Appendix to
"Evidence of Neighborhood Effects from Moving to Opportunity: LATEs of Neighborhood Quality"

Dionissi Aliprantis
Federal Reserve Bank of Cleveland

Francisca G.-C. Richter*
Case Western Reserve University

First Version: March 2012. This Version: May 5, 2016

A V_i Is Identified through a Discrete Ordered Choice Model

Assuming that the unobserved component of each household's neighborhood quality decision is distributed as $V_i \sim \mathcal{N}(0,1)$ is a normalization of our model that anchors estimates under various discretizations of treatment. This Appendix shows that the ordered choice models implied by any two partitions of quality both generate estimates of this same underlying V_i. We first show that a given partition of the continuous neighborhood quality variable into discrete levels implies a specific ordered choice model of neighborhood selection. We then show that given a sequence of partitions of quality, the sequence of approximations $\widehat{V}_i$ implied by the corresponding sequence of ordered choice models converges uniformly to V_i as the norm of the partitions of quality converges to zero.

This result allows for us to move freely between different partitions of quality when estimating the ordered choice model and when estimating causal effects on outcomes. In practice, this means that we use one partition of quality when estimating the ordered choice model, using the sample population to determine where cutpoints are located so as to improve estimation. We use a different partition of quality when estimating causal effects on outcomes, placing cutpoints so as to characterize meaningful margins of neighborhood characteristics.

A.1 Partitions Generating a Discrete Ordered Choice Model

Consider a partition of the continuous quality measure $q \in [\delta, 1] \subset [0,1]$ into K discrete levels:

$$\mathcal{P}_K^q = \left\{ \delta = q_0^K, q_1^K, \ldots, q_{K-1}^K, q_K^K = 1 \right\}.$$

*Address: Federal Reserve Bank of Cleveland, Research Department, PO Box 6387, Cleveland, OH 44101-1387, USA. Phone: +1(216)579-3021. E-mail: *dionissi.aliprantis@clev.frb.org, francisca.richter@case.edu.*

Partition $\mathcal{P}_K^q$ defines the discrete treatment

$$Q_i^K = \begin{cases} 1 & \text{if } q_i \in [q_0^K, q_1^K]; \\ 2 & \text{if } q_i \in (q_1^K, q_2^K]; \\ \vdots & \vdots \quad \vdots \\ K & \text{if } q_i \in (q_{K-1}^K, q_K^K]. \end{cases}$$

In the corresponding discrete model of ordered neighborhood choice, we denote the net benefit of moving from discrete quality level $Q_i^K = k$ to $Q_i^K = k+1$ by MB_{ik}. For interior solutions, household i maximizes the net benefit at $MB_{ik} = k^*$ satisfying $MB_{ik^*-1} \geq 0$ and $MB_{ik^*} < 0$. For the corner solutions $k_i^* = 1$ and $k_i^* = K$, the maximization conditions are, respectively, that $MB_{i1} < 0$ and $MB_{iK-1} \geq 0$. To facilitate exposition, we consider the case in which observed characteristics are fixed at a particular realization $X_i = x_i$, allowing us to consider fixed $\mu(x_i)$. Since we define $MB_{i0} = \max\{\ 0,\ \ \mu(x_i) - C_0 - V_i\ \}$ and we constrain treatment level k to be in the set $\{1, 2, \ldots, K\}$ (both to allow the model to handle corner solutions), these conditions can also be written as:

$$Q_i^K = k \iff \mu(x_i) - C_k < V_i \qquad \text{for} \quad j = 1; \tag{1}$$

$$Q_i^K = k \iff \mu(x_i) - C_k < V_i \leq \mu(x_i) - C_{k-1} \qquad \text{for} \quad k = 2, \ldots, K-1; \tag{2}$$

$$Q_i^K = k \iff V_i \leq \mu(x_i) - C_{k-1} \qquad \text{for} \quad k = K. \tag{3}$$

The cut points of this ordered choice model generate a partition of a closed interval in $\mathbb{R}$:

$$\mathcal{P}_K^C = \{C_1, \ldots, C_{K-1}\}.$$

Equations 1-3 together with partition $\mathcal{P}_K^C$ define a corresponding partition of the closed interval $[\mu(x_i) - C_{K-1}, \mu(x_i) - C_1] \subset \mathbb{R}$:

$$\mathcal{P}_K^V = \{\mu(x_i) - C_{K-1}, \ldots, \mu(x_i) - C_1\}.$$

A.2 Discrete Model Estimates of V_i under Different Partitions of Quality

We utilize two facts in our central result about the estimates of V_i generated by the ordered choice models implied by different partitions of neighborhood quality. We prove these two facts before proving our central result.

Lemma 1. *Suppose that the partitions $\mathcal{P}_M^q$ and $\mathcal{P}_N^q$ share the point $q^{M,N}$ in common, which defines the upper cutoff for $Q^M = m$ for some $m \in \{1, \ldots, M\}$ and $Q^N = n$ for some $n \in \{1, \ldots, N\}$:*

$$q^{M,N} = q_m^M = q_n^N.$$

Then the cutpoints in ordered choice models defined on $\mathcal{P}^q_M$ and $\mathcal{P}^q_N$ are equal:

$$C^M_{m-1} = C^N_{n-1}.$$

Proof. The result follows immediately from the structure of the ordered choice model that

$$\begin{aligned} Pr(q_i^* > q^{M,N}) &= Pr(Q_i^M > m) = \Phi(\mu(x_i) - C^M_{m-1}) \\ &= Pr(Q_i^N > n) = \Phi(\mu(x_i) - C^N_{n-1}), \end{aligned}$$

which implies that the cut points corresponding to the shared point in the partitions are equal, or

$$C^M_{m-1} = C^N_{n-1}. \tag{4}$$

Lemma 2. *If q_i^* is continuously distributed given x_i, then C_k is a continuous function of the chosen cut point q_k.*[1]

Proof. To see why this is true, note that

$$Pr(q_i^* > q_k) = \Phi(\mu(x_i) - C_k), \tag{5}$$

which implies that

$$C_k = \Phi^{-1}\big[1 - F(q_k|x_i)\big] - \mu(x_i), \tag{6}$$

where $F(\cdot|x_i)$ is the CDF of q_i conditional on observed characteristics $X_i = x_i$. Since we assumed that V_i is distributed according to a standard normal distribution, we know that Φ^{-1} is continuous. Thus, C_k is continuous in q_k.

Equation 6 implies that C_k is bounded on $[\underline{q}, \overline{q}] \subset (\delta, 1)$ as long as $F(q|x_i) \in (0,1)$ for all $q \in [\underline{q}, \overline{q}]$. Thus C_k is continuous over the closed interval $[\underline{q}, \overline{q}]$, so C_k is uniformly continuous over $[\underline{q}, \overline{q}]$ (Heine-Cantor Theorem, Aliprantis and Border (2006) Corollary 3.31). That is, given $\epsilon > 0$ there exists $\delta_\epsilon > 0$ such that for all $q_{k-1}, q_k \in [\underline{q}, \overline{q}]$,

$$|q_k - q_{k-1}| < \delta_\epsilon \Rightarrow |C_k - C_{k-1}| < \epsilon. \tag{7}$$

Theorem 3. *Consider a sequence of refinements $\{\Pi^q_K\}^\infty_{K=1}$ of $[\underline{q}, \overline{q}] \subset (\delta, 1)$ defined so that $q^K_1 = \underline{q}$ and $q^K_{K-1} = \overline{q}$ for every refinement in the sequence, and so that the norms of the partitions converge*

[1] We drop the superscripts on C^K_k when the cut point does not depend on the specific partition of quality $\mathcal{P}^q_K$.

to zero. Then $\widehat{V}_i^K$ constructed by linear interpolation converges uniformly to V_i for $V_i \in \big[\mu(x_i) - C_{K-1}, \mu(x_i) - C_1\big]$. In other words:[2]

$$\lim_{K\to\infty}\left(\sup_{\{V_i\,:\,2\le Q_i\le K-1\}}\left|\widehat{V}_i^K - V_i\right|\right) = 0.$$

Proof. Suppose there is a sequence of refinements $\{\Pi_K^q\}_{K=1}^{\infty}$ of $[\underline{q}, \overline{q}] \subset (\delta, 1)$ defined so that $q_1 = \underline{q}$ and $q_{K-1} = \overline{q}$ for every refinement in the sequence, and so that the norms of the partitions converge to zero.[3] Given $\epsilon > 0$, for such a sequence of refining partitions there exists some $N \in \mathbb{N}$ such that $M > N$ implies

$$\max_{k=2,\ldots,M-1}\{|q_k^M - q_{k-1}^M|\} < \delta_\epsilon,$$

which by the (uniform) continuity of C_k with respect to the definition of q_k (Equation 7) implies that

$$\max_{k=2,\ldots,M-1}\{|C_k^M - C_{k-1}^M|\} < \epsilon.$$

Alternatively, this can be stated as:

$$\lim_{K\to\infty}\left(\max_{k=2,\ldots,K-1}\left|C_k^K - C_{k-1}^K\right|\right) = 0$$

for the cut points C_k^K implied by any sequence of refining partitions whose norms converge to 0:

$$\{\Pi_K^q\}_{K=1}^{\infty} \in \left\{ \{\Pi_K^q\}_{K=1}^{\infty} \;\middle|\; \lim_{K\to\infty}||\Pi_K^q|| = 0 \right\}.$$

One approximation of V_i

$$\widehat{V}_i^K \equiv \mu(x_i) - C^K(q_i)$$

is generated by linearly interpolating between the cut points in the partition $\mathcal{P}_K^C$ to construct a

---

[2] Note again that we drop the superscript on C_k^K when the cut point does not depend on the specific partition of quality $\mathcal{P}_K^q$.

[3] Recall that $\mathcal{R}$ is a refinement of partition $\mathcal{P}$ if $\mathcal{P} \subset \mathcal{R}$, and the norm of a partition is defined as

$$||\mathcal{P}_N|| = \max_{n\in\{1,\ldots,N\}}\left|p_n - p_{n-1}\right|.$$

Thus, for example, $||\mathcal{P}_K^C|| = ||\mathcal{P}_K^V||$. An example of a sequence of partitions in $\left\{ \{\Pi_K^q\}_{K=1}^{\infty} \;\middle|\; \lim_{K\to\infty}||\Pi_K^q|| = 0 \right\}$ is the sequence of refining partitions simply dividing each interval in the previous partition in half, so that

$$q_k - q_{k-1} = \frac{\overline{q} - \underline{q}}{2K} \quad \forall \;\; k \in \{2, \ldots, K-1\}.$$

continuous cost function $C^K : [\underline{q}, \overline{q}] \rightarrow \mathbb{R}$ defined by:

$$C^K(q_i) = \left(\frac{C_k^K - C_{k-1}^K}{q_k^K - q_{k-1}^K} \right) (q_i - q_{k-1}^K) + C_{k-1}^K \quad \text{if } q_i \in (q_{k-1}^K, q_k^K].$$

This construction of $\widehat{V}_i^K$ can be made to approximate V_i to any arbitrary degree of accuracy when $V_i \in [\mu(x_i) - C_1, \mu(x_i) - C_{K-1}]$. Since both

$$\widehat{V}_i^K, V_i \in \left(\mu(x_i) - C_k^K, \mu(x_i) - C_{k-1}^K\right] \text{ when } D_i^K = k,$$

it follows that

$$\begin{aligned}
\lim_{K\rightarrow\infty} \left(\sup_{\{V_i \,:\, 2 \leq Q_i \leq K-1\}} \left|\widehat{V}_i^K - V_i\right| \right) &\leq \lim_{K\rightarrow\infty} \left(\max_{k=2,\dots,K-1} \left| [\mu(x_i) - C_{k-1}^K] - [\mu(x_i) - C_k^K] \right| \right) \\
&= \lim_{K\rightarrow\infty} \left(\max_{k=2,\dots,K-1} \left| C_k^K - C_{k-1}^K \right| \right) \\
&= 0.
\end{aligned}$$

□

A.3 Discussion of Intuition

We now seek to illustrate the intuition linking the continuous and discrete models, and how the distributional assumption on V_i acts as a normalization to which the cost function RESPONDS. Suppose there is a continuous measure of neighborhood quality $q \in [\delta, 50]$ for arbitrary $\delta > 0$, and that there are two partitions into discrete levels of quality, where under the first partition

$$Q_i^3 = \begin{cases} 1 & \text{if } q_i \in [q_0, q_1] = [\delta, 10]; \\ 2 & \text{if } q_i \in (q_1, q_2] = (10, 40]; \\ 3 & \text{if } q_i \in (q_2, q_3] = (40, 50], \end{cases}$$

and under the second partition (a refinement of the first partition)

$$Q_i^5 = \begin{cases} I & \text{if } q_i \in [q_0, q_I] = [\delta, 10]; \\ II & \text{if } q_i \in (q_I, q_{II}] = (10, 20]; \\ III & \text{if } q_i \in (q_{II}, q_{III}] = (20, 30]; \\ IV & \text{if } q_i \in (q_{III}, q_{IV}] = (30, 40]; \\ V & \text{if } q_i \in (q_{IV}, q_V] = (40, 50]. \end{cases}$$

Focusing on observed characteristics for a particular realization $X_i = x_i$, we have that

$$\begin{aligned} Pr(q_i^* > 40) = Pr(Q_i^3 = 3) = Pr(q_i^* > q_2) = \Phi(\mu(x) - C_2) \\ = Pr(Q_i^5 = V) = Pr(q_i^* > q_{IV}) = \Phi(\mu(x) - C_{IV}). \end{aligned}$$

Thus $C_2 = C_{IV}$, so the values at the common cut point/knot will be the same under both partitions, with $C(40) = C(40) = C_2 = C_{IV}$. The same logic applies to see that

$$\begin{aligned} Pr(q_i^* > 10) = Pr(Q_i^3 \geq 2) = Pr(q_i^* > q_1) = \Phi(\mu(x) - C_1) \\ = Pr(Q_i^5 \geq II) = Pr(q_i^* > q_I) = \Phi(\mu(x) - C_I). \end{aligned}$$

Thus we will likewise have $C(10) = C(10) = C_1 = C_I$. The partitions and cut points/knots from this example are illustrated below.

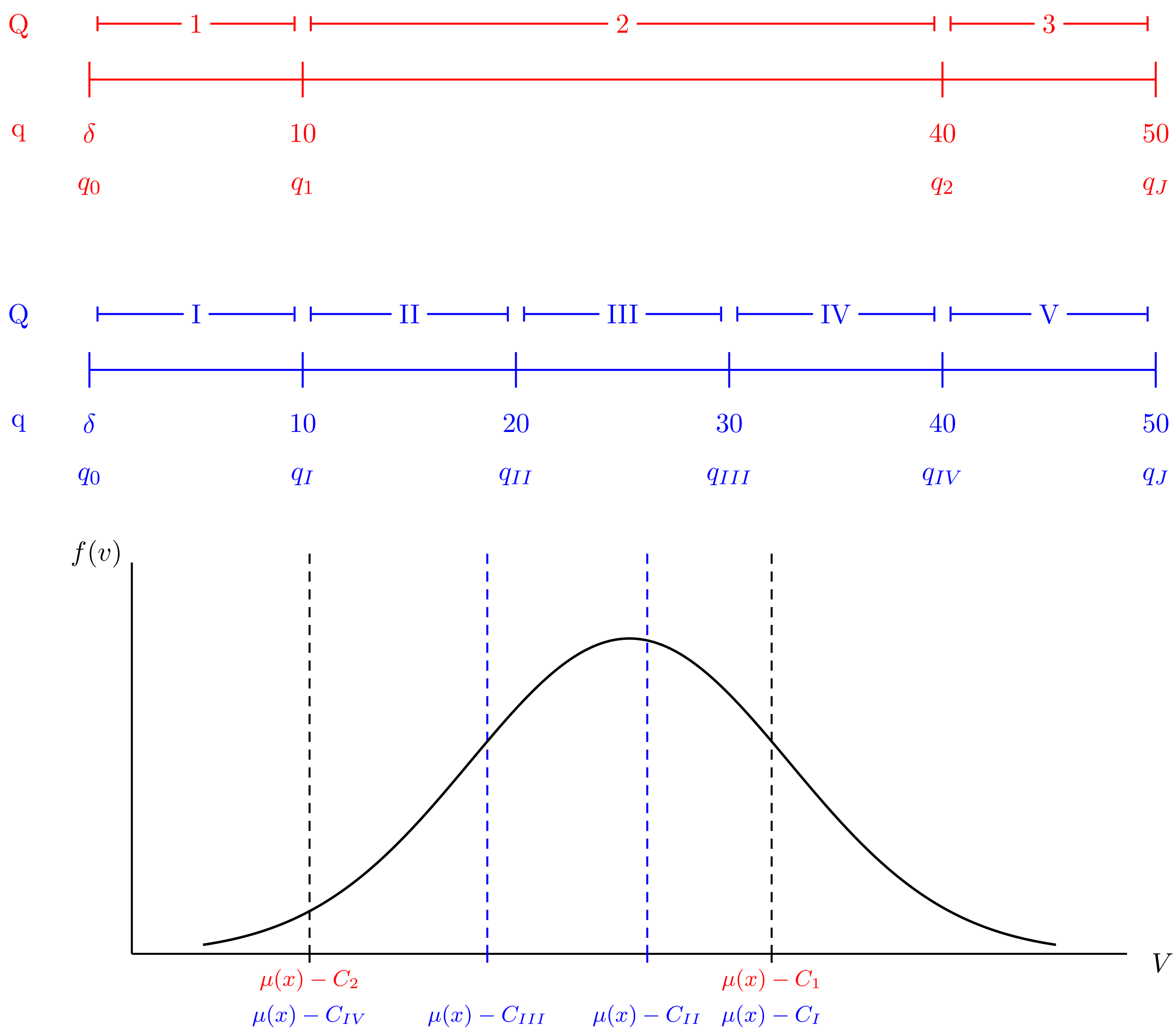

Q
1
2
3
q
δ
10
40
50
q_0
q_1
q_2
q_J
Q
I
II
III
IV
V
q
δ
10
20
30
40
50
q_0
q_I
q_{II}
q_{III}
q_{IV}
q_J
$f(v)$
V
$\mu(x) - C_2$
$\mu(x) - C_1$
$\mu(x) - C_{IV}$
$\mu(x) - C_{III}$
$\mu(x) - C_{II}$
$\mu(x) - C_I$

Figure 1 illustrates the relationship between the linearly interpolated $C^K(q)$ implied by Π^q_K and $C(q)$: When estimated on an infinite sample, the linear interpolation $C^K(q)$ between $C(q_k)$ and $C(q_{k-1})$ for $q \in (q_{k-1}, q_k)$ will approximate the true continuous cost function $C(q)$ to an arbitrary degree of accuracy for all $q \in [q_1, q_{K-1}]$. That is, for all $\epsilon > 0$, there exists some $N \in \mathbb{N}$ such that $M > N$ implies that the norm of the the partition Π^q_M is less than $\delta_\epsilon > 0$, which by the uniform continuity of $C(q)$ and linear interpolation of $C^M(q)$ implies that $\mathbb{E}[\,|C^M(q) - C(q)|\,] < \max_{m \in 2,\ldots,M-1} \mathbb{E}[\,|C(q_m) - C(q_{m-1})|\,] < \epsilon$ for all $q \in [q_1, q_{M-1}]$ when $C^M(q)$ is estimated using the partition Π^q_M.

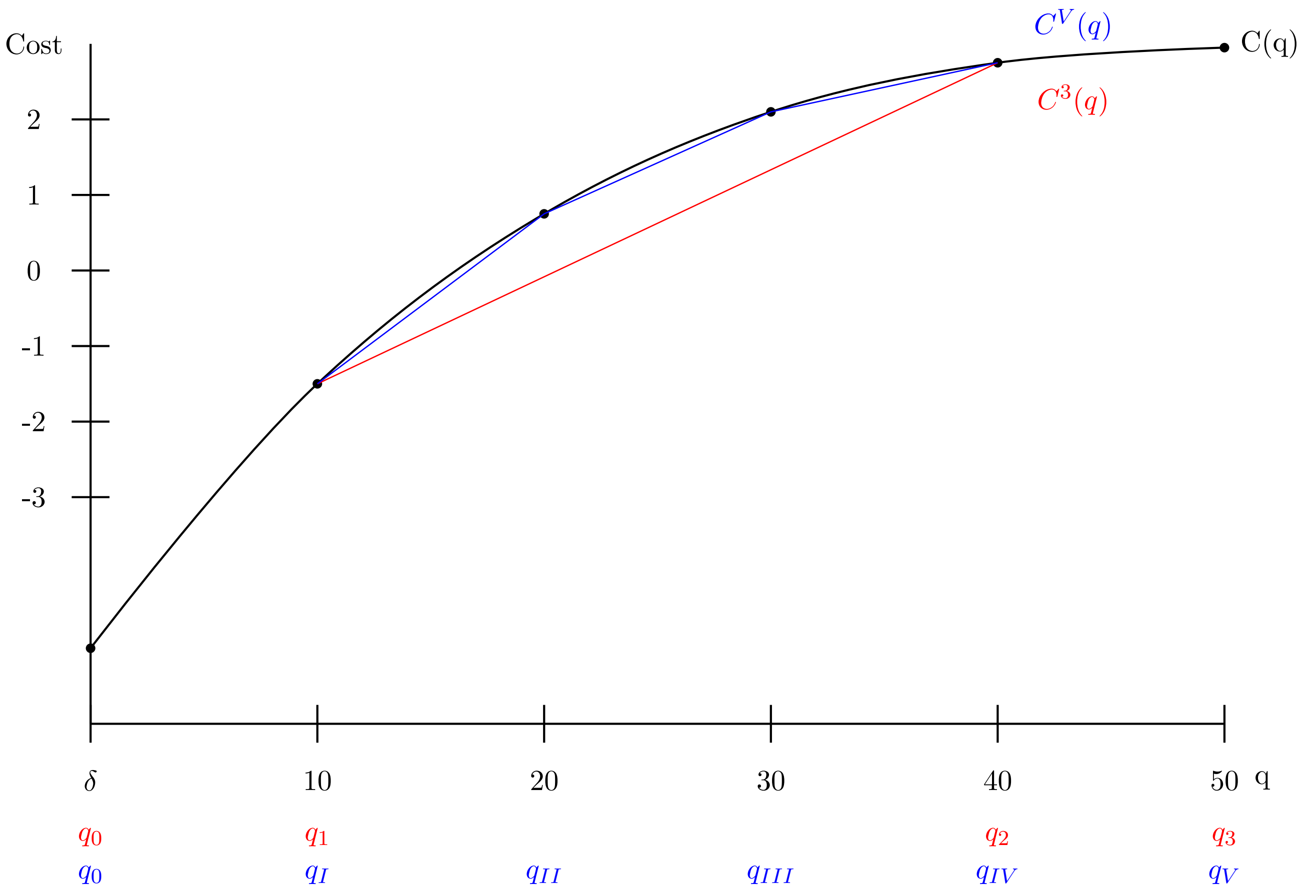


Figure 1: Approximating $C(q)$ to Arbitrary Accuracy in $[q_1, q_{K-1}] \subset (\delta, 50)$

B Specification of the Full Likelihood Function

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:

A6:

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

We stress that the role of Assumption A6 in aiding identification is entirely through the choice model. We also stress that while the normal distribution of V_i may seem like the imposition of a functional form, it is a normalization when the specification of $C(q)$ is flexible.

Assumption A6 implies that the marginal distributions are distributed as follows:

$$V_i \sim \mathcal{N}(0,1), \tag{8}$$

$$(V_i, V_i^S) \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho_S \\ \rho_S & 1 \end{bmatrix} \right), \quad \text{and} \tag{9}$$

$$(V_i, V_i^M) \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho_M \\ \rho_M & 1 \end{bmatrix} \right). \tag{10}$$

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 **??** and the marginal distributions in Equation 8 allow us to express the probability of observing $Q_i = k$ 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

$$\Pr(Q_i = k|X_i, Z_i^S = 0, Z_i^M = 0) = \Phi\big(\ \mu(X_i) - C_{k-1}\ \big) - \Phi\big(\ \mu(X_i) - C_k\ \big) \tag{11}$$

Section 8 Voucher Movers and Non-Movers

$$\begin{aligned}\Pr(Q_i = k|X_i, Z_i^S = 1, Z_i^M = 0, \tau_i^S = 1) = &\ \Phi_2\big(\ \mu(X_i) + \gamma_{k-1}^S - C_{k-1}, \quad \mu^S(X_i); \quad \rho_S\ \big)\\ &- \Phi_2\big(\ \mu(X_i) + \gamma_k^S - C_k, \quad \mu^S(X_i); \quad \rho_S\ \big)\end{aligned}$$

$$\begin{aligned}\Pr(Q_i = k|X_i, Z_i^S = 1, Z_i^M = 0, \tau_i^S = 0) = &\ \Phi_2\big(\ \mu(X_i) - C_{k-1}, \quad -\mu^S(X_i); \quad \rho_S\ \big)\\ &- \Phi_2\big(\ \mu(X_i) - C_k, \quad -\mu^S(X_i); \quad \rho_S\ \big)\end{aligned}$$

MTO Voucher Movers and Non-Movers

$$\begin{aligned}\Pr(Q_i = k|X_i, Z_i^S = 0, Z_i^M = 1, \tau_i^M = 1) = &\ \Phi_2\big(\ \mu(X_i) + \gamma_{k-1}^M - C_{k-1}, \quad \mu^M(X_i); \quad \rho_M\ \big)\\ &- \Phi_2\big(\ \mu(X_i) + \gamma_k^M - C_k, \quad \mu^M(X_i); \quad \rho_M\ \big)\end{aligned}$$

$$\begin{aligned}\Pr(Q_i = k|X_i, Z_i^S = 0, Z_i^M = 1, \tau_i^M = 0) = &\ \Phi_2\big(\ \mu(X_i) - C_{k-1}, \quad -\mu^M(X_i); \quad \rho_M\ \big)\\ &- \Phi_2\big(\ \mu(X_i) - C_k, \quad -\mu^M(X_i); \quad \rho_M\ \big).\end{aligned}$$

These probabilities allow us to identify the parameters of the ordered choice model by expressing its log-likelihood function as:

$$\begin{aligned}\mathcal{LL}(\theta|\mathbf{X}, \mathbf{Z}, \mathbf{Q}, \tau) &= \sum_{i=1}^{N}\sum_{k=1}^{K}\mathbf{1}\{Q_i = k\}\, ln\Big(\Pr\big(Q_i = k\big|X_i, Z_i^S, Z_i^M, \tau\big)\Big)\\ &= \sum_{i=1}^{N^0}\sum_{k=1}^{K}\mathbf{1}\{Q_i = k\}\, ln\Big(\Pr\big(Q_i = k\big|X_i, Z_i^S = 0, Z_i^M = 0\big)\Big)\\ &+ \sum_{i=1}^{N^S}\sum_{k=1}^{K}\mathbf{1}\{Q_i = k\}\, ln\Big(\sum_{t=0}^{1}\Pr\big(Q_i = k\big|X_i, Z_i^S = 1, Z_i^M = 0, \tau_i^S = t\big)1(\tau_i^S = t)\Big)\\ &+ \sum_{i=1}^{N^M}\sum_{k=1}^{K}\mathbf{1}\{Q_i = k\}\, ln\Big(\sum_{t=0}^{1}\Pr\big(Q_i = k\big|X_i, Z_i^S = 0, Z_i^M = 1, \tau_i^M = t\big)1(\tau_i^M = t)\Big).\end{aligned} \tag{12}$$

C 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 $\tau_i^M = 1$ 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.

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 τ_i^M (which is determined by the household's $(\mu^M(X_i), V_i^M)$, and which then determines their $\gamma^M(q)$.). In the restricted version of the model considered in this Appendix for the sake of providing intuition, $\gamma_i^M(q) = \gamma^M(q)$ for all households i for each level of quality q, 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 $\{\gamma^M(q)\}$, 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:

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

where $\underline{u}(Z^M)$ is the minimum u_D at $\mu(x)$ to be induced by receiving the instrument from $D = j$ to $D = j+1$, and $\overline{u}(Z^M)$ is the maximum such u_D at $\mu(x)$.

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 2 and 3. 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

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

The patterns of heterogeneity in response to the instrument allowed under the assumptions in the text ensure that the instrument monotonically increases individuals' latent index, so that $\pi_j^1(X) \geq \pi_j^0(X)$ for all j and for all i. From our assumptions 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):

$$
\begin{aligned}
E[Y|Z=1]-E[Y|Z=0] &= \sum_{j=1}^{J}\int_0^1 \mathbf{1}\{\pi_j^1 \le u_D < \pi_{j-1}^1\}E[Y(D=j)|u_D]du_D \qquad (13)\\
&\quad - \sum_{j=1}^{J}\int_0^1 \mathbf{1}\{\pi_j^0 \le u_D < \pi_{j-1}^0\}E[Y(D=j)|u_D]du_D \\
&= \sum_{j=1}^{J}\Bigg\{\int_{\pi_j^1}^{\pi_{j-1}^1} E[Y(D=j)|u_D]du_D \qquad (14)\\
&\quad - \int_{\pi_j^1}^{\pi_{j-1}^1}\sum_{n=1}^{J}\Big[\mathbf{1}\{\pi_n^0 \le u_D < \pi_{n-1}^0\}E[Y(D=n)|u_D]\Big]du_D\Bigg\}.
\end{aligned}
$$

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

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

allowing us to rewrite Equation 13 as:

$$
\begin{aligned}
E[Y|Z=1]-E[Y|Z=0] &= \sum_{j=1}^{J-1}\int_{\pi_j^0}^{\pi_j^1} \triangle_{j,j+1}^{MTE}(u_D)du_D \\
&= \sum_{j=1}^{J-1}\Big\{\triangle_{j,j+1}^{LATE}(Z^M)\left[\pi_j^1-\pi_j^0\right]\Big\}. \qquad (15)
\end{aligned}
$$

These expressions can be seen in Figures 2-3 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 4 shows that Example I in Figure 2 is just one cross section taken from an interval of the observed, invariant characteristics X (ie, effects for $X = x$).

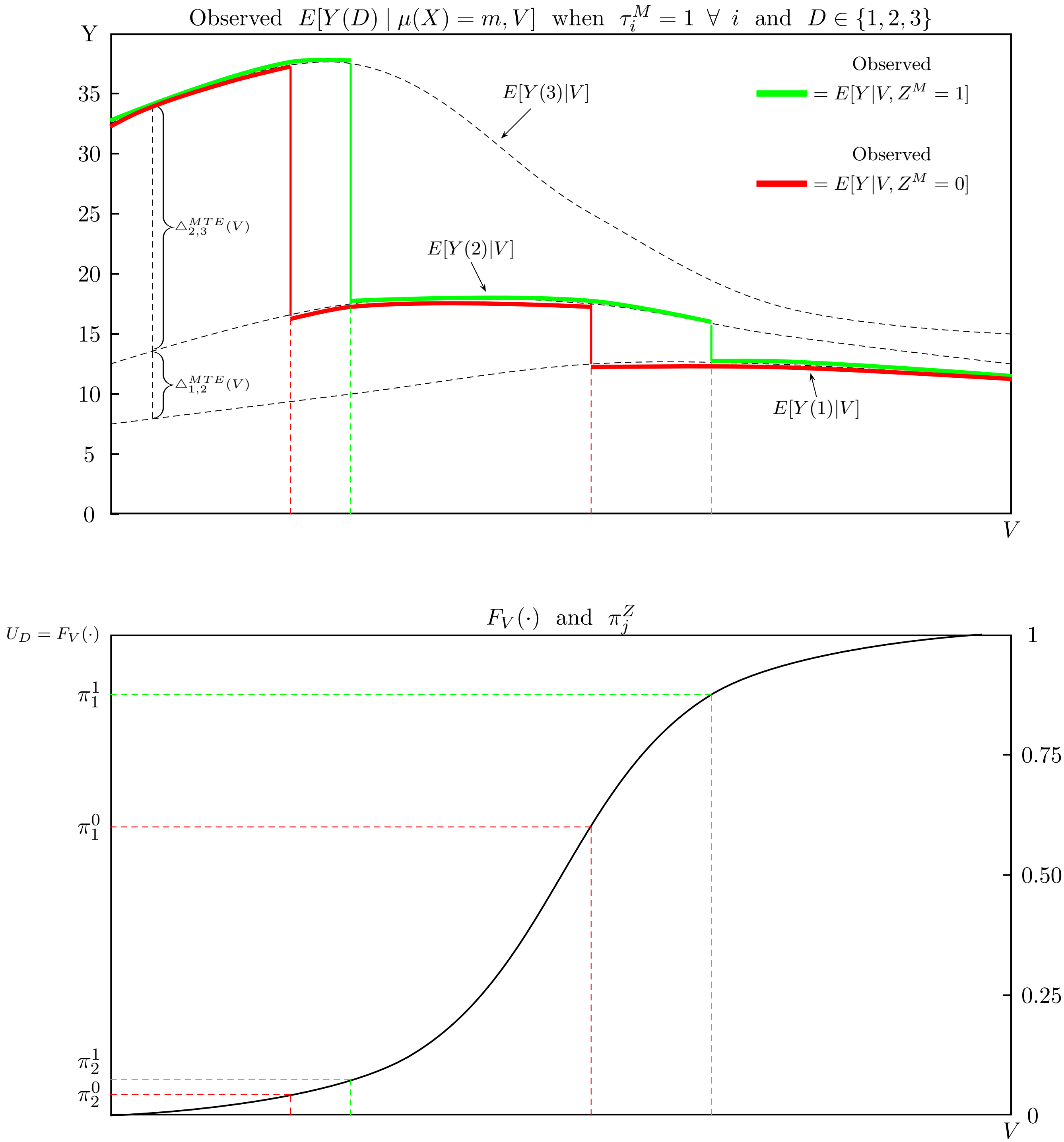


Figure 2: Example I: Potential Outcomes and Marginal Treatment Effects Given $\mu(X) = m$

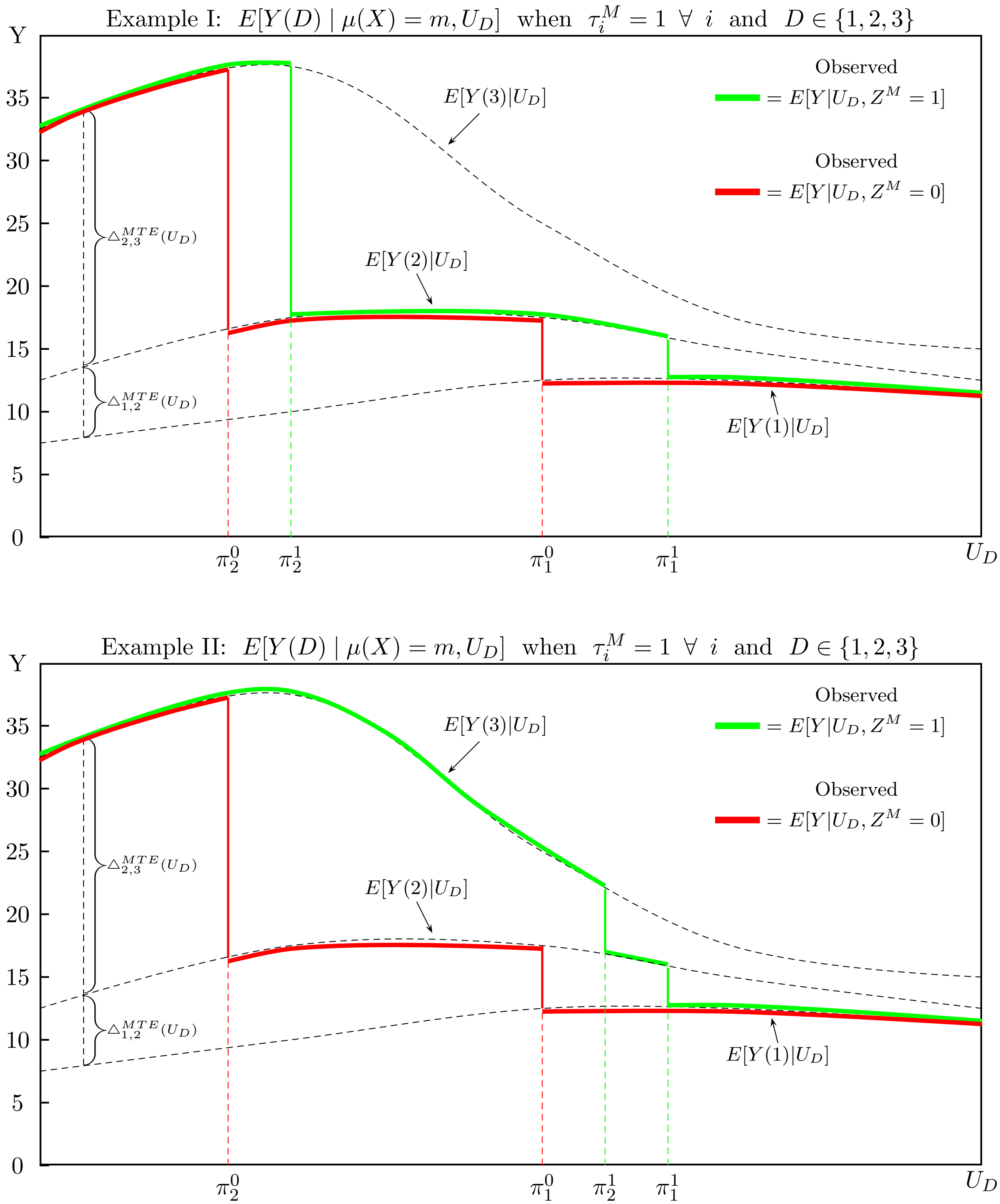


Figure 3: Potential Outcomes and Marginal Treatment Effects Given $\mu(X) = m$

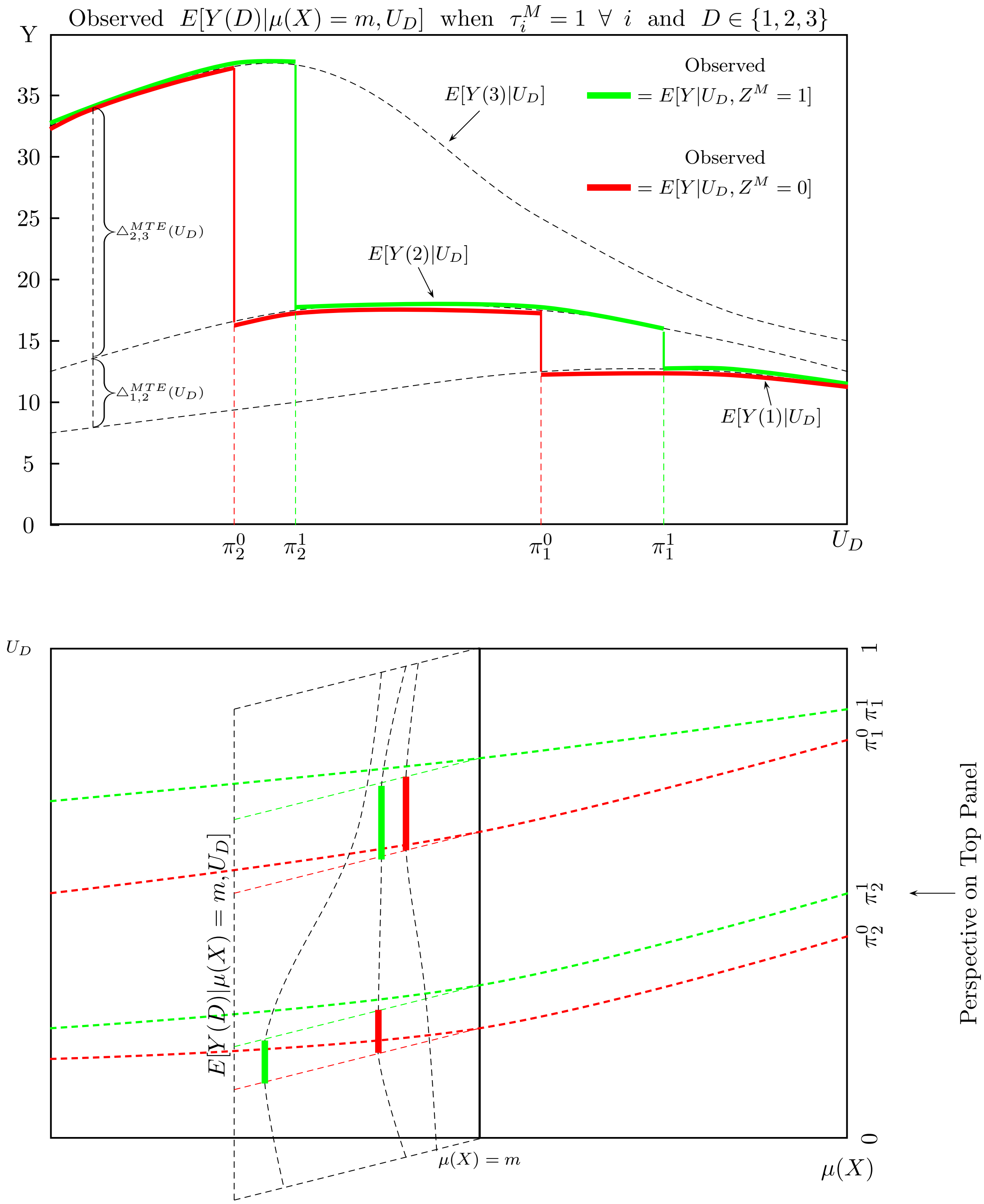


Figure 4: Example I: Potential Outcomes and Marginal Treatment Effects Given $\mu(X)=m$

Now suppose we augment Assumption A5 with:

A5* $\pi^1_j > \pi^0_{j-1}$ for all $j \in \{1, \ldots, J-1\}$.

Example I in Figure 2 shows such an ordering when $J = 3$. Under A5* the right hand side of Equation 15 can be derived quickly since

$$\begin{aligned}
E[Y|Z=1] &= \sum_{j=1}^{J} \int_{\pi^1_j}^{\pi^1_{j-1}} E[Y(D=j)|u_D]du_D \\
&= \int_{\pi^1_1}^{1} E[Y(D=1)|u_D]du_D \\
&\quad + \sum_{j=1}^{J-1} \left\{ \int_{\pi^0_j}^{\pi^1_j} E[Y(D=j+1)|u_D]du_D + \int_{\pi^1_{j+1}}^{\pi^0_j} E[Y(D=j+1)|u_D]du_D \right\}
\end{aligned}$$

and

$$\begin{aligned}
E[Y|Z=0] &= \sum_{j=0}^{J} \int_{\pi^0_j}^{\pi^0_{j-1}} E[Y(D=j)|u_D]du_D \\
&= \int_{\pi^1_1}^{1} E[Y(D=1)|u_D]du_D \\
&\quad + \sum_{j=1}^{J-1} \left\{ \int_{\pi^0_j}^{\pi^1_j} E[Y(D=j)|u_D]du_D + \int_{\pi^1_{j+1}}^{\pi^0_j} E[Y(D=j+1)|u_D]du_D \right\}.
\end{aligned}$$

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

$$\begin{aligned}
E[Y|Z=1] - E[Y|Z=0] &= \sum_{j=1}^{J-1} \left[\int_{\pi^0_j}^{\pi^1_j} E[Y(D=j+1)|u_D]du_D - \int_{\pi^0_j}^{\pi^1_j} E[Y(D=j)|u_D]du_D \right] \\
&= \sum_{j=1}^{J-1} \int_{\pi^0_j}^{\pi^1_j} \triangle^{MTE}_{j,j+1}(u_D)du_D = \sum_{j=1}^{J-1} \left\{ \triangle^{LATE}_{j,j+1}(Z) \quad [\pi^1_j - \pi^0_j] \right\}. \qquad (16)
\end{aligned}$$

Since we can recover $\boldsymbol{\pi}^Z = [\pi^Z_1, \pi^Z_2, \ldots, \pi^Z_{J-1}]$ and we can tell in which $[\pi^0_j, \pi^1_j]$ interval u_{Di} lies from the data, we can estimate these LATEs. The variation in treatment induced by the instrument identifies:

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

and

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

Now suppose that we drop A5* and replace it with the less restrictive original Assumption A5. 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 15 implies

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

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

$$E[Y|u_D \in [a,b], Z=1] - E[Y|u_D \in [a,b], Z=0] = \frac{\int_a^b \sum_{j=n}^{m-1} \triangle_{j,j+1}^{MTE}(u_D) du_D}{b-a} \tag{17}$$

That is, conditional on X_i, the empirical pattern of selection into treatment determines the identification support set $\mathcal{S}_{j,j+1}^M$ 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 $\pi_j^1 > \pi_{j-1}^0$ for some j, the corresponding LATE parameter is still separately identified over the interval

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

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

Comparing Example I and Example II in Figure 3 helps to illustrate how the ordering of the π_j^Z determines identification (ie, the boundaries of the identification support set $\mathcal{S}_{j,j+1}^M$). Since $\pi_2^1 > \pi_0^1$ in Example II, the instrument identifies

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

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

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

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

LATE parameter:

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

Figure 3 shows these LATEs graphically. These identification issues are why determination of the identification support set $\mathcal{S}_{j,j+1}^M$ 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).

D Comparison with Other Identification Strategies

D.1 Using Ignorability to Identify Homogeneous Effects

For the sake of exposition we maintain the assumption from Appendix C that $\gamma_i^M(q) = \gamma^M(q)$ for all households i (ie, that $\tau_i^M = 1$ for all i). One approach to identifying treatment effects would be to strengthen the assumptions in the text. One particular assumption would allow us to estimate average treatment effects over the support of the distribution of the continuous treatment q using the generalized propensity score as developed in Imai and van Dyk (2004) or Hirano and Imbens (2005). However, while relatively standard, the Ignorability assumption necessary for identification in Imai and van Dyk (2004) is restrictive relative to the framework we have adopted that more closely resembles the model in Heckman et al. (2006).[4] When f denotes the distribution of potential outcomes, Ignorability can be written as:

$$f\{y_j|X\} = f\{y_j|X, U_D \in [a,b]\} \text{ for all } j \in \{1,\dots,J\}. \tag{18}$$

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

$$f\{u_j|X\} = f\{u_j|X, U_D \in [a,b]\} \text{ for all } j \in \{1,\dots,J\}, \tag{19}$$

or

$$U_j \perp\!\!\!\perp U_D|X \text{ for all } j \in \{1,\dots,J\}. \tag{20}$$

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

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

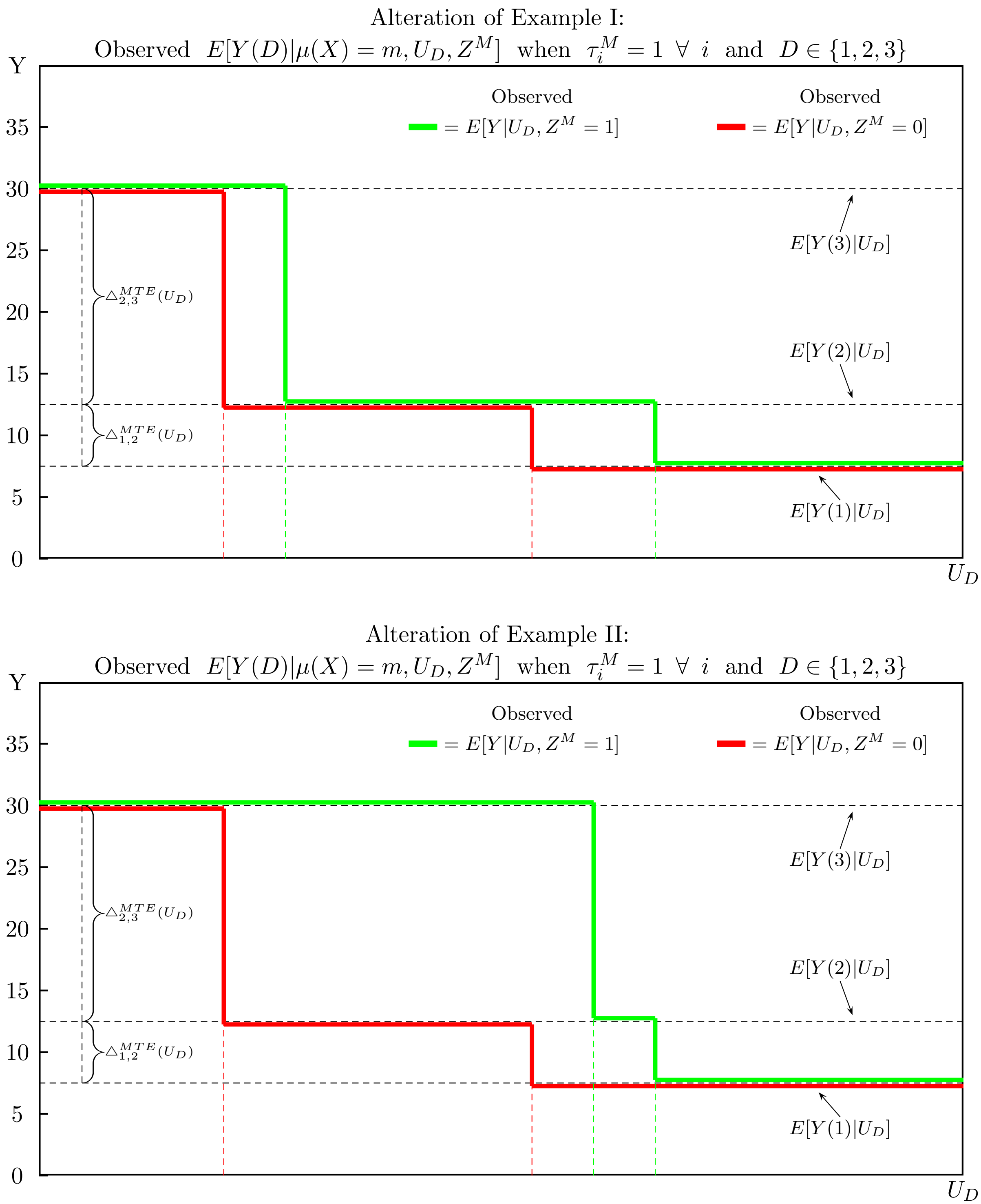


Figure 5: Altering Examples I and II to Satisfy Ignorability

Figure 6 shows a binary example to help illustrate the differences between the heterogeneity in treatment effects allowed under our assumptions in the text and under Ignorability. 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).[5] 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 (2015), Clampet-Lundquist and Massey (2008), Sampson (2008)).

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

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

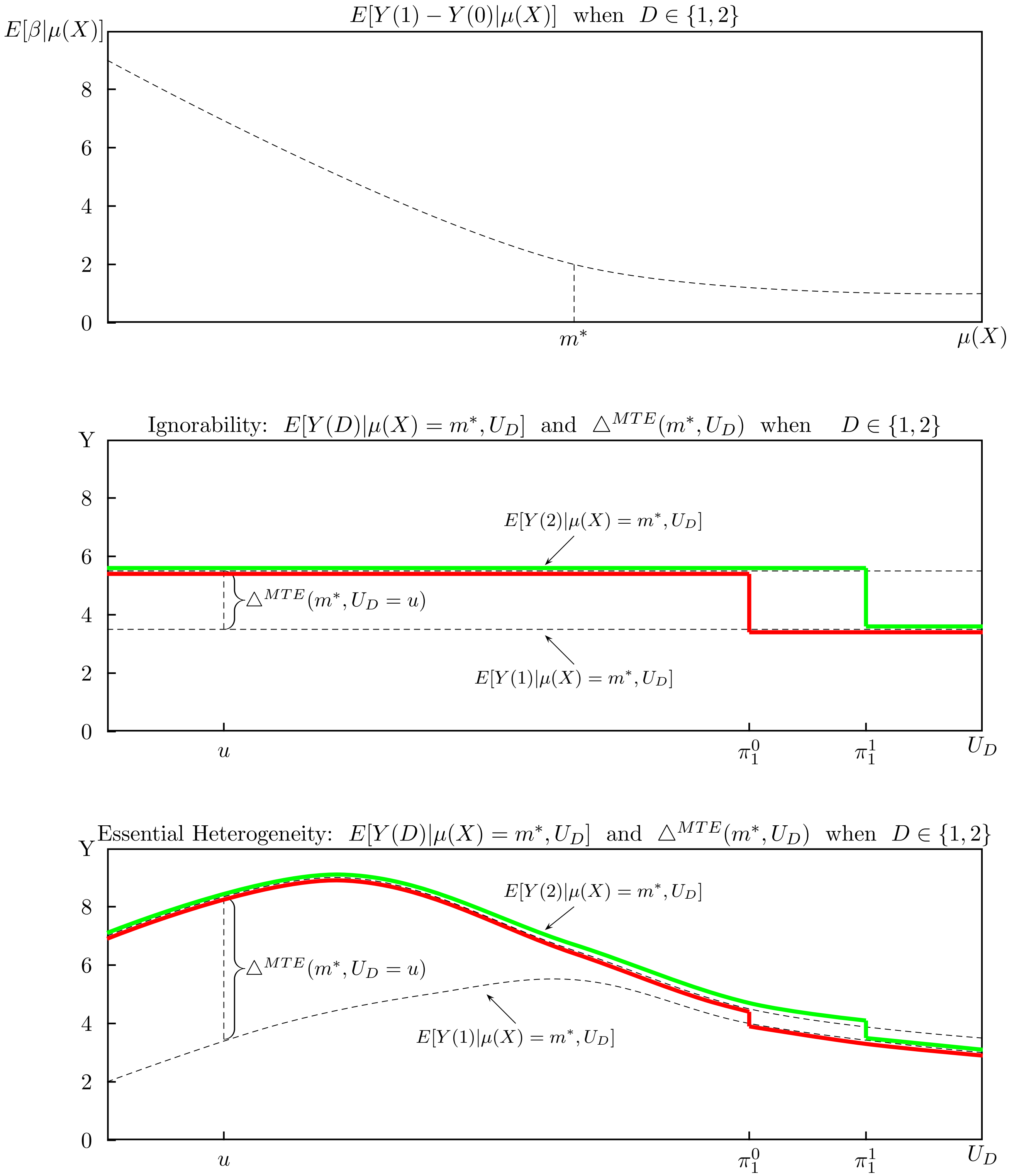


Figure 6: Binary Example with and without Ignorability

D.2 Using Transition-Specific Instruments to Identify Heterogeneous Transition-Specific Effects

Another approach to identifying MTE parameters in this model is to augment our assumptions, 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 $j = 1, \dots, J-1$ such that the distribution of $C_j(W_j)$, conditional on X, Z, and $\{C_h : h \neq j\}$, is nondegenerate and continuous. Under this assumption, which does not hold in our model or empirical application, 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 $\boldsymbol{\pi} \equiv [\pi_1^Z(X), \pi_2^Z(X), \dots, \pi_{J-1}^Z(X)]$, 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 $k \neq j$:

$$\triangle_{j,j+1}^{MTE}(x, \pi_j) = \frac{\partial E[Y|X = x, \boldsymbol{\pi}^Z(x) = \boldsymbol{\pi}]}{\partial \pi_j}. \tag{21}$$

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 21.

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.

D.3 Using a Binary Instrument to Identify the Average of Multiple Heterogeneous, Transition-Specific Effects

Another alternative would be to weaken the assumptions in the text, which would allow us to estimate the Average Causal Response (ACR) parameter introduced in Angrist and Imbens (1995). Under the assumptions in the text 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 Assumption A1 to allow for $\gamma^M(q)$ to be household-specific, 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.

E Interpretation of Neighborhood Choice Model

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.[6] 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:[7]

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

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

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

[6] 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.

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

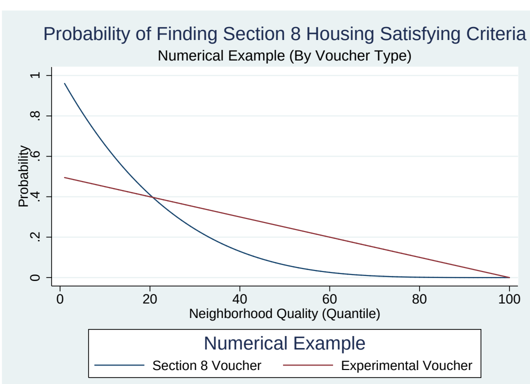
where

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

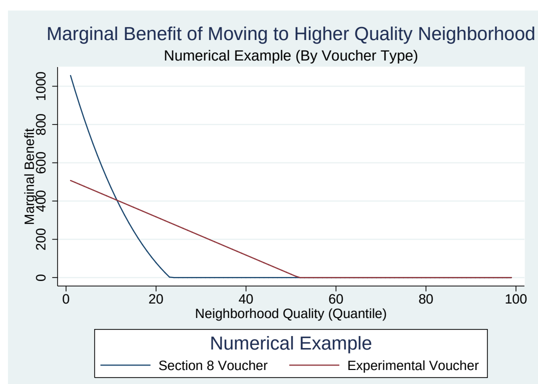
$$E\big[\,NB(q|Z^S=1,Z^M=0)\big] = E\Bigg[\sum_k w^k Y^k(q)\Bigg]$$
$$- \Bigg[Pr(S8|q,Z^S=1)0.30 \times \text{Income} + \big(1 - Pr(S8|q,Z^S=1)\big)E[\text{rent}(q)]\Bigg]$$

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

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

Figure 7a shows two numerical examples of $Pr(S8|q,Z^S=1)$ and $Pr(S8|q,Z^M=1)$ satisfying this qualitative description, and Figure 7b shows the resulting Marginal Benefit functions.[8] We

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

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 $\{C_k\}$ 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 $\{\gamma_k^S\}$ represent elements of the cost function $E[C(q|Z^S = 1)]$ only, and the $\{\gamma_k^M\}$ represent elements of $E[C(q|Z^M = 1)]$. We refer readers interested in the interpretation of these parameters to the discussions on pages 72-78 and 139-150 of de Souza Briggs et al. (2010).

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

References

Aliprantis, C. D. and K. C. Border (2006). *Infinite Dimensional Analysis* (Third ed.). Berlin: Springer.

Aliprantis, D. (2015). Assessing the evidence on neighborhood effects from Moving to Opportunity. *Federal Reserve Bank of Cleveland Working Paper 15-06*.

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*.

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.

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.