CHOICE SET FORMATION IN RESIDENTIAL MOBILITY

AND ITS IMPLICATIONS FOR SEGREGATION DYNAMICS:

ONLINE APPENDIX: MODELING THE PROBABILITY OF CONSIDERING

NEIGHBORHOODS IN A GIVEN PRICE STRATUM AND GEOGRAPHIC REGION

Elizabeth Bruch
Department of Sociology & Population Studies
University of Michigan

Joffre Swait
Erasmus School of Health Policy & Management and
Erasmus Choice Modelling Centre
Erasmus University Rotterdam

March 20, 2019

In this section, we discuss the mathematical specification of the choice set formation process in detail. A person's probability of a given choice set, $Q_n(C_k^*)$, is the product of the probability that mover *n* considers neighborhoods within a given price range (W_{gn}) and the probability that mover *n* considers neighborhoods within a given set of geographic regions and possibly also the current housing unit (Q_{Cn}). We model the probability of considering a particular price stratum, W_{gn}, as a function of household characteristics (H_n) that affect one's propensity to consider neighborhoods in a given price range, independent of geographic location. We use an ordinal logistic model to capture the probability that the n*th* neighborhood seeker considers the g*th* price stratum:

$$W_{gn} = \begin{cases} G\left(\tau_1 - \psi' H_{gn}\right) & \text{if } g = 1 \\ G\left(\tau_g - \psi' H_{gn}\right) - G\left(\tau_{g-1} - \psi' H_{gn}\right) & \text{if } 1 < g < G, \\ 1 - \sum_{g=1}^{G-1} W_{gn} & \text{if } g = G \end{cases} \tag{3a}$$

where

$$G(x) = \frac{1}{1+\exp(-x)} \quad . \tag{3b}$$

Here, the likelihood of consideration is a function of household and affordability characteristics (H_{gn}) that affect one's propensity to consider stratum g independent of location and associated affordability parameters ψ, and τ_g is the propensity threshold for stratum *g.* For example, if H_{gn} includes the ratio of the household income of person *n* to average housing costs in stratum g, and its respective coefficient is found to be positive, this implies that higher income households are more likely to consider neighborhoods containing more expensive housing.

We model the probability of choosing a particular regional choice set, Q_{Cn}, as a function of attributes of its component regions interacted (where appropriate) with attributes of the potential mover. For example, the probability of considering housing in a particular configuration of geographic regions might be a function of the racial composition and distance from the potential mover's place of origin. Note that for estimation purposes we treat the current housing unit as a special region (r_{R+1}).

One challenge in applying CSF models to neighborhood choice is the potentially unwieldy number of regional choice sets. For example, we identify 7 geographic regions in Los Angeles as candidates for consideration that, when combined with the option to stay, give rise to $2^8 - 1 = 255$ unique choice set combinations. A strategy for reducing the total number of choice sets is to pre-specify which combinations of macro-regions are most likely. For example, one might specify that people consider only spatially adjacent macro-regions. However, evidence suggests that—especially for movers with severe income constraints—considered neighborhoods may not be contiguous (Piazzesi, Schneider, and Stroebel 2015; Huff 1986).

Thus, we impose a different restriction: the probability of adding any given macro-region to the choice set does not depend on what other macro-regions are included (net of observable characteristics of macro-regions, such as distance from the origin region). In the absence of more detailed knowledge of spatial neighborhood selection strategies, we believe this "independent availability" assumption is a reasonable starting point (Swait and Ben-Akiva 1987a,b). Formally, the assumption implies that the probability of including a particular configuration of macro-regions in the choice set (where the current housing unit, if included, is its own macro-region) is the product of the individual probabilities of choosing each of those macro-regions.[1] That is, the probability of a given choice set is the product of the individual inclusion probabilities for all included macro-regions times the product of individual exclusion probabilities for all omitted macro-regions, conditioned on a non-empty choice set:

$$Q_{Cn} = \frac{\left(\prod_{r\in C} \pi_{rn}\right)\left(\prod_{l\notin C}(1-\pi_{ln})\right)}{1-\left(\prod_{r\in C}(1-\pi_{rn})\right)} \tag{4a}$$

where

$$\pi_{rn} = \frac{1}{1+\exp\left(-\theta' Z_{rn}\right)}, \quad \text{for } r = 1, \dots, R+1. \tag{4b}$$

Here, π_{rn} is the probability of the n*th* person including macro-region *r*, where Z_{rn} are attributes of macro-regions and/or individuals relevant for choice set formation, and θ are associated weights governing the relationship between these attributes and inclusion propensity. For example, the probability of a given macro-region appearing in the choice set might be modeled as a function of its average income or distance from the place of origin. The current housing unit includes only a single

[1] This assumption also implies that the scale factor for a given geographic region choice set combination is the product of the scale factors for each individual geographic region. In other words, the scale factor associated with regional choice sets, $\lambda_C = \prod_{r\in C} \phi_r$ where ϕ_r denotes the scale factor for the r*th* region.

covariate within Z_{rn}, which is a dummy variable denoting the average propensity for the potential mover to include the current housing unit within the choice set.[2]

[2] Putting it all together, our model adheres to the same basic structure of Manski's theoretical formulation shown in Equation 1. The unconditional probability of choosing a neighborhood is the weighted sum of the probability that a neighborhood is chosen averaged over all possible choice sets: $P_n(i) = \sum_{C_k^* \in \Gamma^*} P_n(i, C_k^*) = \sum_{C_k^* \in \Gamma^*} P_n(i|C_k^*) Q_n(C_k^*)$, where Γ^* is the set of all possible choice sets that have been augmented to include in some cases the option to remain in the current housing unit.

REFERENCES

Huff, J. 1986. "Geographic Regularities in Residential Search Behavior." *Annals of the Association of American Geographers* 76: 208-227.

Piazzesi, M., M. Schneider, and J. Stroebel. 2015. “Segmented Housing Search.” Working Paper No. 20823. National Bureau of Economic Research.

Swait, J. and M. Ben-Akiva. 1987a. Incorporating Random Constraints in Discrete Choice Models of Choice Set Generation, *Transportation Research Part B,* 21, 91-102.

_______. 1987b. Empirical Test of a Constrained Choice Discrete Model: Mode Choice in São Paulo, Brazil, *Transportation Research Part B*, 21, 103-115.