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

Dionissi Aliprantis* Francisca G.-C. Richter⋆

October 26, 2018

A Proofs

A.1 The j to $j+1$ LATE Is Identified by the Wald Estimator

Proposition 1. *Assume the model from the text. Consider the set of observed and unobserved characteristics in the ordered choice model that would result in (i) selection into treatment level j when not receiving the instrument, (ii) treatment level j or $j+1$ with the instrument, and (iii) a positive probability of selection into treatment level $j+1$ with the instrument. In the text we define this identification support set first using*

$$\Omega_j \equiv \left\{ \; i \in \Omega \;\middle|\; D_0(i) = j, \;\; D_1(i) \in \{j, j+1\}, \;\; Pr(D_1(i) = j+1) > 0 \; \right\},$$

and then

$$\mathcal{S}_j^M \equiv \left\{ \; \big(\mu(x_i), U_D(i)\big) \;\middle|\; i \in \Omega_j \; \right\}.$$

Applying the Wald estimator to the subsample of experimental and control households in $\mathcal{S}_j^M$ identifies the j to $j+1$ transition-specific LATE:

$$\frac{\mathbb{E}\big[\, Y \;\big|\; Z^M = 1, (\mu(X), U_D) \in \mathcal{S}_j^M \,\big] \; - \; \mathbb{E}\big[\, Y \;\big|\; Z^M = 0, (\mu(X), U_D) \in \mathcal{S}_j^M \,\big]}{\mathbb{E}\big[\, D \;\big|\; Z^M = 1, (\mu(X), U_D) \in \mathcal{S}_j^M \,\big] \; - \; \mathbb{E}\big[\, D \;\big|\; Z^M = 0, (\mu(X), U_D) \in \mathcal{S}_j^M \,\big]} = \triangle_{j,j+1}^{LATE}(Z^M).$$

*: GATE/University of Lyon, *dionissi.aliprantis@gmail.com*

⋆: Case Western Reserve University, +1(216)368-8686, *francisca.richter@case.edu*

Disclaimer: The views stated herein are those of the authors and are not necessarily those of the Federal Reserve Bank of Cleveland or the Board of Governors of the Federal Reserve System.

Proof. Recall that $W_Z^M(i)$ is an indicator for individual i's counterfactual voucher take up. When offered a voucher ($Z_i^M = 1$), a household in $\mathcal{S}_j^M$ can possibly respond with any of the following mutually exclusive options:

1 Not use the MTO voucher and remain in a neighborhood of quality j; we denote this set of households by $S_j^M(W_1^M(i) = 0, D(i) = j)$.

2 Move with the MTO voucher, but not to a higher-quality neighborhood, denoted by $S_j^M(W_1^M(i) = 1, D(i) = j)$.[1]

3 Move with the MTO voucher to a higher quality neighborhood, denoted by $S_j^M(W_1^M(i) = 1, D(i) = j + 1)$.

We can then classify households into non-compliers and compliers as follows:

$$\mathcal{NC}_j^M = S_j^M(W_1^M(i) = 0, D(i) = j) \ \sqcup \ S_j^M(W_1^M(i) = 1, D(i) = j)$$

$$\mathcal{C}_j^M = S_j^M(W_1^M(i) = 1, D(i) = j + 1),$$

where $\sqcup$ represents a disjoint union. We denote the probability of being a complier as $\pi(\mathcal{C}_j^M)$.

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

$$\begin{aligned}
&\frac{\mathbb{E}\left[\, Y \mid Z^M = 1, (\mu(X), U_D) \in \mathcal{S}_j^M \,\right] - \mathbb{E}\left[\, Y \mid Z^M = 0, (\mu(X), U_D) \in \mathcal{S}_j^M \,\right]}{\mathbb{E}\left[\, D \mid Z^M = 1, (\mu(X), U_D) \in \mathcal{S}_j^M \,\right] - \mathbb{E}\left[\, D \mid Z^M = 0, (\mu(X), U_D) \in \mathcal{S}_j^M \,\right]} \\
&\quad = \frac{\mathbb{E}\left(Y_{j+1} - Y_j \mid (\mu(X), U_D, W_1^M) \in \mathcal{C}_j^M \right) \pi(\mathcal{C}_j^M)}{\pi(\mathcal{C}_j^M)} \\
&\quad = \mathbb{E}\left[\, Y_{j+1} - Y_j \mid (\mu(X), U_D, W_1^M) \in \mathcal{C}_j^M \,\right] \\
&\quad \equiv \triangle_{j,j+1}^{LATE}(Z^M),
\end{aligned} \tag{1}$$

[1] One possibility generating this case is that the household moves to a low-poverty neighborhood of the same quality. Another possibility is that because the interim study was conducted four to seven years after randomization, households could have moved more than once, with their final move being to a neighborhood of quality level j.

where the equality in Equation 1 is derived as follows,

$$
\begin{aligned}
&\mathbb{E}[\, Y \mid Z^M = 1, (\mu(X), U_D) \in \mathcal{S}_j^M \,] \;-\; \mathbb{E}[\, Y \mid Z^M = 0, (\mu(X), U_D) \in \mathcal{S}_j^M \,] \\
&\quad = \Big\{ \mathbb{E}[\, Y_{j+1} \mid (\mu(X), U_D, W_1^M) \in \mathcal{C}_j^M \,] \pi(\mathcal{C}_j^M) \quad + \quad \mathbb{E}[\, Y_j \mid (\mu(X), U_D, W_1^M) \in \mathcal{NC}_j^M \,] (1 - \pi(\mathcal{C}_j^M)) \Big\} \\
&\qquad - \Big\{ E[\, Y_j \mid (\mu(X), U_D, W_1^M) \in \mathcal{C}_j^M \,] \pi(\mathcal{C}_j^M) \quad + \quad \mathbb{E}[\, Y_j \mid (\mu(X), U_D, W_1^M) \in \mathcal{NC}_j^M \,] (1 - \pi(\mathcal{C}_j^M)) \Big\} \\
&\quad = \Big(\mathbb{E}[\, Y_{j+1} \mid (\mu(X), U_D, W_1^M) \in \mathcal{C}_j^M \,] \quad - \quad \mathbb{E}[\, Y_j \mid (\mu(X), U_D, W_1^M) \in \mathcal{C}_j^M \,] \Big) \pi(\mathcal{C}_j^M) \\
&\qquad + \Big(\mathbb{E}[\, Y_j \mid (\mu(X), U_D, W_1^M) \in \mathcal{NC}_j^M \,] \quad - \quad \mathbb{E}[\, Y_j \mid (\mu(X), U_D, W_1^M) \in \mathcal{NC}_j^M \,] \Big) (1 - \pi(\mathcal{C}_j^M)) \\
&\quad = \Big(\mathbb{E}[\, Y_{j+1} \mid (\mu(X), U_D, W_1^M) \in \mathcal{C}_j^M \,] \quad - \quad \mathbb{E}[\, Y_j \mid (\mu(X), U_D, W_1^M) \in \mathcal{C}_j^M \,] \Big) \pi(\mathcal{C}_j^M)
\end{aligned}
$$

and likewise,

$$
\begin{aligned}
&\mathbb{E}[\, D \mid Z^M = 1, (\mu(X), U_D) \in \mathcal{S}_j^M \,] \;-\; \mathbb{E}[\, D \mid Z^M = 0, (\mu(X), U_D) \in \mathcal{S}_j^M \,] \\
&\quad = \Big\{ (j+1)\pi(\mathcal{C}_j^M) + (j)(1 - \pi(\mathcal{C}_j^M)) \Big\} \quad - \quad \Big\{ (j)\pi(\mathcal{C}_j^M) + (j)(1 - \pi(\mathcal{C}_j^M)) \Big\} \\
&\quad = \Big[(j+1) - j \Big] \pi(\mathcal{C}_j^M) + \Big[j - j \Big] \big(1 - \pi(\mathcal{C}_j^M)\big) \\
&\quad = \pi(\mathcal{C}_j^M).
\end{aligned}
$$

$\square$

Note that this entire identification strategy is predicated on identifying V_i, as it is required for identifying $\mathcal{S}_j^M$ and applying the Wald estimator to households in this set.

A.2 Equivalence of the Unobserved Component of Choice Specified Under Discrete and Continuous Models of Neighborhood Quality

Suppose that we observe a continuous measure of neighborhood quality $q \in [0, 1]$. If we partition quality to generate discrete treatment levels, there will be an associated ordered choice model. We can estimate the unobserved component V of the ordered choice model using the identification strategy in the text.

We interpret estimates obtained in this way to be estimates of the same random variable V regardless of the partition of quality used in estimation. We justify this interpretation by showing that a sequence of $\{V^n(i)\}_{n=1}^{\infty}$ derived from a sequence of refinements of quality converging in the norm will converge to the random variable $V(i)$ from a continuous model (Corollary 2).

In practice, this result allows us to move freely between different partitions of quality when estimating the ordered choice model and when estimating causal effects on outcomes. 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.

Before stating our results, we will first provide some relevant notation and definitions. Let

$$MB(q) = \mu(x_i) - C(q) - V(i) = 0 \tag{2}$$

be the First Order Condition determining neighborhood quality selection q in a continuous model of choice, where all variables are defined as in the text. Consider a partition of the continuous quality measure $q \in [\alpha, 1] \subset (0, 1]$ into K discrete levels:

$$\mathcal{P}^q(K) = \{\alpha = q_0, q_1, \ldots, q_{K-1}, q_K = 1\}.$$

Partition $\mathcal{P}^q(K)$ defines the discrete treatment

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

Given a continuous cost function $C(q) : [\alpha, 1] \to B \subset \mathcal{R}$, partition $\mathcal{P}^q(K)$ implies a partition of the image of C

$$\mathcal{P}^C(K) = \{C_0, C_1, \ldots, C_{K-1}, C_K\}\,,$$

where $C_k = C(q_k)$.

$\mathcal{P}^q(K)$ and $\mathcal{P}^C(K)$ define a partition of the closed interval $[\mu(x_i) - C_K,\ \mu(x_i) - C_0] \subset \mathbb{R}$:

$$\mathcal{P}^V(K) = \{\mu(x_i) - C_K, \ldots, \mu(x_i) - C_0\}\,.$$

The selection conditions for the discrete model of neighborhood quality choice associated with the continuous model 2 and partition $\mathcal{P}^q(K)$ are:

$$D^K(i) = k \iff \mu(x_i) - C(q_k) < V(i) \leq \mu(x_i) - C(q_{k-1}) \qquad \text{for} \qquad k = 1, \ldots, K. \tag{3}$$

Some definitions we use in our proof are as follows:

Partition-Specific Discrete Model: We say that the ordered choice model given by Condition 3 is the $\mathcal{P}^q(K)$-discrete model associated with the continuous choice model in Equation 2.

Refinement of a Partition: $\mathcal{P}^q_n$ is a refinement of $\mathcal{P}^q(K_0)$ if $\mathcal{P}^q(K_0) \subset \mathcal{P}^q_n$, with $\subset$ representing strict inclusion.

Sequence of Refinements of a Partition: $\{\mathcal{P}^q_n\}_{n=1}^{\infty}$ is a sequence of refinements of $\mathcal{P}^q(K_0)$ if $\mathcal{P}^q(K_0) \subset \mathcal{P}^q_1$ and $\mathcal{P}^q_{n-1} \subset \mathcal{P}^q_n$ for $n = 2, \ldots, \infty$. The n^{th} refinement in the sequence is denoted by $\mathcal{P}^q_n = \{\alpha = q_{0,n}, q_{1,n}, \ldots, q_{N_n,n} = 1\}$, where $N_n + 1$ is the cardinality of the refinement.

Norm of a Partition: The norm of $\mathcal{P}^q_n$ is $\max_{k \in \{1,\ldots,N_n\}} |q_{k,n} - q_{k-1,n}|$.

Convergence in the Norm: The norm of the refinements in the sequence $\{\mathcal{P}^q_n\}_{n=1}^{\infty}$ converges to zero if

$$\lim_{n\to\infty} \max_{k \in \{1,\ldots,N_n\}} |q_{k,n} - q_{k-1,n}| = 0. \tag{4}$$

Proposition 2. *Let the $\mathcal{P}^q(K_0)$-discrete model be associated with the continuous choice model in Equation 2. Define a sequence of refinements of $\mathcal{P}^q(K_0)$ indexed by n, $\{\mathcal{P}^q_n\}_{n=1}^{\infty}$, such that the norm of the refinements converges to zero. The sequence of refinements of $\mathcal{P}^C(K_0)$ generated by $\{\mathcal{P}^q_n\}_{n=1}^{\infty}$ also converges to zero in the norm.*

Proof. We want to show that

$$\lim_{n\to\infty} \max_{k\in\{1,\dots,N_n\}} |C(q_{k,n}) - C(q_{k-1,n})| = 0. \tag{5}$$

This is equivalent to:

$$\forall\ \epsilon > 0, \exists\, m^* \text{ such that } |C(q_{k,n}) - C(q_{k-1,n})| < \epsilon,\ \ \forall\, k \in \{1,\dots,N_n\},\ \text{when } n > m^*. \tag{6}$$

We know that every continuous function on a closed and bounded interval is uniformly continuous (Heine-Cantor Theorem, Aliprantis and Border (2006) Corollary 3.31). Thus, $C : [\alpha, 1] \to B \subset \mathcal{R}$ satisfies:

$$\forall\ \epsilon > 0, \exists\, \delta > 0 \text{ such that } |q - p| < \delta \Rightarrow |C(q) - C(p)| < \epsilon,\ \ \forall\ p, q \in [\alpha, 1]. \tag{7}$$

The condition that the norm of refinements in the series $\{\mathcal{P}^q_n\}_{n=1}^{\infty}$ converges to zero (condition 4) can be restated as

$$\forall\ \delta > 0, \exists\ n^* \text{ such that } |q_{k,n} - q_{k-1,n}| < \delta,\ \ \forall\, k \in \{1,\dots,N_n\},\ \text{when } n > n^*. \tag{8}$$

Take $\epsilon_0 > 0$. By uniform continuity of $C(q)$ there exists $\delta(\epsilon_0)$ that satisfies

$$|C(q) - C(p)| < \epsilon_0,\ \ \forall\ p, q \in [\alpha, 1] \text{ whenever } |q - p| < \delta(\epsilon_0).$$

Given Condition 8, we can find $n^*(\delta(\epsilon_0))$ that satisfies

$$|q_{k,n} - q_{k-1,n}| < \delta(\epsilon_0),\ \ \forall\, k \in \{1,\dots,N_n\},\ \text{when } n > n^*(\delta(\epsilon_0)).$$

Thus, given ϵ_0, we can find $m^* = n^*(\delta(\epsilon_0))$ such that

$$|C(q_{k,n}) - C(q_{k-1,n})| < \epsilon_0,\ \ \forall\, k \in \{1,\dots,N_n\},\ \text{when } n > m^* = n^*(\delta(\epsilon_0)),$$

which satisfies Condition 6. □

Corollary 1. *The sequence of refinements of $\mathcal{P}^V(K_0)$ converges to zero in the norm if it is generated by a sequence of partitions $\{\mathcal{P}^q_n\}_{n=1}^{\infty}$ that converges to zero in the norm.*

Proof. By stating the definition and using the fact that taking the limit is compatible with algebraic operations, we know that

$$\lim_{n\to\infty}\left(\max_{k=1,\dots,K_n}\quad\Big|[\mu(x_i)-C(q_{k-1,n})]-[\mu(x_i)-C(q_{k,n})]\Big|\right)=\lim_{n\to\infty}\quad\left(\max_{k=2,\dots,K-1}\quad\big|C(q_{k,n})-C(q_{k-1,n})\big|\right).$$

By Proposition 2, we know that the right hand side of this equation is equal to 0. ☐

Corollary 2. *For a sequence of partitions $\{\mathcal{P}^q_n\}_{n=1}^{\infty}$ that converges in the norm, a sequence of $\{V^n(i)\}_{n=1}^{\infty}$ satisfying Equation 3 converges to $V(i)$ in Equation 2.*

Proof. By Equation 3 we know that for the $V(i)$ in Equation 2,

$$V(i)\in\big(\mu(x_i)-C(q_{k,n}),\mu(x_i)-C(q_{k-1,n})\big]\;\text{ when }\; D^K(i)=k.$$

By construction, the $V^n(i)$ in our identification strategy is also in this same interval:

$$V^n(i)\in\big(\mu(x_i)-C(q_{k,n}),\mu(x_i)-C(q_{k-1,n})\big]\;\text{ when }\; D^K(i)=k.$$

It follows that

$$\lim_{n\to\infty}\;\big(\big|V^n(i)-V(i)\big|\big)\le\lim_{n\to\infty}\left(\max_{k=1,\dots,K_n}\quad\Big|[\mu(x_i)-C(q_{k-1,n})]-[\mu(x_i)-C(q_{k,n})]\Big|\right).$$

We know from Corollary 1 that the right hand side of this inequality is 0. ☐

A.2.1 Discussion of Intuition

Corollary 2 can be interpreted as the discrete choice conditions of Equation 3 converging to the continuous first order condition of Equation 2. The implication for interpreting the estimates in our model comes from the fact that in practice, $V(i)$ is estimated from a given partition of quality. Corollary 2 assures us that a sequence of $V^n(i)$'s derived from a series of refinements of that partition satisfying Condition 4 will converge to $V(i)$ from the same continuous model, regardless of the initial partition.

We now seek to add intuition to the link between the continuous and discrete models, and to discuss how a distributional assumption on V can be seen as a normalization due to the flexibility of the cost function. Suppose there is a continuous measure of neighborhood quality $q \in [\alpha, 50]$ for arbitrary $\alpha > 0$, and that there are two partitions into discrete levels of quality, where under the first partition

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

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

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

Focusing on observed characteristics for a particular realization $X(i) = x_i$, we have that

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

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

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

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

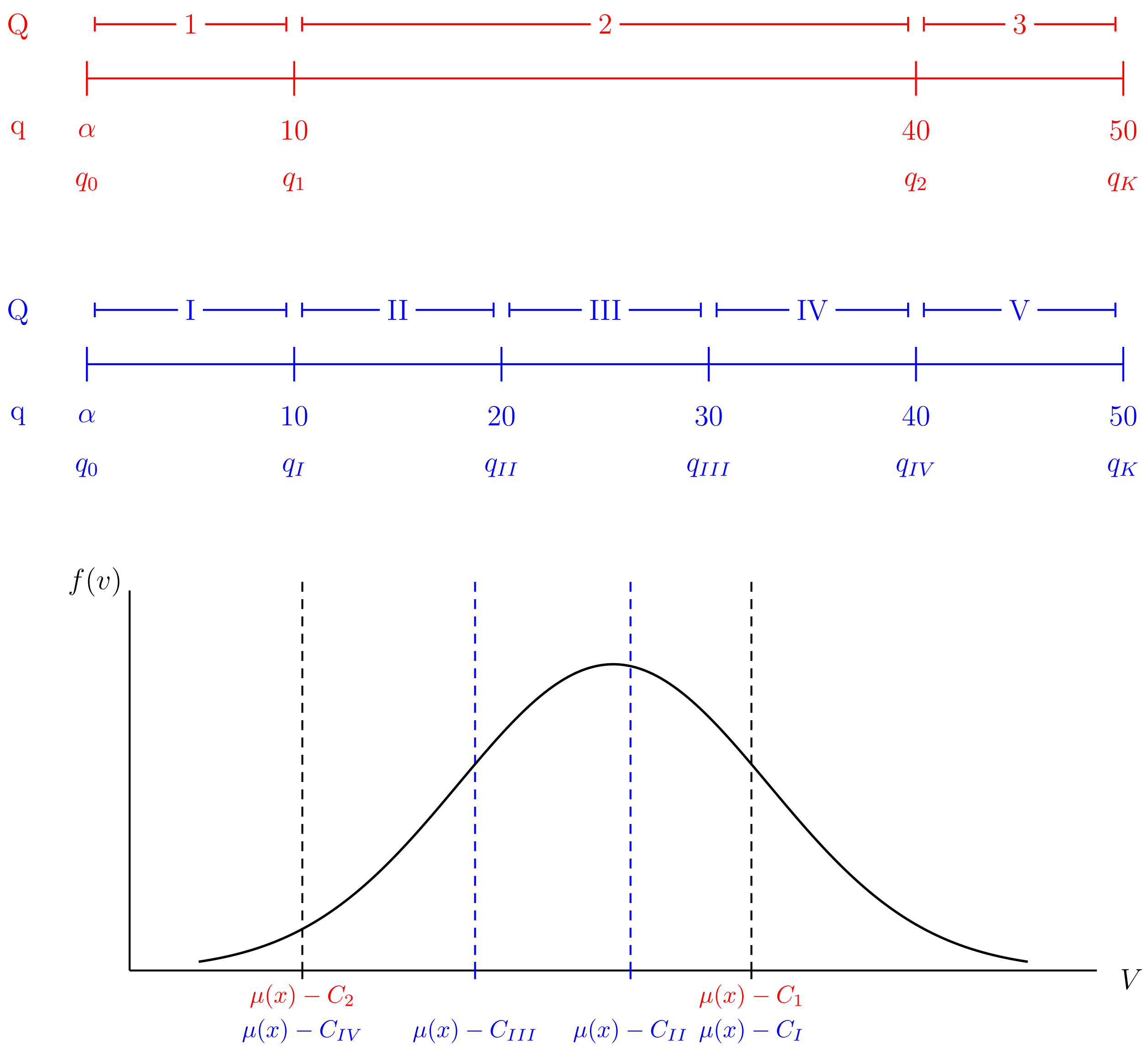


Figure 1 illustrates the relationship between the linear interpolation implied by the partitions $\mathcal{P}_3^q$ and $\mathcal{P}_V^q$ along with $C(q)$. The linear interpolation $C_n(q)$ between $C(q_{k,n})$ and $C(q_{k-1,n})$ for $q \in (q_{k-1}, q_k)$ can be made to approximate the true continuous cost function $C(q)$ to an arbitrary degree of accuracy. That is, for all $\epsilon > 0$, there exists some $m^* \in \mathbb{N}$ such that $n > m^*$ implies that the norm of the partition $\mathcal{P}_n^q$ is less than $\delta_\epsilon > 0$. By the uniform continuity of $C(q)$, if C is increasing, then linear interpolation of $C_n(q)$ implies that $|C_n(q) - C(q)| \leq \max_{k\in\{1,\ldots,K\}} |C(q_{k,n}) - C(q_{k-1,n})|\,] < \epsilon$ for all $q \in [\alpha, 50]$ when $C_n(q)$ is estimated using the partition $\mathcal{P}_n^q$.

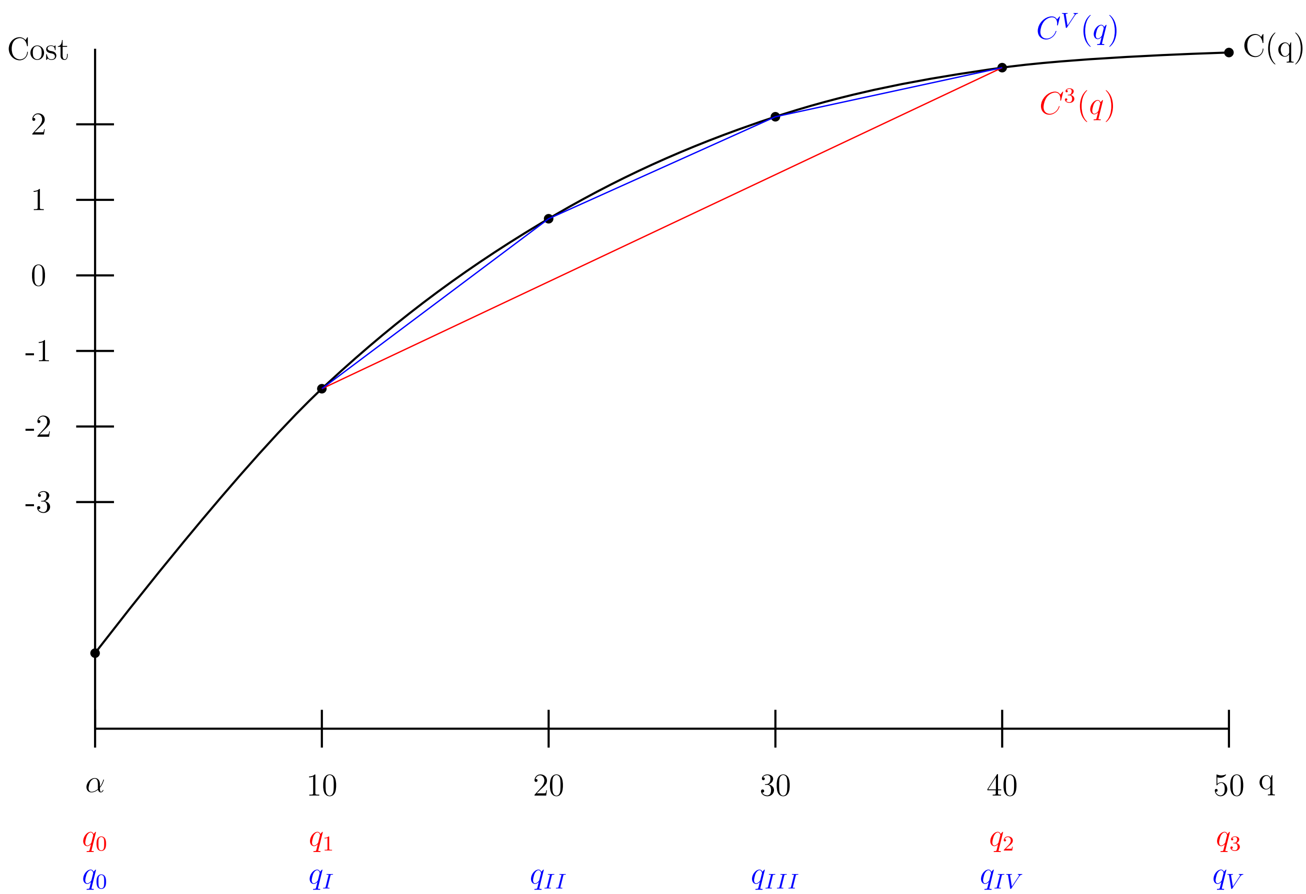


Figure 1: Approximating $C(q)$ to Arbitrary Accuracy

B Estimation Algorithm

The general estimation algorithm is as follows:

Step 1 Estimate the ordered choice model to obtain $\widehat{\mu}(x_i)$, $\{\widehat{C}_k\}$, $\{\widehat{\gamma}_k^S\}$, and $\{\widehat{\gamma}_k^M\}$.

Step 2 Linearly interpolate to obtain $\widehat{C}(q)$, $\widehat{\gamma}^S(q)$, and $\widehat{\gamma}^M(q)$. In the case of $\widehat{C}(q)$,

$$\widehat{C}(q) = \widehat{C}_k + (q - \overline{q}_k)\left(\frac{\widehat{C}_{k+1} - \widehat{C}_k}{\overline{q}_{k+1} - \overline{q}_k}\right) \quad \text{for } q \in (\overline{q}_k, \overline{q}_{k+1}).$$

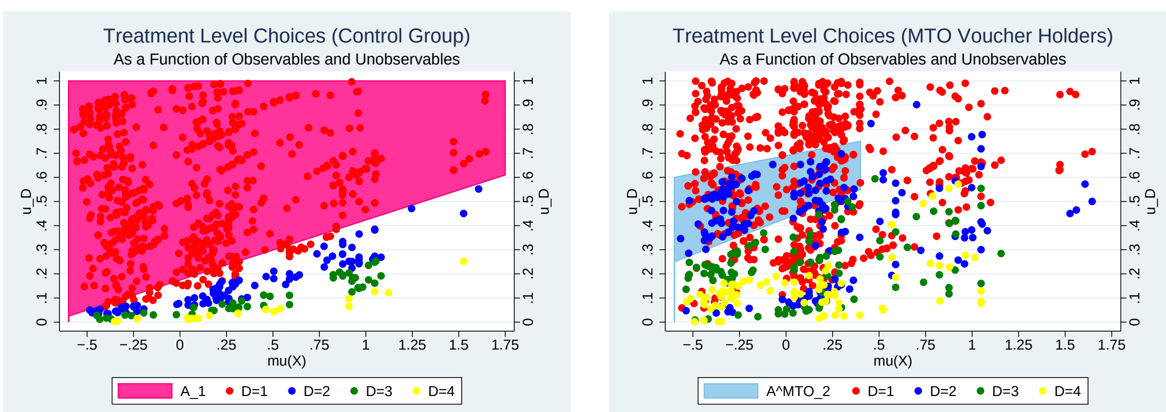
Step 3 Estimate $\widehat{V}$ using the FOC

$$\widehat{V}(i) = \widehat{\mu}(x_i) - \widehat{C}(q_i^*) + \widehat{\gamma}^S(q_i^*)z_i^S w_i^S + \widehat{\gamma}^M(q_i^*)z_i^M w_i^M$$

and $\widehat{U}_D(i)$ via

$$\widehat{U}_D(i) = \Phi(\widehat{V}(i)).$$

Step 4 Using the control group, find an area $\widehat{\mathcal{A}}_j \subset \mathcal{M} \times [0,1]$ such that households with $(\widehat{\mu}(x_i), \widehat{u}_{Di}) \in \widehat{\mathcal{A}}_j$ would select into neighborhood quality $D_i = j$ without any voucher. Using the MTO voucher group, find the subset $\widehat{\mathcal{A}}_{j,j+1}^M$ for which some households would select into neighborhood quality $D_i = j+1$ with an MTO voucher. The identification support set is $\widehat{\mathcal{S}}_j^M \equiv \widehat{\mathcal{A}}_j \cap \widehat{\mathcal{A}}_{j,j+1}^M$.



(a) Control Group and $\widehat{\mathcal{A}}_1$

(b) MTO Voucher Holders and $\widehat{\mathcal{A}}_{1,2}^M$

Figure 2: Selection into Treatment and Counterfactual Areas $\widehat{\mathcal{A}}_1$ and $\widehat{\mathcal{A}}_{1,2}^M$

Step 5 Estimate the j to $j+1$ transition-specific LATE over $\widehat{\mathcal{S}}_j^M$ using the Wald estimator from Equation 1 applied to $\widehat{\mathcal{S}}_j^M$.

Step 6 Bootstrap by repeating 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 3: Calculate $\mathbb{E}[\widehat{\triangle}^{LATE}_{j,j+1}(Z^M)|(\widehat{\mu}(x_i), \widehat{u}_{Di}) \in \widehat{\mathcal{S}}^M_j]$ on the new sample where the set $\widehat{\mathcal{S}}^M_j$ maintains the definition determined in Step 2 for the original sample

Construct standard errors using the T parameter estimates.

C Specification of the Full Likelihood Function

C.1 A Simple Model Illustrating Identification

The only distributional assumption required to identify the parameters of the choice model needed to identify LATE parameters is

$$V(i) \sim \mathcal{N}(0,1). \tag{9}$$

For the sake of exposition, we show this in a model with only the control and MTO groups; the result extends trivially to the case with both standard Section 8 and experimental MTO voucher groups. Recall that $V(i)$ represents the unobserved cost for household i of moving up in the absence of a voucher program and we use $V^M(i)$ to denote unobserved variables influencing household i's take up of the MTO voucher.

Assumptions about the joint distribution $(V(i), V^M(i))$ are only invoked to understand what predicts take up (by identifying parameters of the take-up model). To see why this is the case, we introduce the random variables

$Z^M(i) : \Omega \to \{0,1\}$, the MTO voucher assignment; and

$W^M(i) : \Omega \to \{0,1\}$, an indicator for taking up the MTO voucher.

Observing the realizations of d_i, x_i, z_i^M, and w_i^M for all individuals, we have that

$$\begin{aligned}
\mathcal{LL}(\theta) &= \sum_{i=1}^{N} ln\Big(\Pr\big(d_i \big| x_i, z_i^M, w_i^M\big)\Big) \\
&= \sum_{i=1}^{N}\sum_{k=1}^{K} \mathbf{1}\{d_i = k\}\, ln\Big(\Pr\big(d_i = k \big| x_i, z_i^M, w_i^M\big)\Big) \\
&= \sum_{i=1}^{N}\sum_{k=1}^{K} \mathbf{1}\{d_i = k\}\, ln\Big(\sum_{\zeta=0}^{1}\sum_{\omega=0}^{1} \mathbf{1}\{z_i^M = \zeta, w_i^M = \omega\} \times \Pr\big(d_i = k \big| x_i, z_i^M = \zeta, w_i^M = \omega\big)\Big).
\end{aligned}$$

Our parametric specification of the ordered choice model in Equation 1 in the text along with the normality assumption in Equation 9 allow us to write the choice probabilities as:

$$\begin{aligned}
\Pr\big(d_i = k \mid x_i, z_i^M = 0, w_i^M = 0\big) &= \Phi\big(\mu(x_i) - C_{k-1}\big) - \Phi\big(\mu(x_i) - C_k\big) \\
\Pr\big(d_i = k \mid x_i, z_i^M = 1, w_i^M = 0\big) &= \Phi\big(\mu(x_i) - C_{k-1}\big) - \Phi\big(\mu(x_i) - C_k\big) \\
\Pr\big(d_i = k \mid x_i, z_i^M = 0, w_i^M = 1\big) &= \Phi\big(\mu(x_i) + \gamma_{k-1}^M - C_{k-1}\big) - \Phi\big(\mu(x_i) + \gamma_k^M - C_k\big); \\
\Pr\big(d_i = k \mid x_i, z_i^M = 1, w_i^M = 1\big) &= \Phi\big(\mu(x_i) + \gamma_{k-1}^M - C_{k-1}\big) - \Phi\big(\mu(x_i) + \gamma_k^M - C_k\big).
\end{aligned}$$

No assumption about the distribution of V^M has been invoked here, while we have identified all parameters in the choice model needed to identify LATEs.

C.2 Full Model Likelihood

In the full model we would additionally like to learn about what predicts take-up of vouchers, and in particular, features of the joint distribution of (V, V^S, V^M), where V^S and V^M denote unobserved variables influencing the cost-reduction to household i from a Section 8 and MTO voucher, respectively. To address these issues we invoke additional assumptions about the joint distribution of (V, V^S, V^M). As demonstrated above, we add assumptions about the joint distribution of unobservables only so that we can make some inference about this distribution, not for the sake of identifying LATEs.

In the full model we will also add the standard Section 8 group to the estimation sample. In order to do so, we introduce the random variables

$Z^S(i) : \Omega \to \{0, 1\}$, the standard Section 8 voucher assignment; and

$W^S(i) : \Omega \to \{0, 1\}$, an indicator for taking up the standard Section 8 voucher.

Take-up is now modeled as an endogenous variable, so the likelihood becomes

$$\begin{aligned}
\mathcal{LL}(\theta) &= \sum_{i=1}^{N} ln\Big(\Pr\big(d_i, w_i^S, w_i^M \big| x_i, z_i^S, z_i^M\big)\Big) \\
&= \sum_{i=1}^{N}\sum_{k=1}^{K} \mathbf{1}\{d_i = k\}\, ln\Big(\Pr\big(d_i = k, w_i^S, w_i^M \big| x_i, z_i^S, z_i^M\big)\Big) \\
&= \sum_{i=1}^{N}\sum_{k=1}^{K} \mathbf{1}\{d_i = k\}\, ln\Big(\sum_{\omega^S=0}^{1}\sum_{\omega^M=0}^{1}\sum_{\zeta^S=0}^{1}\sum_{\zeta^M=0}^{1} \mathbf{1}\{w_i^S = \omega^S, w_i^M = \omega^M, z_i^S = \zeta^S, z_i^M = \zeta^M\} \\
&\qquad\qquad \times \Pr\big(d_i = k, w_i^S = \omega^S, w_i^M = \omega^M \big| x_i, z_i^S = \zeta^S, z_i^M = \zeta^M\big)\Big).
\end{aligned}$$

In order to specify the $\Pr\big(d_i = k, w_i^S = \omega^S, w_i^M = \omega^M \big| x_i, z_i^S = \zeta^S, z_i^M = \zeta^M\big)$, we first note that the selection equation in the text, Equation 1, in the full model becomes

$$\begin{aligned}
&D_{Z^S Z^M}(i) = k \iff \qquad (10) \\
&C_{k-1} - W_{Z^S}(i)\gamma_{k-1}^S - W_{Z^M}(i)\gamma_{k-1}^M \le \mu(X(i)) - V(i) < C_k - W_{Z^S}(i)\gamma_k^S - W_{Z^M}(i)\gamma_k^M.
\end{aligned}$$

Recall from the discussion of A2 in the text that to prevent always-takers, which would require that some households take up a voucher without it first being assigned, we assume:

$$W_{Z^S}(i) = \mathbf{1}\{\mu^S(X(i)) - 1,000,000 \times (1 - Z(i)) - V^S(i) \ge 0\} \quad \text{and} \qquad (11)$$

$$W_{Z^M}(i) = \mathbf{1}\{\mu^M(X(i)) - 1,000,000 \times (1 - Z(i)) - V^M(i) \ge 0\} \qquad (12)$$

under the distributional assumption:

$$\mathbf{V}(i) \equiv \big(V(i), V^S(i), V^M(i)\big) \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho^S & \rho^M \\ \rho^S & 1 & \rho^{SM} \\ \rho^M & \rho^{SM} & 1 \end{bmatrix}\right). \tag{13}$$

The marginal distributions implied by this joint distribution are as follows:

$$\begin{aligned} V(i) &\sim \mathcal{N}(0,1), \\ (V(i), V^S(i)) &\sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho_S \\ \rho_S & 1 \end{bmatrix}\right), \qquad \text{and} \\ (V(i), V^M(i)) &\sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho_M \\ \rho_M & 1 \end{bmatrix}\right). \end{aligned}$$

In our data we observe that individuals are assigned at maximum one type of voucher and take up at maximum one type of voucher (ie, $z_i^S = 1$ and $z_i^M = 1$ are mutually exclusive as are $w_i^S = 1$ and $w_i^M = 1$). As well, in our data we do not observe any voucher take up when not assigned a voucher (ie, $w_i^S = 0$ for all i with $z_i^S = 0$ and $w_i^M = 0$ for all i with $z_i^M = 0$). This implies that for the sake of estimation we only need to specify one conditional probability for the control group and two conditional probabilities for each voucher group. Along with our parametric specification of the ordered choice model in Equation 10, the take-up models in Equations 11 and 12 and the joint normality assumption in Equation 13 allow us to write the required conditional choice probabilities as:

Control Group

$$\Pr\big(d_i = k, w_i^S = 0, w_i^M = 0 \mid x_i, z_i^S = 0, z_i^M = 0\big) = \Phi\big(\ \mu(x_i) - C_{k-1}\ \big) - \Phi\big(\ \mu(x_i) - C_k\ \big)$$

Section 8 Group

$$\begin{aligned}
\Pr\big(d_i = k, w_i^S = 0, w_i^M = 0 \mid x_i, z_i^S = 1, z_i^M = 0\big) &= \Phi_2\big(\ \mu(x_i) - C_{k-1},\quad -\mu^S(x_i);\quad \rho_S\ \big) \\
&\quad - \Phi_2\big(\ \mu(x_i) - C_k,\quad -\mu^S(x_i);\quad \rho_S\ \big) \\
\Pr\big(d_i = k, w_i^S = 1, w_i^M = 0 \mid x_i, z_i^S = 1, z_i^M = 0\big) &= \Phi_2\big(\ \mu(x_i) + \gamma_{k-1}^S - C_{k-1},\quad \mu^S(x_i);\quad \rho_S\ \big) \\
&\quad - \Phi_2\big(\ \mu(x_i) + \gamma_k^S - C_k,\quad \mu^S(x_i);\quad \rho_S\ \big)
\end{aligned}$$

MTO Group

$$\begin{aligned}
\Pr\big(d_i = k, w_i^S = 0, w_i^M = 0 \mid x_i, z_i^S = 0, z_i^M = 1\big) &= \Phi_2\big(\ \mu(x_i) - C_{k-1},\quad -\mu^M(x_i);\quad \rho_M\ \big) \\
&\quad - \Phi_2\big(\ \mu(x_i) - C_k,\quad -\mu^M(x_i);\quad \rho_M\ \big) \\
\Pr\big(d_i = k, w_i^S = 0, w_i^M = 1 \mid x_i, z_i^S = 0, z_i^M = 1\big) &= \Phi_2\big(\ \mu(x_i) + \gamma_{k-1}^M - C_{k-1},\quad \mu^M(x_i);\quad \rho_M\ \big) \\
&\quad - \Phi_2\big(\ \mu(x_i) + \gamma_k^M - C_k,\quad \mu^M(x_i);\quad \rho_M\ \big).
\end{aligned}$$

Note that the voucher group probabilities are similar to those in Equation 4 of Greene et al. (2014).

D 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 (2018) 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, for example, one random variable $Y^k(q)$ is the social network one has access to when living in a neighborhood of quality q.[2] 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:[3]

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

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

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

[2]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.

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

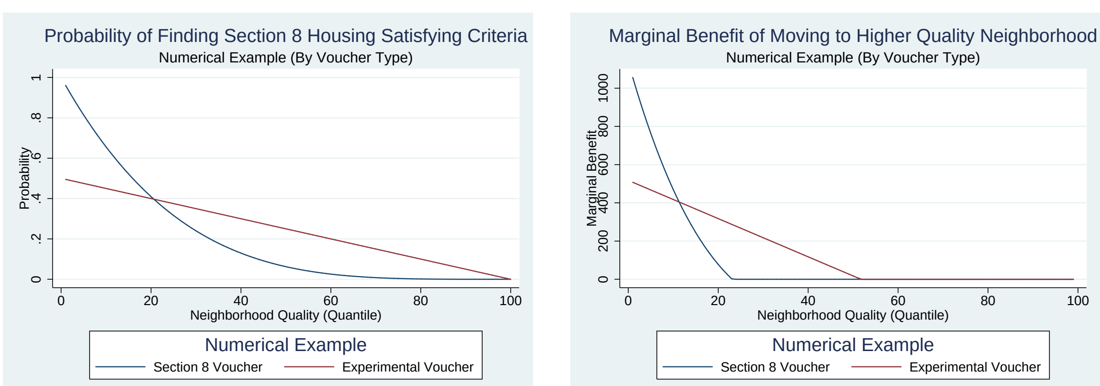
where

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

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

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

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

Figure 3a shows two numerical examples of $Pr(S8|q, Z^S = 1)$ and $Pr(S8|q, Z^M = 1)$ satisfying this qualitative description, and Figure 3b shows the resulting Marginal Benefit functions.[4] We 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).

References

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

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.

Collinson, R. and P. Ganong (2018). How do changes in housing voucher design affect rent and neighborhood quality? *American Economic Journal: Economic Policy 10*(2), 62–89.

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.

Greene, W., M. N. Harris, B. Hollingsworth, and P. Maitra (2014). A latent class model for obesity. *Economics Letters 123*(1), 1 – 5.

[4]The precise parameterization used in this numerical example is: $E[B(q)] = 25,000 + 1,000\,q$; $E[\text{rent}(q)] = 1,000\,q$; $Pr(S8|q, Z^S = 1) = \left(\frac{100-q}{100}\right)^4$; $Pr(S8|q, Z^M = 1) = 0.5\left(\frac{100-q}{100}\right)$; Income $= 15,000$.