big) \equiv \ E\big[Y(j+1) \, \big| \, \mu(x), \ u_D \in [\underline{u}, \overline{u}] \big] \ - \ E\big[Y(j) \, \big| \, \mu(x), \ u_D \in [\underline{u}, \overline{u}] \big]$$
$$= \ \frac{\int_{\underline{u}}^{\overline{u}} \triangle^{MTE}_{j,j+1}\big(\mu(x), \ u_D \big) \, du_D}{\overline{u} - \underline{u}}.$$

In this version of the model there is still heterogeneity in response to the instrument, but this heterogeneity is now restricted. Under these restrictions identification is instrument specific, but the definition of parameters is not. Consider the case of effects for a given value of the observed, invariant characteristics X (ie, $\mu(X) = m^*$). Two such examples with $D \in \{1, 2, 3\}$ are illustrated in Figures 1 and 2. We begin 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 $\boldsymbol{\pi}^1(X)$, where

$$\pi^Z_j(X) \equiv Pr(D > j | X, Z).$$

The patterns of heterogeneity in response to the instrument allowed under Assumptions A1-A8 ensure that the instrument monotonically increases individuals' latent index, so that $\pi^1_j(X) \geq \pi^0_j(X)$ for all j and for all i. From A9 we can attribute the variation in average outcomes induced by the instrument 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):

$$E[Y|Z=1] - E[Y|Z=0] = \sum_{j=1}^{J} \int_0^1 \mathbf{1}\{\pi_j^1 \le u_D < \pi_{j-1}^1\} E[Y(D=j)|u_D]du_D \qquad (1)$$

$$- \sum_{j=1}^{J} \int_0^1 \mathbf{1}\{\pi_j^0 \le u_D < \pi_{j-1}^0\} E[Y(D=j)|u_D]du_D$$

$$= \sum_{j=1}^{J} \left\{ \int_{\pi_j^1}^{\pi_{j-1}^1} E[Y(D=j)|u_D]du_D \qquad (2) \right.$$

$$\left. - \int_{\pi_j^1}^{\pi_{j-1}^1} \sum_{k=1}^{J} \left[\mathbf{1}\{\pi_k^0 \le u_D < \pi_{k-1}^0\} E[Y(D=k)|u_D] \right] du_D \right\}.$$

As long as $\tau_i^M = 1$ for all i or $V_i^M \perp\!\!\!\perp V_i$,

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

allowing us to rewrite Equation 1 as:

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

$$= \sum_{j=1}^{J-1} \left\{ \triangle_{j,j+1}^{LATE}(\pi_j^0, \pi_j^1) \ \left[\pi_j^1 - \pi_j^0 \right] \right\}. \qquad (3)$$

These expressions can be seen in Figures 1-2 for two examples with $J = 3$.

An important issue to remember is that the preceding and ensuing results implicitly condition on observable characteristics X. Figure 3 shows that Example I in Figure 1 is just one cross section taken from an interval of the observed, invariant characteristics X (ie, effects for $X = x$).

Now suppose we augment Assumption A7 with:

A7* $\pi_j^1 > \pi_{j-1}^0$ for all $j \in \{1, \ldots, J-1\}$.

Example I in Figure 1 shows such an ordering when $J = 3$. Under A7* the right hand side of Equation 3 can be derived quickly since

$$E[Y|Z=1] = \sum_{j=1}^{J} \int_{\pi_j^1}^{\pi_{j-1}^1} E[Y(D=j)|u_D]du_D$$

$$= \int_{\pi_1^1}^{1} E[Y(D=1)|u_D]du_D$$

$$+ \sum_{j=1}^{J-1} \left\{ \int_{\pi_j^0}^{\pi_j^1} E[Y(D=j+1)|u_D]du_D + \int_{\pi_{j+1}^1}^{\pi_j^0} E[Y(D=j+1)|u_D]du_D \right\}$$
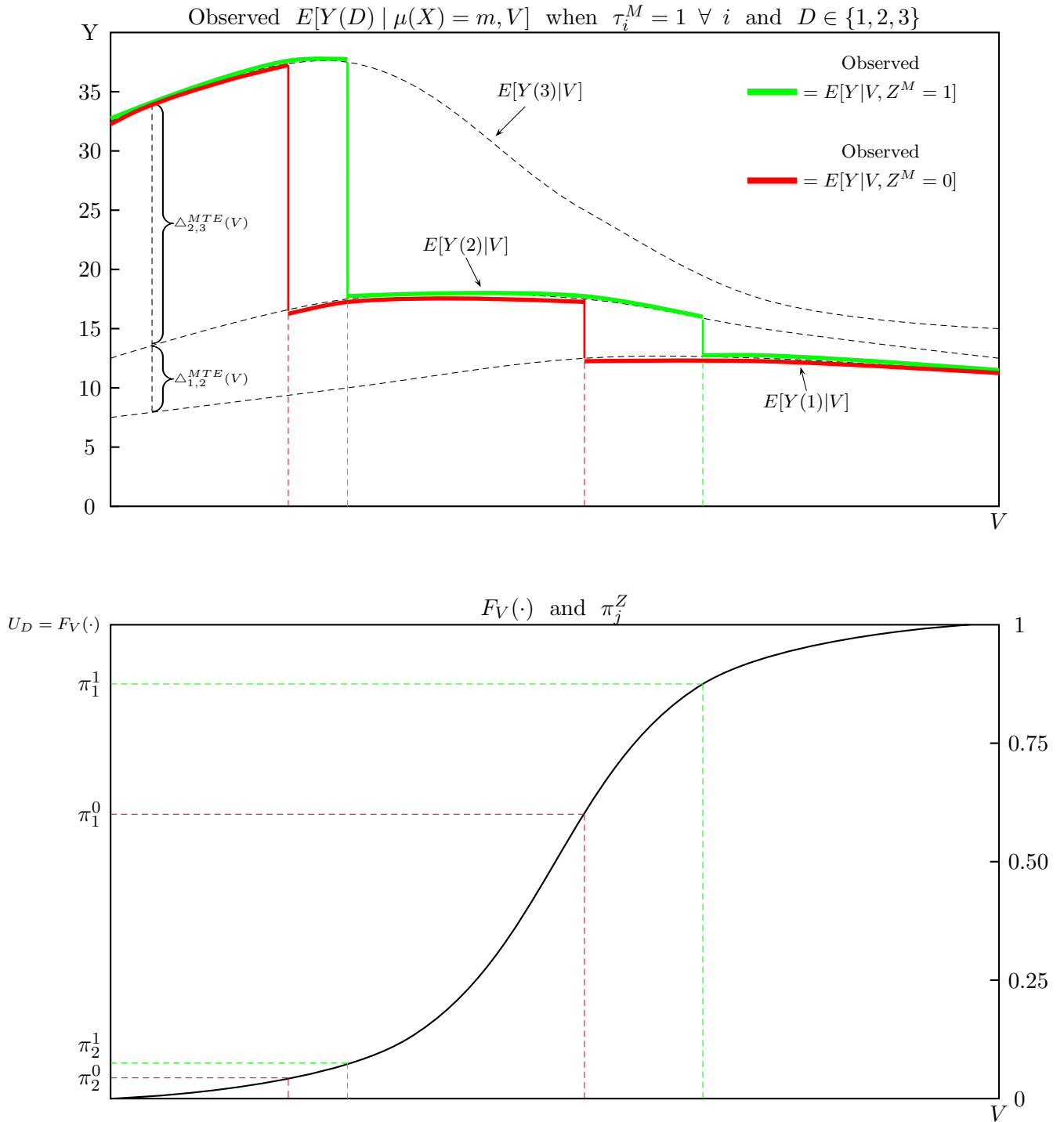
Figure 1: Example I: Potential Outcomes and Marginal Treatment Effects Given $\mu(X) = m$
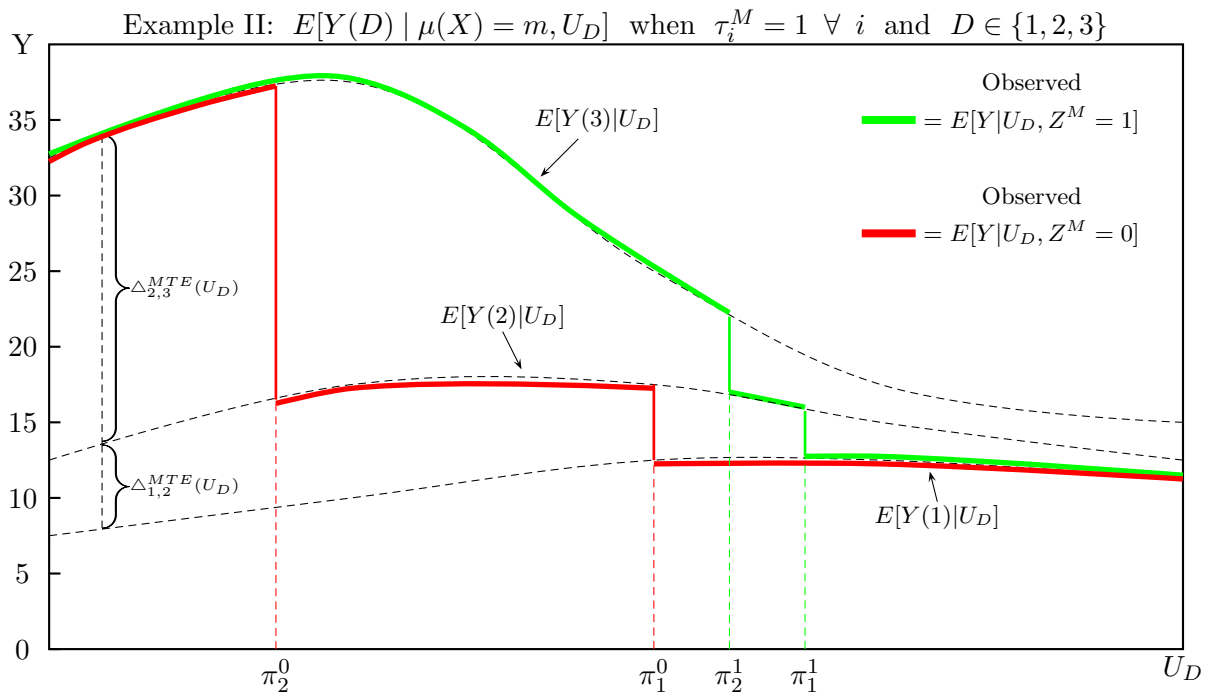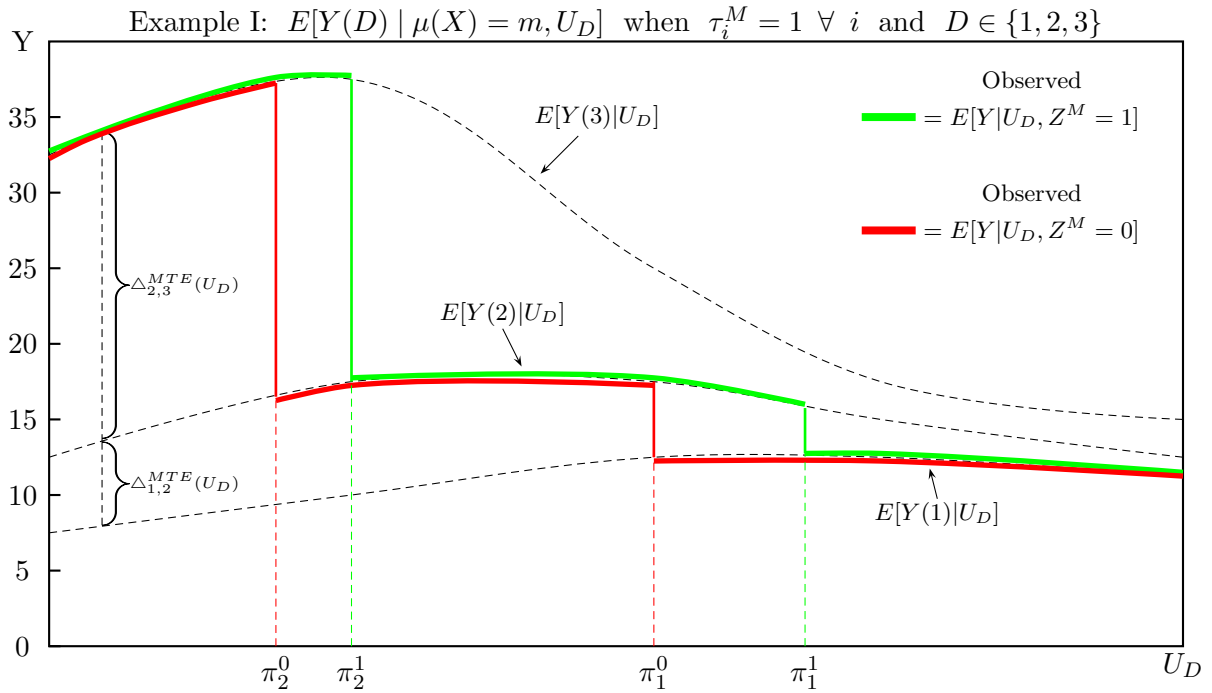
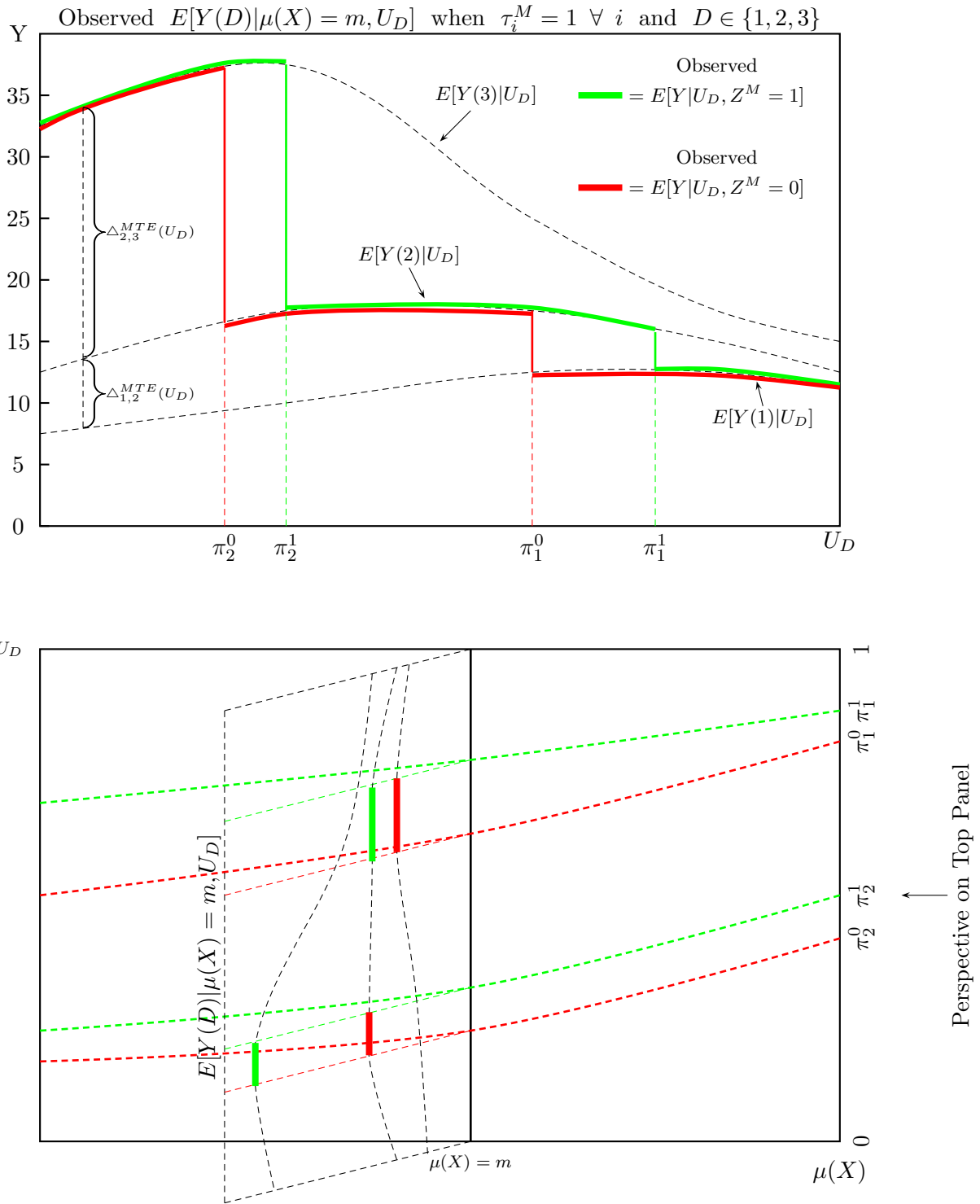Figure 2: Potential Outcomes and Marginal Treatment Effects Given $\mu(X) = m$

Figure 3: Example I: Potential Outcomes and Marginal Treatment Effects Given $\mu(X) = m$

and

$$E[Y|Z=0] = \sum_{j=0}^{J} \int_{\pi_j^0}^{\pi_{j-1}^0} E[Y(D=j)|u_D]du_D$$

$$= \int_{\pi_1^1}^{1} E[Y(D=1)|u_D]du_D$$

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

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

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

$$= \sum_{j=1}^{J-1} \int_{\pi_j^0}^{\pi_j^1} \triangle_{j,j+1}^{MTE}(u_D)du_D = \sum_{j=1}^{J-1} \left\{ \triangle_{j,j+1}^{LATE}(\pi_j^0, \pi_j^1) \ \left[\pi_j^1 - \pi_j^0 \right] \right\}. \quad (4)$$

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_{Di} lies from the data, we can estimate these LATEs. 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

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

Now suppose that we drop A7* and replace it with the less restrictive original Assumption A7. 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 3 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:

$$E[Y|u_D \in [a,b], Z=1] - E[Y|u_D \in [a,b], Z=0] = \frac{\int_a^b \sum_{j=n}^{m-1} \triangle_{j,j+1}^{MTE}(u_D)du_D}{b-a} \quad (5)$$

That is, conditional on X_i, the empirical pattern of selection into treatment determines the iden-

tification support set $\mathcal{S}^M_{j,j+1}$ by way of the the interval $[a, b]$.

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 $\pi^1_j > \pi^0_{j-1}$ for some j, the corresponding LATE parameter is still separately identified over the interval

$$\left[\max\{\pi^0_j, \pi^1_{j+1}\}, \min\{\pi^0_{j-1}, \pi^1_j\} \right].$$

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

Comparing Example I and Example II in Figure 2 helps to illustrate how the ordering of the π^Z_j determines identification (ie, the boundaries of the identification support set $\mathcal{S}^M_{j,j+1}$). Since $\pi^1_2 > \pi^1_0$ in Example II, the instrument identifies

$$\triangle^{LATE}_{1,2}(\pi^1_2, \pi^1_1) = \frac{\int^{\pi^1_1}_{\pi^1_2} \triangle^{MTE}_{1,2}(u_D) du_D}{\pi^1_1 - \pi^1_2} = E[Y|u_D \in [\pi^1_2, \pi^1_1], Z = 1] - E[Y|u_D \in [\pi^1_2, \pi^1_1], Z = 0]$$

However, over the interval $[\pi^0_1, \pi^1_2] = [\pi^0_1, \min\{\pi^1_2, \pi^1_1\}]$ we cannot separately identify each transition-specific LATE. Instead the instrument identifies:

$$\begin{aligned} \triangle^{LATE}_{1,3}(\pi^0_1, \pi^1_2) &= \triangle^{LATE}_{1,2}(\pi^0_1, \pi^1_2) + \triangle^{LATE}_{2,3}(\pi^0_1, \pi^1_2) \\ &= E[Y|u_D \in [\pi^0_1, \pi^1_2], Z = 1] - E[Y|u_D \in [\pi^0_1, \pi^1_2], Z = 0] \\ &= \frac{\int^{\pi^1_2}_{\pi^0_1} \left(\triangle^{MTE}_{1,2}(u_D) + \triangle^{MTE}_{2,3}(u_D) \right) du_D}{\pi^1_2 - \pi^0_1}. \end{aligned}$$

But over the interval $[\pi^0_2, \pi^0_1] = [\pi^0_2, \min\{\pi^0_1, \pi^1_2\}]$ the instrument does again separately identify the LATE parameter:

$$\triangle^{LATE}_{2,3}(\pi^0_2, \pi^0_1) = E[Y|u_D \in [\pi^0_2, \pi^0_1], Z = 1] - E[Y|u_D \in [\pi^0_2, \pi^0_1], Z = 0] = \frac{\int^{\pi^0_1}_{\pi^0_2} \triangle^{MTE}_{2,3}(u_D) du_D}{\pi^0_1 - \pi^0_2}.$$

Figure 2 shows these LATEs graphically. These identification issues are why determination of the identification support set $\mathcal{S}^M_{j,j+1}$ is such a crucial step in estimation.

Readers interested in further discussion of these issues are referred to the standard references Heckman et al. (2006), Heckman and Vytlacil (2005), and Imbens and Angrist (1994).

9 Appendix B: Comparison with Other Identification Strategies

9.1 Using Ignorability to Identify Homogeneous Effects

For the sake of exposition we maintain the assumption from Appendix A that $\{\gamma_{ji}^M\} = \{\gamma_j^M\}$ for all households i for all transitions j (ie, that $\tau_i^M = 1$ for all i). One approach to identifying treatment effects would be to strengthen assumptions A1-A10. 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.[1] When f denotes the distribution of potential outcomes, SI can be written as:

$$f\{y_j|X\} = f\{y_j|X, U_D \in [a,b]\} \quad \text{for all} \ \ j \in \{1, \ldots, J\}. \tag{6}$$

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

$$f\{u_j|X\} = f\{u_j|X, U_D \in [a,b]\} \quad \text{for all} \ \ j \in \{1, \ldots, J\}, \tag{7}$$

or

$$U_j \perp\!\!\!\perp U_D|X \quad \text{for all} \ \ j \in \{1, \ldots, J\}. \tag{8}$$

When calculating $E[Y(D=j)|U_D]$ as in Figures 1 and 2, 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 4. 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.

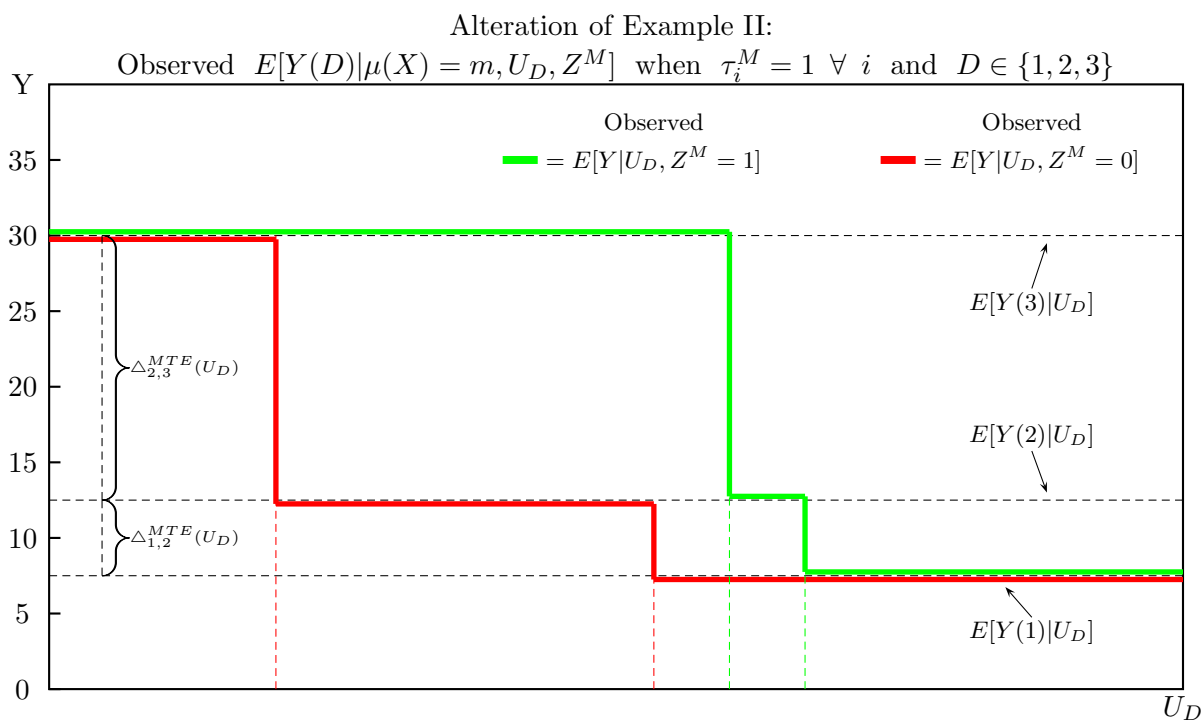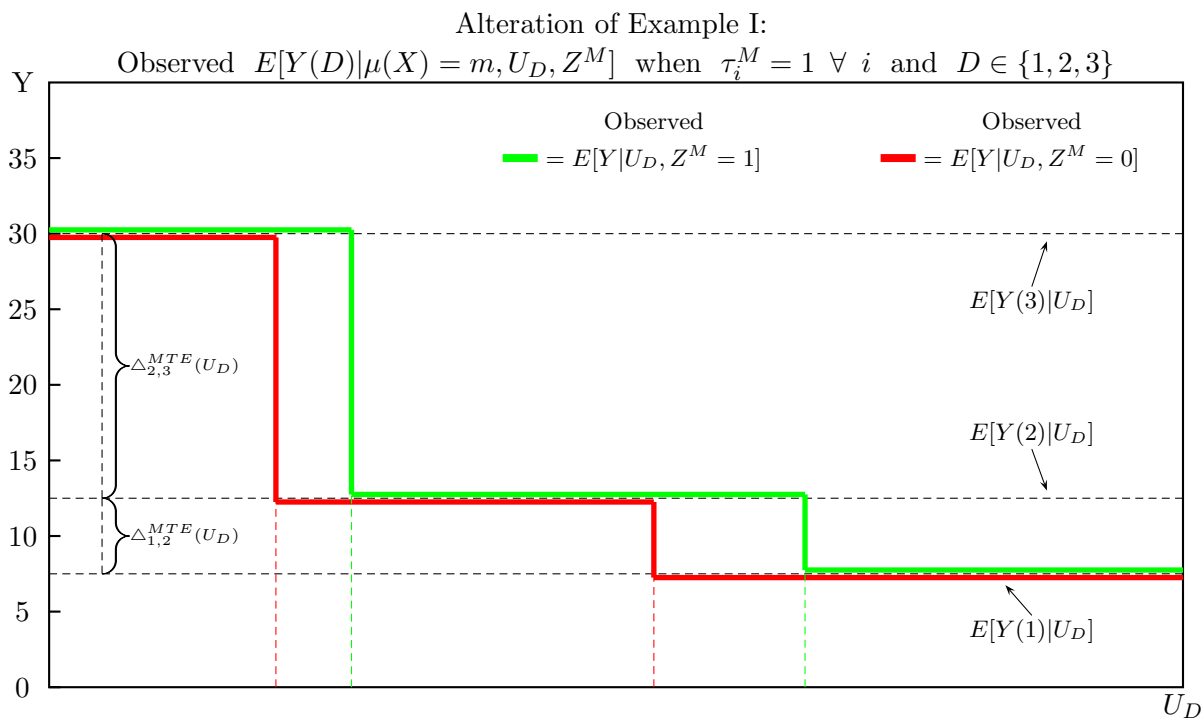[1] See Imbens (2004) for a discussion of the SI assumption in models with a binary treatment.

Figure 4: Altering Examples I and II to Satisfy Strong Ignorability

Figure 5 shows a binary example to help illustrate the differences between the heterogeneity in treatment effects allowed under our assumptions A1-A10 and under SI. Let $\theta = \mu(X)$ be an index of observed characteristics. The average effect of treatment varies across observed characteristics 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).[2] 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 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 (2013), Clampet-Lundquist and Massey (2008), Sampson (2008)).

---

[2]Essential heterogeneity between levels j and $j + 1$ is defined as

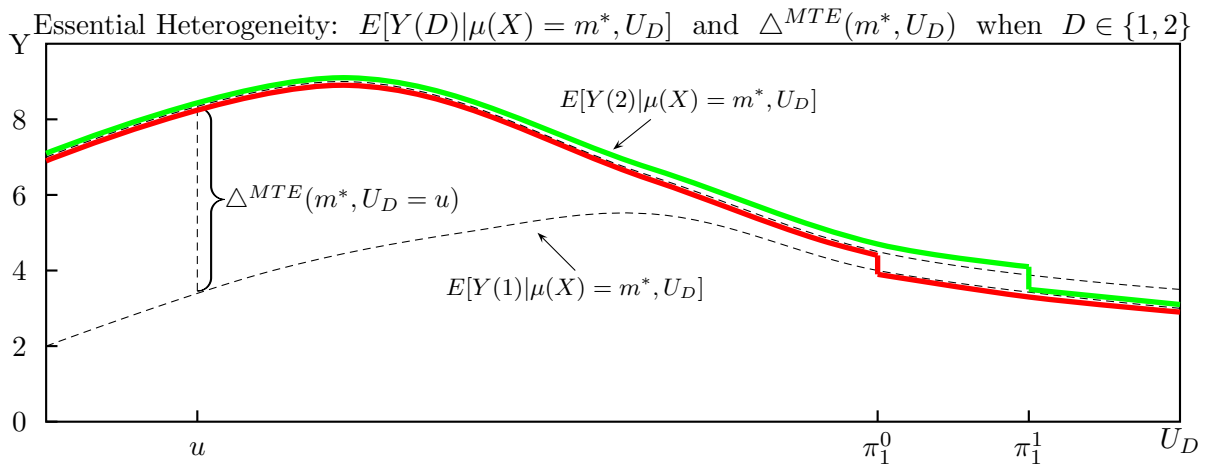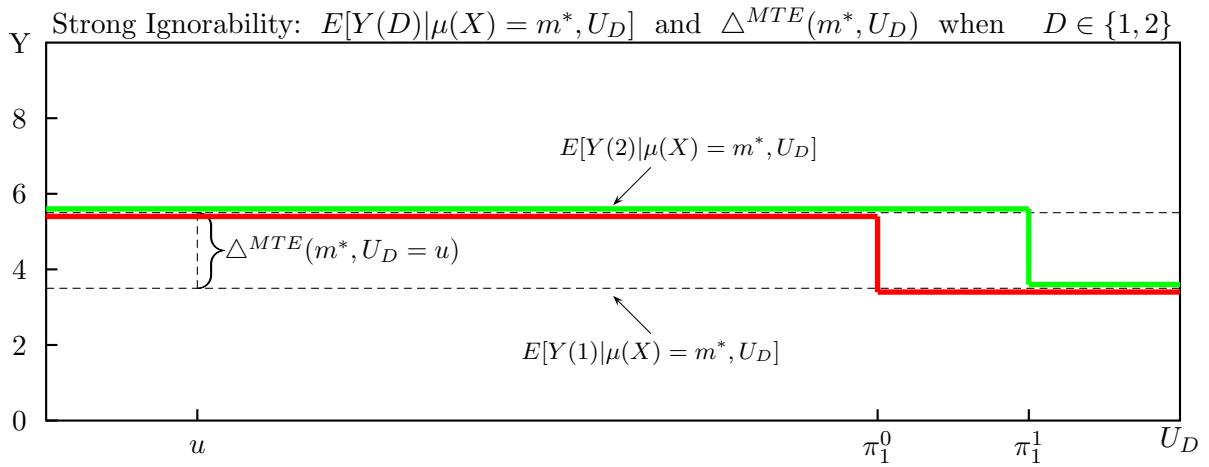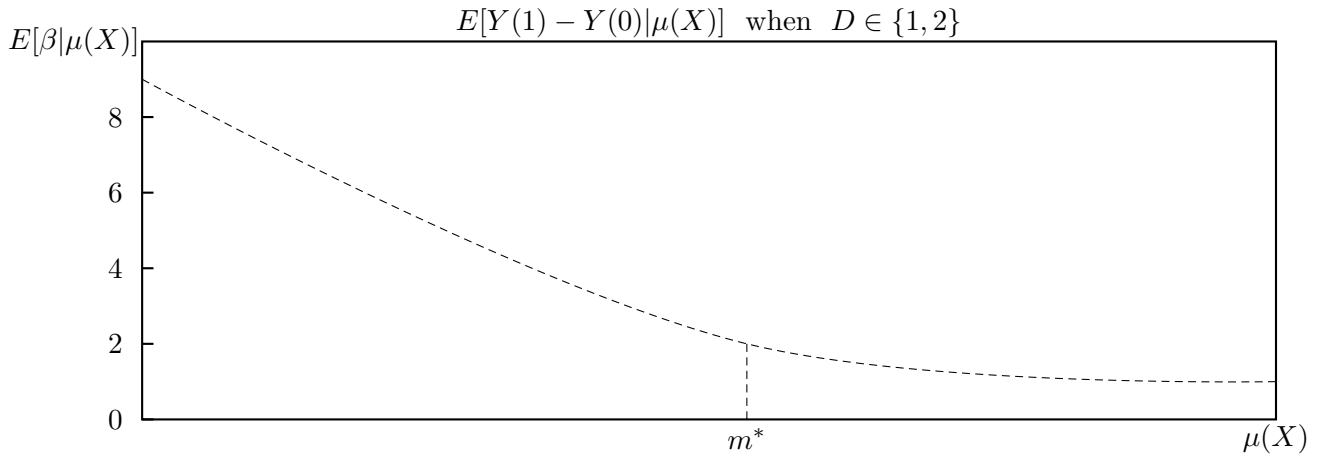**EH** $COV(U_{j+1} - U_j, U_D) \mid X \neq 0.$

Figure 5: Binary Example with and without Strong Ignorability

9.2 Using Transition-Specific Instruments to Identify Heterogeneous Transition-Specific Effects

Another approach to identifying MTE parameters in this model is to augment assumptions A1-A10, as done by Heckman et al. (2006) (HUV), with an assumption about the ordered choice model. HUV assumes there exist instrumental variables W_j for each margin of selection $j = 1, \ldots, J-1$ such that the distribution of $C_j(W_j)$, conditional on X, Z, and $\{C_h : h \neq j\}$, is nondegenerate and continuous. We label this assumption, which does not hold in our model, A11. Under A11 each margin of choice can be varied 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 $\boldsymbol{\pi} \equiv [\pi_1^Z(X), \pi_2^Z(X), \ldots, \pi_{J-1}^Z(X)]$, and that $E[Y|\boldsymbol{\pi}^Z]$ is differentiable under some distributional assumptions. 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 $k \neq j$:

$$\triangle_{j,j+1}^{MTE}(x, \pi_j) = \frac{\partial E[Y|X = x, \boldsymbol{\pi}^Z(x) = \boldsymbol{\pi}]}{\partial \pi_j}. \tag{9}$$

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 9.

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 transition-specific instrument W_j would vary the choice margin over some interval, but not over the entire unit interval. A discussion of related issues can be found in Carneiro et al. (2011) for a binary context.

9.3 Using a Binary Instrument to Identify the Average of Multiple Heterogeneous, Transition-Specific Effects

Another alternative would be to weaken A1-A10, which would allow us to estimate the Average Causal Response (ACR) parameter introduced in Angrist and Imbens (1995). Under Assumptions A1-A10 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. An unattractive feature of ACRs identified under these weaker assumptions, however, is that they yield only one summary parameter that is quite difficult to interpret. In contrast, imposing the structure of our choice model allows 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. Nevertheless, the ACR can still be of great interest, and in Appendix D we estimate ACR parameters from the MTO data.

10 Appendix C: Ordered Choice Model Parameters

Here we present a simple numerical example to illustrate why the probability of *feasibly* entering into a Section 8 contract in a neighborhood of quality q is central to modeling neighborhood selection in MTO, which is the reason we leave rents and housing prices out of our model (See Collinson and Ganong (2013) for a related model.). The numerical example also illustrates the interpretation of parameters of our ordered choice model in terms of some of the factors driving the Marginal Benefit function for the Section 8 and Experimental voucher groups.

Suppose that the benefit of living in a neighborhood of quality q is a weighted average over a set of potential outcomes

$$B(q) = \sum_k w^k Y^k(q),$$

where one random variable $Y^k(q)$ is the social network one has access to when living in a neighborhood of quality q.[3] Additionally, let $Pr(S8|q)$ be the probability of *feasibly* entering into a Section 8 contract in a neighborhood of quality q. Then the expected cost of living in a neighborhood of quality q is 30 percent of income if a household finds Section 8 housing, and the expected market rent otherwise:[4]

$$
\begin{aligned}
E[C(q|Z^S, Z^M)] = \mathbf{1}\{Z^S = 0, Z^M = 0\} \ &E[\text{rent}(q)] \\
+ \mathbf{1}\{Z^S = 1, Z^M = 0\} \ &\Big[Pr(S8|q, Z^S = 1)0.30 \times \text{Income} \\
&+ \ \big(1 - Pr(S8|q, Z^S = 1)\big) E[\text{rent}(q)]\Big] \\
+ \mathbf{1}\{Z^S = 0, Z^M = 1\} \ &\Big[Pr(S8|q, Z^M = 1)0.30 \times \text{Income} \\
&+ \ \big(1 - Pr(S8|q, Z^M = 1)\big) E[\text{rent}(q)]\Big]
\end{aligned}
$$

Thus the expected net benefit at any neighborhood quality q for Section 8 and experimental voucher holders is:

$$E[NB(q|Z^S, Z^M)] = E[B(q)] - E[C(q|Z^S, Z^M)],$$

---

[3]See Blume et al. (2011) for a related discussion on the importance of disrupting social networks for housing mobility programs like MTO. Although we consider social networks and other outcomes as part of the benefit of living in a neighborhood of quality q, we might just as easily categorize this outcome and others as costs.

[4]Recall that $Z^S = 1 \Rightarrow Z^M = 0$ and $Z^M = 1 \Rightarrow Z^S = 0$.

where

$$E\big[NB(q|Z^S = 0, Z^M = 0)\big] = E\left[\sum_k w^k Y^k(q)\right] - \left[E\big[\text{rent}(q)\big]\right]$$

$$E\big[NB(q|Z^S = 1, Z^M = 0)\big] = E\left[\sum_k w^k Y^k(q)\right]$$
$$- \left[Pr(S8|q, Z^S = 1)0.30 \times \text{Income} + \big(1 - Pr(S8|q, Z^S = 1)\big)E[\text{rent}(q)]\right]$$

$$E\big[NB(q|Z^S = 0, Z^M = 1)\big] = E\left[\sum_k w^k Y^k(q)\right]$$
$$- \left[Pr(S8|q, Z^M = 1)0.30 \times \text{Income} + \big(1 - Pr(S8|q, Z^M = 1)\big)E[\text{rent}(q)]\right].$$

To illustrate the importance of the probability of entering a Section 8 contract, here we consider a particular specification and parameterization of net benefit functions capturing particular cost functions. Suppose $E[B(q)]$ and $E[C(q)]$ were both increasing functions of q, with $E[C(q)]$ rising faster than $E[B(q)]$. At low q, due to the 10 percent poverty restriction they face, the MTO voucher group faces a restricted set of neighborhoods relative to the standard Section 8 voucher group. The counseling offered to the MTO voucher group does not offset this restriction, so $Pr(S8|q, Z^S = 1) > Pr(S8|q, Z^M = 1)$ at these low levels of q. As quality increases, though, the set of neighborhoods satisfying the experimental restrictions starts getting closer to the full set of neighborhoods with Section 8 housing. At some $\widetilde{q}$, eligible neighborhoods become sufficiently similar so that due to the counseling offered to the experimental group, the probabilities switch, and now it is actually the case that $Pr(S8|q, Z^S = 1) < Pr(S8|q, Z^M = 1)$ for $q > \widetilde{q}$.
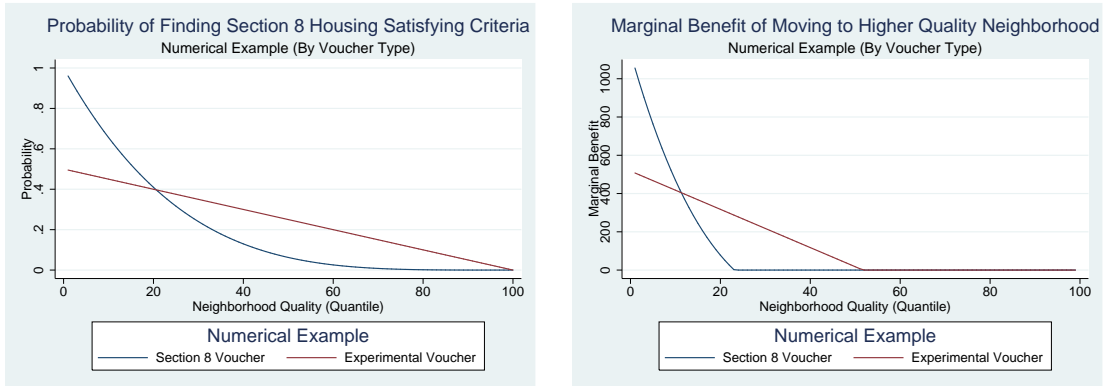
(a) Probability of Finding Section 8 Housing

(b) Marginal Benefit Functions, Conditional on Voucher Type

Figure 6: Probability of Feasibly Finding Section 8 Housing and Marginal Benefit Functions

Figure 6a shows two numerical examples of $Pr(S8|q, Z^S = 1)$ and $Pr(S8|q, Z^M = 1)$ satisfying this qualitative description, and Figure 6b shows the resulting Marginal Benefit functions.[5] We

---

[5]The precise parameterization used in this numerical example is: $E[B(q)] = 25,000 + 1,000\,q; \quad E[\text{rent}(q)] =$

can see that at low levels of quality, those holding the Section 8 voucher are more likely to move to a higher quality neighborhood. However, at $\widetilde{q}$, the MTO voucher becomes more effective than the ordinary Section 8 voucher.

This numerical example highlights the flexibility and interpretation of our ordered choice model, especially when $Pr(S8|q, Z^S = 1)$ and $Pr(S8|q, Z^M = 1)$ are not observed in the data. The cost and marginal benefit functions in the model can very flexibly characterize the effects of the Section 8 and MTO vouchers, in this example even allowing the effectiveness of the programs to cross at some quality level $\widetilde{q}$. In terms of the parameters of our model, the $\{C_j\}$ represent elements of both benefits $E[B(q)]$ and costs $E[C(q|Z^S, Z^M)]$ (regardless of the values taken by Z^S and Z^M), while the $\{\gamma_j^S\}$ represent elements of the cost function $E[C(q|Z^S = 1)]$ only, and the $\{\gamma_j^M\}$ represent elements of $E[C(q|Z^M = 1)]$.

$1,000\, q;\quad Pr(S8|q, Z^S = 1) = \left(\frac{100-q}{100}\right)^4;\quad Pr(S8|q, Z^M = 1) = 0.5\left(\frac{100-q}{100}\right);\quad \text{Income} = 15,000.$

References

Aliprantis, D. (2013). Assessing the evidence on neighborhood effects from Moving to Opportunity. *Federal Reserve Bank of Cleveland Working Paper 12-33r*.

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.

Blume, L., W. Brock, S. Durlauf, and Y. Ioannides (2011). Identification of social interactions. In J. Benhabib, A. Bisin, and M. Jackson (Eds.), *Handbook of Social Interactions*, pp. 854–964. North-Holland.

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.

Collinson, R. and P. Ganong (2013). Incidence and price discrimination: Evidence from housing vouchers. *Harvard University Joint Center for Housing Studies Working Paper W13-7*.

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.

Imbens, G. W. and J. D. Angrist (1994). Identification and estimation of Local Average Treatment Effects. *Econometrica 62*(2), 467–475.

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

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.