



Methodological Issues in the Analysis of Residential Preferences, Residential Mobility, and Neighborhood Change

Elizabeth E. Bruch¹ and Robert D. Mare²

Abstract

This paper reviews methods for analyzing both individual preferences and choices about where to live, and the implications of these choices for residential patterns. Although these methods are discussed in the context of residential choice, they can be applied more broadly to individual choices in a range of social contexts where behavior is interdependent. We also discuss specific problems with residential mobility data, including the treatment of one's current location as a potential choice, the aggregation of units and the need to take into account variations in neighborhood size, the problem of very large choice sets, and the link between residential mobility and patterns of neighborhood change.

Keywords

Discrete choice modeling, residential segregation, social dynamics

1. INTRODUCTION

This paper reviews methods for analyzing individual preferences and choices about where to live as well as the implications of these choices for residential patterns.¹

¹University of Michigan

²University of California, Los Angeles

Corresponding Author:

Elizabeth E. Bruch, Department of Sociology, University of Michigan, 500 S. State Street, Ann Arbor, MI 48109

Email: ebruch@umich.edu

Residential mobility is a key determinant of the spatial distribution of populations; the segregation of persons who differ in socioeconomic status, race, and ethnicity; and the stability and quality of children's homes and neighborhoods. Patterns of residential choice have implications for the persistence of racial segregation and the concentration of neighborhood poverty. We can use data on residential preferences and mobility to investigate how different characteristics of neighborhoods (e.g., their race-ethnic and economic composition) affect the desirability of that area. Such studies examine either preferences for neighborhood characteristics (as observed in vignette studies such as Farley et al. [1978], Mare and Bruch [2003], and Charles [2005]) or the relationship between neighborhood characteristics and the actual choices made by individuals (Quillian 1999; Crowder and South 2008). We can also use residential choice data to explore the extent to which people's choices are constrained by discrimination, low income, or lack of information (e.g., Pager and Shepherd 2008). Mobility studies combine information on residential choices of individuals with population data on neighborhoods to infer the population dynamics and residential patterns that are implied by the residential preferences and choices of individuals. Such studies may focus on the processes that underpin segregation and population dynamics (Schelling 1969, 2006; Bruch and Mare 2006, 2009) or examine how housing policies, natural disasters, and other exogenous factors affect mobility behavior and population redistribution (e.g., Kingsley and Johnson 2003; Basolo and Nguyen 2005; Clark 2005; Groen and Polivka 2010; Fussell, Sastry, and VanLandingham 2010).

In reviewing methodological issues in the analysis of residential preference and residential mobility, we focus on how individuals respond to the race-ethnic composition of their neighborhoods, although the methods discussed here may be used to model choices based on any dimension of neighborhoods. For the purposes of discussion, we will refer to choices by "individuals," but, with suitable modification, these methods can take into account the fact that households, families, or other social units may make mobility decisions. We review a variety of types of data on residential preferences and mobility and discuss appropriate statistical models for these data. We discuss the analysis of ranked and other types of clustered data; functional form issues; problems of unobserved heterogeneity in individuals and in neighborhoods; and strengths and weaknesses of stated preference data versus observations of actual mobility behavior. We also discuss specific problems with residential mobility data, including the treatment of one's current location as a potential choice; how to specify the choice set of potential movers, the aggregation of units (such as dwelling units into neighborhoods) and the need to take into account variations in neighborhood size; the problem of very large choice sets and possible sampling solutions, and the link between residential mobility and patterns of neighborhood change.

This paper makes several contributions to the existing literature. First, although the basic discrete choice model is a well-known social science tool, sociological studies of residential preferences and residential mobility have made little use of these models. Researchers working in these areas typically employ regression models for the effects of individuals' characteristics on their probabilities of moving to/from a neighborhood or focus only on a single dimension of neighborhoods. These models do not naturally represent residential mobility as a choice that is constrained by

available options and motivated by the differential attractiveness of destinations across multiple dimensions. The analysis of residential mobility requires a number of specific adaptations to the basic choice model that we discuss below. Second, we suggest how the discrete choice framework may be used to develop more behaviorally sophisticated models of residential choice behavior, including how people respond to past experience and neighborhood change. Third, the models discussed in this paper provide a common analytic framework for both actual mobility behavior and stated residential preferences (as typically elicited through vignettes). Finally, we show how statistical models of individual preference and choice provide a foundation for the analysis of aggregate patterns of neighborhood change and segregation.

Our key assumption in this paper is that neighborhood characteristics attract, repel, constrain, and enable individuals of varying kinds to move or stay. The effects of neighborhood characteristics on decisions whether or not to move into neighborhoods are the main focus of analysis. This is in contrast to the more common approach in the sociological literature, which is to emphasize the types of individuals who move into a given neighborhood type (e.g., South and Crowder 1998). Analyses that focus on what types of individuals move into what kinds of neighborhoods are useful for describing group differences in transition probabilities. For a broad array of questions, however, it is preferable to focus instead upon how variation in neighborhood characteristics accounts for population movement. This approach shows not only which individuals are more or less likely to move into different neighborhood types, but also how the moves of individuals lead to changes in neighborhoods, which alter both residential patterns and also the relative attractiveness of neighborhoods for future movers.

Of course, individuals do vary in their preferences for different kinds of neighborhoods. For example, blacks may respond to the proportion of persons in a neighborhood who are black in a substantially different way from how whites respond. Moreover, individuals may have unique responses to neighborhood characteristics that are not measured by characteristics such as their race-ethnicity and socioeconomic status. For analytic purposes, the latter type of variation may be regarded as unobserved random heterogeneity in individuals' responses. Whether systematic or random, however, these kinds of variations enter our models as *interactions* between individual characteristics and the attributes of neighborhoods.

Once a set of residential preference or choice models have been estimated, we may draw inferences about aggregate neighborhood change (e.g., Farley and Frey 1994). In some studies this is done by inspection of the coefficients or predicted probabilities derived from elementary regression models. However, this approach does not take into account the fact that residential mobility evolves dynamically through the interdependent actions of a population of individuals. Individuals as well as households both respond to and also affect the composition of their origin and destination neighborhoods. The set of choices confronted by individuals or households in any moment is generated from the choices of others in the past. For this reason, we present a more elaborate set of methods that link individual choice to aggregate change. The models of residential preference and choice discussed in this paper provide a basis for this type of extrapolation from individual behavior to neighborhood change.

With suitable modification, the methods and analytical models introduced here are more generally applicable to the study of individual choice in a social context. In many instances individuals choose from a set of alternatives, such as the decision to go to college or to take a particular job, the choice of a dating or marriage partner, and decisions to join a social movement or vote in a particular way. In most of these cases, the choices of one person may affect the opportunities and choices of others. Our models are related to other models of social influence that have also been developed for the study of interdependent behavior and social dynamics, including social interaction models for the study of the effects of group or neighborhood membership (Brock and Durlauf 2001) and dynamic models of social networks and group formation (Steglich, Snijders, and Pearson 2010). Our models focus on group (neighborhood) choice by individuals and the aggregate implications of individual choices.

In Section 2 we describe two types of data available to estimate models of residential choice: (1) stated preferences data, based on vignettes, and (2) actual move data, based on mobility histories. In Section 3 we introduce the general discrete choice model for residential choice. In Sections 4 and 5, we detail a range of practical issues that come up when estimating choice models from residential mobility data, including the selection of an appropriate functional form for linking neighborhood characteristics to individual choices, specifying the appropriate geographic units chosen (e.g., neighborhoods, regions of metro areas, housing units), the independence from irrelevant alternatives assumptions, and techniques for exploring how people may evaluate their current place of residence differently from other destinations. In Section 6 we discuss how to incorporate the effect of housing costs (prices) into models of residential choice. Section 7 provides empirical examples of some of the methods discussed in the paper. Section 8 discusses methods for making the link between the residential choices of individuals and aggregate neighborhood change, including agent-based models, interactive Markov models, and general equilibrium models. Section 9 concludes the paper with a brief discussion of future research on methods for the study of residential choice and mobility.

2. TYPES OF DATA

Most studies of residential choice are based on either stated residential preferences or observations of actual residential moves. Stated residential preference data are typically obtained through individuals' interview responses, and they measure their evaluation of or willingness to move into hypothetical neighborhoods that vary along one or more neighborhood characteristics. Actual move data, obtained through residential histories, are reports of the location decisions made by individuals. They reflect both individuals' preferences about where to live and the constraints they face in making residential decisions. Both types of data can be analyzed within a common framework of choice.

2.1. Stated Preferences

An example of stated preference data involves measures of residential race-ethnic preferences from the 1992–1994 Multi-City Study of Urban Inequality (MCSUI)

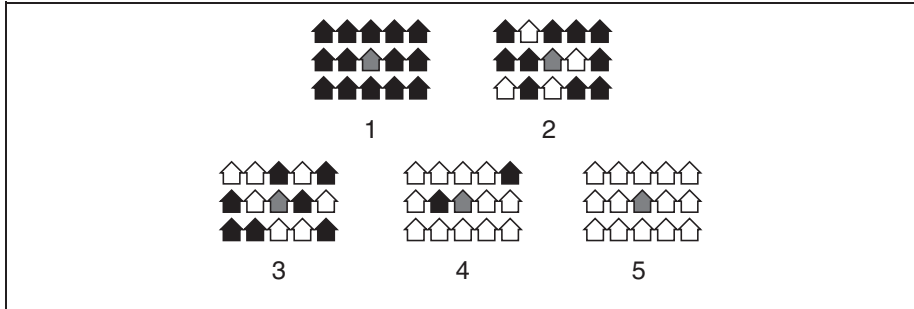


Figure 1. Neighborhood vignettes shown to black respondents asked about white neighbors

(Bobo et al. 2000, appendix D). The MCSUI presented survey respondents with cards depicting five neighborhood vignettes of 14 houses that vary in their race-ethnic composition. The respondent's house is located in the center of the neighborhood. Although the study as a whole examined four groups (whites, blacks, Asians, and Hispanics), each card shows only two groups, the respondent's group and one other group. Figure 1 shows the cards shown to black respondents concerning white neighbors. The survey used a split-ballot design in Boston and Los Angeles, such that each respondent had a one-third probability of being shown a particular vignette out-group. The data include three measures of racial residential preferences. First, for each of the five neighborhood vignettes, each respondent is asked whether he or she would move into that neighborhood. (Whites were asked if they would move out of the neighborhood.) The data consist of five binary responses, each one corresponding to a different proportion own-group and featured out-group. Second, respondents were asked to rank the five vignettes in order of attractiveness. Finally, respondents were given another card with the same configuration of 14 empty houses, but they were asked to assign each house to one of the four race-ethnic groups according to his or her "ideal" neighborhood composition. Exact wording of these three types of questions is shown in Appendix A. The second of these three types of response provide a full ranking of alternatives. The binary responses to the "would you move in/out" question provide a partial ranking of the five neighborhoods. The neighborhoods that the respondent would move into are ranked higher than the ones that the respondent would not move into, but the relative desirability beyond this dichotomy is unknown to the analyst. The ideal neighborhood ethnic configuration response indicates that the chosen configuration is preferred to all other possible configurations, but the relative desirability of the configurations that were not chosen is unknown to the analyst.

These data have been analyzed using a variety of approaches, including descriptive statistics, OLS regression, and categorical response models of various types (e.g., Farley 1978; Farley et al. 1978, 1993, 1994; Charles 2000, 2005; Krysan and Farley 2002). Although these analyses have been informative, they typically do not make full use of information available in the data. In contrast to these approaches,

the discrete choice models proposed in this paper make full use of the quantitative information about race-ethnic composition in these data, allow for full examination of complex interactions among race-ethnic groups, generalize to data that include more dimensions of neighborhood variation than just race-ethnic makeup, and provide a natural comparison to analyses of actual residential choices.

The MCSUI vignettes contain information only on neighborhood racial composition; all other neighborhood characteristics are ignored. Thus, it is difficult to know whether to interpret these data as representing the degree of an individual's true ethnic "tolerance" or a response to other neighborhood characteristics (e.g., crime, school quality, and housing costs) associated with race (Quillian 1995; Harris 1999). Emerson, Chai, and Yancey (2001) use vignette neighborhoods that vary along a number of dimensions: school quality, ethnic composition, property values, and crime rate. They find that, after controlling for non-race/ethnic neighborhood characteristics, whites' aversion to a predominantly Hispanic or Asian neighborhood is no longer statistically significant in their sample, but whites' apparent aversion to black neighborhoods remains. Krysan et al. (2009) construct video vignettes that vary the race of actors portraying neighborhood residents, but they hold constant key visual indicators of the socioeconomic composition of neighborhoods (e.g., the upkeep of yards and the types of cars in driveways). Multidimensional vignette data in principle allow the analyst to "control for" any potential confounding neighborhood characteristics. However, it is hard to represent multidimensional neighborhoods using pictures, and complex verbal descriptions are difficult for respondents to understand. A more straightforward way of exploring how multiple factors affect residential choice is to use data on actual moves.

2.2. Mobility Histories

Residential choices and preferences may also be observed in actual mobility behavior. Information about mobility and neighborhood choice may be obtained from cross section data, such as the U.S. Decennial Census, which documents both current neighborhood of residence and also year moved into current unit (to identify recent movers). Alternatively, mobility data may come from retrospective survey questions that ask individuals to recall their previous addresses over some specified time period. For example, wave 1 of the Los Angeles Family and Neighborhood Survey (LA FANS) asked individuals to report all moves and addresses lived in over the past two years, and wave 2 asked for a residential history between wave 1 and wave 2 (Sastry et al. 2006). Residential mobility data may also be prospective, identifying respondents at the beginning of a time period and tracking their subsequent moves.

For example, the Panel Study of Income Dynamics (PSID) records where a respondent lives at the time of each interview. The population represented by a set of mobility data, of course, depends on the survey instrument. For example, the data may be nationally representative data, as in the Census or the PSID, or data focused on a particular metropolitan area, as in the LA FANS.

Several studies have used the PSID panel data to examine neighborhood mobility. Some treat the decision to move out of one's current neighborhood (e.g., analyses of "white flight") as a binary outcome variable (e.g., South and Crowder 1997; Rosenbaum and Friedman 2001), whereas others use the demographic (typically race-ethnic) composition of the destination neighborhood as a polytomous or quantitative outcome variable (Crowder, South, and Chavez 2006; Crowder and South 2008). The outcome is often characterized by its racial composition (e.g., its percentage of white, black, or Hispanic). Typically the outcome is modeled using a binary logit (did or did not move out) or multinomial logit (with destinations categorized into types). The goal of these analyses is to predict choice of destination conditional on individual and/or household characteristics, characteristics of the current residential census tract, and characteristics of the metropolitan area as a whole.

Although these studies usefully describe mobility among neighborhood types and covariates of this mobility, they are ill-suited to the study of residential decision-making by individuals and the impact of these decisions on segregation or other aspects of population distribution. Whereas analyses of mobility rates among neighborhoods with varying percentages of a given ethnic group examine only a single dimension of destination neighborhoods, households evaluate potential destination neighborhoods on several dimensions—for example, racial composition, economic level, housing price, and school quality—when making residential decisions. Any single dimension, when considered by itself, may be confounded with other distinct but correlated dimensions. Additionally, these studies only allow respondents' own characteristics, characteristics of their current neighborhood, and the racial composition of the *chosen* tract to affect destinations, omitting the possible effects of the comparative characteristics of potential destinations on mobility decisions. As we show below, a fruitful alternative approach is to adapt models for discrete choice to the analysis of residential decision making. This approach incorporates the effects of both neighborhood and individual characteristics on residential location choice, a multidimensional approach to measuring neighborhood attractiveness, and a natural way to extrapolate to aggregate neighborhood change. Additionally, it allows us to examine both stated preferences and actual mobility decisions within a common analytic framework.

2.3. Stated Preferences Versus Mobility Histories

Stated preference (vignette) and mobility history data have several complementary strengths and weaknesses. The most important advantage of stated preference data is that the hypothetical characteristics of neighborhoods are under the control of the investigator. Thus, it is possible to assign descriptions of neighborhoods that vary along one or more dimensions to different individuals or to administer to the same individual an array of possible neighborhood configurations. Randomization combined with observations of repeated choices can control for unmeasured differences among individuals. This is a relatively low-cost means of data collection inasmuch as it does not require the collection of residential mobility histories or large samples

of individuals, only a fraction of whom have moved in the recent past. It also allows for the specification of relatively rare types of neighborhoods that would otherwise require an extremely large sample of actual moves. Furthermore, stated preference designs elicit individuals' preferences; in theory these preferences are unconstrained by affordability constraints, housing supply, discrimination, and other factors that affect actual moves.

The weaknesses of neighborhood vignettes arise because they are administered in interviews, which poorly approximate the contexts in which actual choices are made. First, preference for neighborhoods that vary in their racial makeup is potentially a sensitive subject and thus respondents may express socially desirable preferences. Second, vignettes are typically administered to individuals, but mobility decisions may be made collectively by multiple household members. Third, it is usually impractical to vary more than two or three dimensions of neighborhood desirability in vignette studies (e.g., racial makeup, poverty rate, age of housing), precluding the investigation of complex interactions among determinants of housing desirability (Harris 1999). Fourth, because neighborhood vignettes are hypothetical, stated preferences abstract from the virtually limitless array of alternatives that people may have in a real choice situation, as well as their substantial proclivity not to move (that is, to choose their current residence) as a result of the search and moving costs. Finally, as discussed further in Section 7, stated preferences may be sensitive to how interview questions are phrased.

Actual mobility histories also have their own advantages and disadvantages. On the one hand, they provide true measures of real mobility decisions, albeit subject to constraints. Additionally, because they measure choices made by heterogeneous individuals for neighborhoods that vary in a wide range of attributes, they allow the analyst to represent mobility using a rich set of individual and neighborhood covariates. Finally, probability samples of individuals and households include both movers and nonmovers and, in individual mobility histories, periods of stable residence as well as episodes of mobility. This enables the analyst to examine differences in how decision makers evaluate their own locations relative to other potential destinations, and to thus explore how origins as well as destinations affect choice.

On the other hand, actual moves are not pure measures of residential preferences. Rather, they result from preferences about desired locations in the context of constraints on residential options. If the analyst can specify the true choice set for each individual, this will reduce the extent to which constraints dominate the choice process. In practice, however, we seldom know an individual's true range of alternatives. Additionally, mobility histories are comparatively expensive to collect. Because recent mobility is usually a relatively rare event, large amounts of data must be collected, whether through lengthy retrospective mobility histories, long prospective panels, or shorter residential histories obtained from large samples of individuals. The need for large numbers of observations is exacerbated, moreover, when the analyst wishes to look at the selection of relatively rare neighborhoods.

In principle, we can combine the strengths of stated and revealed preference data by pooling them into one model. Louviere, Hensher, and Swait (2000) discuss this

possibility for studying consumer choice. To our knowledge, this approach has not yet been taken in the study of residential choice.

3. DISCRETE CHOICE MODELS

Discrete choice models represent behavior in which individuals choose one or more options from a set of given alternatives, typically under the assumption that they select the option(s) with the greatest utility. Ben-Akiva and Lerman (1993), Louviere et al. (2000), and Train (2003) discuss each of these models in detail. In this section we review their essential properties before discussing the special adaptations required for the study of residential mobility. Our discussion builds on the work of McFadden (1978), who first applied discrete choice models to the study of location decisions. In discrete choice models of residential mobility, the choice set may consist of housing units, neighborhoods, or other potential destinations. The outcome of interest is the specific location chosen, given the set of available alternatives. Although our discussion typically refers to the choices of individuals, in practice the choosers may be individuals, families, households, or other decision makers.

3.1. Residential Mobility as a Market Process

In most of the models discussed below, we represent residential choice as a “demand-side” process whereby individuals or households select from an array of possible destinations. This is a partial view of residential mobility inasmuch as moves in fact result from interactions between buyers and sellers or landlords and renters who negotiate the exchange of housing units. Discrete choice models capture housing demand conditional on housing supply, but these models do not represent how the actions or motivations of housing suppliers (e.g., the steering decisions of real-estate agents, the lending decisions of banks, or the building decisions of developers) affect the number and type of available units. For the limited purpose of analyzing individual choice, it suffices to assume that housing vacancies and housing prices are given and a one-sided approach is sufficient. For studying the realistic aggregate dynamics of a housing market, it may be necessary to take the supply as well as the demand side of the market into account. In later sections, we discuss how to incorporate prices into models of *individual* residential choice and to use price equilibrium assumptions to assess the effects of changes in *aggregate* demand. (An alternative modeling strategy is to model explicitly the interactions between housing suppliers and housing seekers. Such models could rely on optimal matching of housing seekers and providers [e.g., Roth and Sotomayor 1990] and use extensions of available “two-sided” statistical models for joint decisions of actors on both the supply and demand sides of a market [Logan 1996, 1998; Logan, Hoff, and Newton 2008]. Specification and implementation of such a model for housing markets is beyond the scope of this paper.)

Table 1. Data Structure Used in Estimation of Discrete Choice Models

ID	Own Race	City	Neighborhood	Other Race	Rank	Move
1	black	Detroit	0	white	4	0
1	black	Detroit	29	white	2	1
1	black	Detroit	50	white	1	1
1	black	Detroit	86	white	3	1
1	black	Detroit	100	white	5	0
2	white	Detroit	0	black	.	1
2	white	Detroit	7	black	.	1
.
.

Note: The variables are defined as own-race (respondent's own race), city (survey city), neighborhood (proportion other-race), other-race (other racial group in vignette neighborhood), rank (attractiveness ranking assigned to vignette neighborhood), and move (whether or not the respondent would move into this vignette neighborhood).

3.2. Outcome Variable and Data Structure

In discrete choice models, the outcome is either a single choice (representing the “best” possible outcome given available opportunities) or a set of ranked choices. Rankings contain more information on preferences than single choices, which reveal the top-ranked choice but not the relative desirability of the remaining options. In data on actual choices, we typically observe only a single choice (or a series of choices made over some period of time). In stated preferences, respondents may rank neighborhoods in order of desirability. The models used to estimate parameters based on these two outcomes are similar, except that the ranked outcome model includes additional elements to the likelihood function, one for each ranking given the current set of unranked items. We discuss this in more detail below.

Table 1 shows the data structure for estimating discrete choice models. Each of the I individuals has J lines of data, one for each of potential destination alternatives. We refer to each line of data as an “individual-alternative” and the set of J alternatives as the individual’s choice set. In the example shown in Table 1, $J = 5$ for all individuals, but in general it is possible for the size of choice set to vary across individuals. Individual characteristics (X_i) are constant within individuals, but features of neighborhood alternatives (Z_j), such as neighborhood proportion own-race, vary across alternatives within individuals.

3.3. Conditional Logit Model

Let Y_{ij} be an indicator variable denoting which neighborhood (indexed by j) is chosen by the i th individual ($i = 1, \dots, I; j = 1, \dots, J$). Let U_{ij} denote the (latent) utility or attractiveness that the i th individual attaches to the j th neighborhood. Let p_{ij} denote the probability that the i th individual chooses the j th neighborhood. The utility of a

neighborhood for an individual depends on neighborhood characteristics, possibly interacting with characteristics of individuals. These characteristics may or may not be known by the researcher, but they are known to the individuals to whom they apply. Let \bar{Z}_j be a vector of observed (to the analyst) characteristics of the j th neighborhood (e.g., the race-ethnic makeup of the neighborhood). Let \bar{X}_i denote a vector of observed characteristics of the i th individual or household. These characteristics include fixed demographic characteristics such as race and sex, and time-varying characteristics such as income, employment status, housing roster, and residential history. Let ϕ_{ij} represent the contribution of unobserved attributes of individuals and potential neighborhoods to utility. The attractiveness of neighborhoods is represented as

$$U_{ij} = F(\bar{Z}_j, \bar{X}_i, \phi_{ij}). \quad (1)$$

If F is a linear random utility model, then, for example, for a single observed neighborhood and personal characteristic (Z and X respectively), the model is

$$U_{ij} = \beta Z_j + \gamma Z_j X_{ij} + \phi_{ij}, \quad (2)$$

where β and γ are parameters to be estimated. When individuals choose where to live, they implicitly compare neighborhoods in their choice set—that is, neighborhoods that they know about and where they may move with a nonzero probability. The difference in utility between the j th and the k th neighborhood is

$$U_{ij} - U_{ik} = \beta(Z_j - Z_k) + \gamma(Z_j - Z_k)X_i + (\phi_{ij} - \phi_{ik}). \quad (3)$$

Utility differences among neighborhoods for a given individual are thus a function of differences in observed and unobserved characteristics of neighborhoods and individuals. Because utility comparisons take place *within* individuals, their characteristics X_i do not affect the utility comparison additively. These characteristics, however, may interact with neighborhood characteristics. For example, the effect of differences in the proportion of persons in a neighborhood in a given ethnic group on the relative attractiveness of the neighborhoods may differ between individuals who are members of that ethnic group and those who are not. Unmeasured characteristics of individuals may also modify the effects of neighborhood characteristics, as we show below. These unmeasured characteristics can induce random variation in the effects of measured neighborhood characteristics β . For example, the effect of the proportion of persons in the neighborhood who are ethnic minorities may depend on an individual's level of tolerance, which is unobserved to the analyst.

Given data on the characteristics of individuals and neighborhoods and the behaviors or stated preferences of individuals for neighborhoods and an assumed probability distribution of the unobserved characteristics of individuals and neighborhoods, it is possible to estimate the parameters of the discrete choice model. If the ϕ_{ij} follow a type I extreme value (Gumbel) distribution, we obtain the conditional logit model²

$$p_{ij}(Z_j, X_i, C_{(i)}) = \frac{\exp(\beta Z_j + \gamma Z_j X_i)}{\sum_{k \in C_{(i)}} \exp(\beta Z_k + \gamma Z_k X_i)} \quad (4)$$

where $C_{(i)}$ denotes the choice set for the i th individual, which may be restricted to incorporate discrimination, prices, or information constraints (McFadden 1978).³ For example, the choice set may be restricted to units within a given radius of a person's current home, to units in neighborhoods that are at least 10% own-race, or to units where monthly rent or mortgage payments would be less than some fraction of individuals' incomes. Typically these models are estimated using maximum likelihood, where the likelihood is

$$L = \prod_{i=1}^N \prod_{j=1}^J \frac{y_{ij} \exp(\beta Z_j + \gamma Z_j X_i)}{\sum_{k \in C_{(i)}} \exp(\beta Z_k + \gamma Z_k X_i)}. \quad (5)$$

Early applications of the basic discrete choice model to residential mobility analysis include McFadden (1978) and Lerman (1975). Gabriel and Rosenthal (1989) use a multinomial logit model to examine how race and other traits of individuals affect residential mobility among five counties in the Washington, D.C., area. Sermons and Koppelman (2001) estimate a discrete choice model of residential choice that explores how men and women differ in their sensitivity to commuting time.⁴

3.4. Independence from Irrelevant Alternatives Assumption

The conditional logit form of the discrete choice model assumes independence from irrelevant alternatives (IIA). It is a model for pairwise comparison and assumes that the odds of preferring an alternative in a pairwise comparison is unaffected by the other available alternatives. That is, after accounting for observable features of choices, the remaining (unobserved) features of choices are uncorrelated (that is, $E[\varphi_{ij}, \varphi_{ik}] = 0$). IIA is really an assumption about proper model specification, which implies that there is no omitted variable bias and also that the choice set is exhaustive and well defined (McFadden, Train, and Tye 1981). The IIA property implies that the ratio of probabilities for any two choices is unaffected by the utilities of all other alternatives implying that the ratio is not affected by the addition or exclusion of alternatives. The conditional probability of choosing the j th neighborhood given a choice between neighborhood j or k is

$$\Pr(Y_{ij} = 1 | Y_{ij} \in \{j, k\}) = \frac{\Pr(Y_{ij} = 1)}{\Pr(Y_{ij} = 1) + \Pr(Y_{ik} = 1)} = \frac{\exp(\beta Z_j + \delta Z_j X_i)}{\exp(\beta Z_j + \delta Z_j X_i) + \exp(\beta Z_k + \delta Z_k X_i)} \quad (6)$$

This probability does not depend on the traits of neighborhoods other than j and k . If valid, this assumption makes it possible to estimate choice models on a subset of alternatives in the choice set. Additionally, we can make out-of-sample predictions

because the parameter estimates from the model are invariant to the inclusion or exclusion of alternatives in individuals' choice sets.

However, in practice the IIA assumption is often not met. We rarely observe all attributes of destinations that affect mobility behavior. Some neighborhoods have similar characteristics and, if one of them were omitted, individuals would disproportionately choose a similar neighborhood rather than distribute themselves proportionately across both similar and dissimilar neighborhoods. Unless the sources of similarity and dissimilarity among neighborhoods are controlled in the choice model, the model is likely to yield incorrect predictions about the effects of omitting one of the neighborhoods. The most common way of testing for IIA is through partitioning the choice set, and comparing estimates from a full model with those from a model estimated using a subset of the choice set (Hausman and McFadden 1984; Small and Hsiao 1985).⁵

There are three ways of dealing with IIA violations. First, we can ignore violation of the IIA assumption but recognize that the estimated parameters are at best an approximation of choice behavior and are not appropriate for making inferences about substitution patterns. Second, we can in principle modify the discrete choice model by adding additional covariates that represent sources of neighborhood resemblance. However, usually we cannot capture all the unobserved correlation in choice behavior explicitly. Finally, if available data permit, we can use a mixed logit specification, preferably with panel data that permit identification of unobserved time invariant neighborhood heterogeneity. We discuss these models in more detail below.

3.5. Unmeasured Heterogeneity

Even neighborhoods that are identical on measured characteristics may vary in their desirability to individuals. For example, neighborhoods may vary in amenities that have not been measured (nearness to the ocean or availability of charming coffee shops). Additionally, even among individuals who have identical measured attributes, we may observe variation in their mobility behavior. Unaccounted for features of individuals or neighborhoods that affect choice behavior can lead to correlations in the disturbance φ_{ij} across alternatives. Another form of unobserved heterogeneity arises if we incorrectly assume that people select one neighborhood directly from a given choice set when in fact they decide sequentially, systematically narrowing down their options based on some criterion. For example, choosers may first select part of a city, then select a neighborhood within that part, and then a house within the neighborhood. In this case, all neighborhoods within the chosen region and all vacant houses within the chosen neighborhood have a higher than average probability of selection irrespective of their measured characteristics. When the number of alternatives is small, we can represent the average level of attractiveness of each residential choice by including alternative-specific constants, which enter as dichotomous variables in the choice model. However, when the choice set is large, when we seek to parameterize the effects of measured attributes of neighborhoods on choice probabilities, or when the concern is with unobserved attributes of individuals that influence

choice behavior, it is more appropriate to estimate a model that allows for correlation in the attractiveness of observations within or among individuals. Several models are available to represent correlation of attractiveness across observations, including the nested and mixed logit models. We discuss these in turn.

3.5.1. Nested Logit Models. Nested logit models may solve the problem of unmeasured neighborhood heterogeneity if unmeasured characteristics of alternatives can be accounted for by conditioning on the appropriate choice subset. For example, if the choice set is all neighborhoods within the Detroit Metropolitan Area, but all the neighborhoods within the Grosse Pointe area of Detroit share key attributes (zoning regulation, funding for schools, etc.), at least some of which are unmeasured, we can treat Grosse Pointe neighborhoods as a subset. Subsets or “nests” are alternatives that are similar along one or more dimensions not accounted for in the formal discrete choice model. The nested logit model partitions the choice set C into N “nests,” C_n such that the complete choice set $C = \bigcup_{n=1}^N C_n$. Nests can represent a decision sequence (e.g., people first choose a region of the country, then a city, and then a neighborhood) or account for attributes of alternatives that make them more similar in both their observed and unobserved characteristics. The nests are constructed such that, for any two alternatives that are within the same nest, the ratio of probabilities is independent of the existence of all other alternatives.⁶

The nesting structure assumes that (1) neighborhoods that are in the same nests share unobserved features and (2) neighborhoods across nests do not share these unobserved features. That is, choices may have correlated unobservables within nests but not between them.⁷ Whereas in the simple conditional logit model, disturbances are independent and follow a univariate extreme value distribution, in the nested logit the *marginal* distribution of the disturbances across nests follows a univariate extreme value distribution but the disturbances may be correlated within nests (Train 2003, Ch. 4). To estimate the nested logit model, the nesting structure must be known to the analyst in advance, which is often not the case.⁸ Resemblance of alternatives on unobserved traits for any subset of alternatives, moreover, is often not an all-or-nothing matter but rather a matter of degree. These considerations give rise to the need for a more flexible model for unobserved heterogeneity.

3.5.2. Mixed Logit Model. Mixed logit models are a more general class of models that can accommodate both alternative- and individual-specific unmeasured heterogeneity, and they are useful if the analyst believes that the unobserved heterogeneity is correlated with observable characteristics of neighborhoods. The model is an extension of Equation (4). In particular, the error component φ_{ij} is broken out into two parts; that is,

$$U_{ij} = \beta Z_j + \gamma Z_j X_i + \mu_i W_{ij} + \varepsilon_{ij}, \quad (7)$$

where μ_i is an individual-specific (alternative invariant) random vector with zero mean, the W_{ij} are one or more vectors of data related to the j th alternative, and $\varphi_{ij} = \mu_i W_{ij} + \varepsilon_{ij}$. The W_{ij} represent characteristics of alternatives that may or may not include interactions

with individual-level variables, and the ε_{ij} follow a type I extreme value distribution. The specification of the W_{ij} generates correlation in alternatives over the unobserved portion of utility because the covariance between any two alternatives is

$$E([\mu_i W_{ij} + \varepsilon_{ij}][\mu_i W_{ik} + \varepsilon_{ik}]) = W'_{ij} V(\mu) W_{ik}, \quad (8)$$

where $V(\mu)$ is the covariance matrix for μ . Given some value μ_i , the conditional choice probability follows the logistic distribution since the remaining error component ε_{ik} follows an extreme value distribution:

$$p_{ij}(\mu_i) = \frac{\exp(\beta Z_j + \gamma Z_j X_i + \mu_i W_{ij})}{\sum_{k \in c(i)} \exp(\beta Z_k + \gamma Z_k X_i + \mu_i W_{ik})}. \quad (9)$$

Because the μ_i are unobserved, the unconditional probability is the logit formula integrated over all possible values for μ_i , weighted by the density of μ .

$$p_{ij} = \int p_{ij}(\mu) f(\mu | \Omega) d\mu \quad (10)$$

where Ω denotes support for the distribution of μ . These models are referred to as “mixed logit” because their probabilities are heterogeneous with f as the mixing distribution (Train 2003). The mixing distribution is assumed by the analyst, and it can be normal, lognormal, or other shapes. Because choice probabilities do not have closed form solutions, they cannot be estimated directly. Instead, the probabilities can be simulated by drawing values of μ , from its assumed distribution, using a Gibbs sampler, EM algorithm, or some other form of iterative estimation (see Train 2003, Chs. 8–10). These models can be estimated using specialized software for discrete choice estimation, such as the NLOGIT package for LIMDEP.

The choice probabilities depend on parameters β , γ , and Ω . Different patterns of correlation are specified based on the choice of W_{ij} . For example, in the nested logit model with N nests, W_{ij} is a set of dummy variables, d_j^c , indicating whether the j th alternative belongs in the c th nest ($\overline{W} = \{d_j^1, d_j^2, \dots, d_j^N\}$). In this case, the μ_i are IID random deviates where $V(\mu)$ is a diagonal matrix with elements σ_n , $n = 1, 2, \dots, N$. The unobserved component is correlated within but not between nests, with covariances $E([\mu_i W_{ij} + \varepsilon_{ij}][\mu_i W_{ik} + \varepsilon_{ik}]) = \sigma_n$ if alternatives j and k are both in the n th nest, and equals zero otherwise.

If the pattern of unobserved heterogeneity across alternatives is unknown, the W_{ij} can be specified as error components that, along with ε_{ij} , make up the random component of utility. In the usual conditional logit model, W_{ij} are zero, which means there is no correlation in utility over alternatives after conditioning on observables. When $W_{ij} \neq 0$, utility is correlated over alternatives, even when the error components are independent across observations such that $V(\mu)$ is a diagonal matrix.

Because this specification includes no measured neighborhood characteristics to identify the correlation across observations, it requires strong assumptions about the distribution of the W_{ij} random deviates.

Mixed logit models can also represent heterogeneity in individual behavior by assuming that $W_{ij} = Z$ (or $Z_j X_i$ when the random coefficient refers to interaction between alternative- and individual-specific variables) such that $U_{ij} = \beta Z_j + \gamma Z_j X_i + \mu_i Z_j + \varepsilon_{ij}$. Under this circumstance, $\beta_i = \beta + \mu_i$ and thus the coefficients of beta vary over individuals, with mean β and deviations μ_i . Elements of Z_j that do not enter into W_{ij} have fixed parameters that do not vary over the population. Similarly, elements of W_{ij} that do not enter into Z_j are variables whose parameters vary within the population but have means of 0. This is analogous to the standard random coefficient framework for linear models. For example, if W_{ij} includes a variable that is the difference between the j th neighborhood's median income and the i th individual's household income, the estimated model would allow for individual variation in response to neighborhood median income, potentially reflecting unobserved differences in consumption patterns.

While mixed logit models are widely used in transportation and land-use research, there are only a few studies that apply them specifically to the analysis of residential choice. In their analysis of Dallas County households' choices to live in a particular land-use zone, Bhat and Guo (2004) estimate a mixed spatially correlated logit that allows for both unobserved taste variation among movers and also spatial correlation among adjacent zones. More recently, Hoshino (2011) uses a mixed logit model to analyze stated preference data collected in Tokyo.

3.5.3. Estimating Unobserved Heterogeneity in Alternatives with Repeated Measures Data. When the goal is to estimate unobserved heterogeneity across individual movers, or when the correlation in unobservables across alternatives is well defined (for example, in the nested logit specification and other special cases), the mixed logit model is an elegant way of parameterizing unobserved heterogeneity in the choice model. If we believe that there is unobserved heterogeneity across alternatives but do not know the structure of this heterogeneity, the model is not generally identified.

If, however, we observe more than one choice by at least a subset of individuals, identification can be achieved. A typical form of repeated measures comes through panel observations, in which individuals make repeated decisions about whether and where to move. This requires that we observe the same individuals making mobility decisions over a period during which observable characteristics of neighborhoods change. This enables the analyst to control for unobserved time-invariant characteristics of alternatives (e.g., proximity to beach or neighborhood history). With repeated measures, either fixed effects or correlated random effects specifications are available. The fixed effects specification is tantamount to incorporating a dummy variable for every alternative. The random effects specification assumes a distribution for the unobservables but uses the assumed time invariance of the distribution to identify its correlation with time-varying characteristics of the alternatives. These models are applications of standard methods for discrete response models with panel or other

clustered data (Chamberlain 1980; Maddala 1983). Equivalently, discrete choice models with unmeasured heterogeneity and repeated measures can be regarded as a species of multilevel model, in which the levels include individuals, alternatives in the choice sets of individuals, and time-specific alternatives. Issues of identification and estimation of these models for residential choice parallel those for the general multilevel model. (See Skrondal and Rabe-Hesketh [2004] for a more detailed discussion of discrete choice models with unmeasured heterogeneity and their relationship to other multilevel models.)

3.6. Functional Form

Discrete choice models allow the analyst to specify a variety of ways that people may respond to characteristics of neighborhoods. For example, in models of the relationship between neighborhood racial composition and the probability of entering or leaving a neighborhood, it is not just the average level of tolerance that matters but also the shape of the response curve. Schelling (1971, 1978) showed that a high level of segregation results when individuals have a threshold response to the proportion own-group in their neighborhood—that is, when people are indifferent to neighborhood characteristics within some interval and only care about whether a neighborhood characteristic is above or below the threshold. In a simple model where only neighborhood characteristic Z_j enters into the choice equation, the utility in a threshold specification is

$$U_{ij} = \begin{cases} 1 & \text{if } Z_j > \text{threshold} \\ 0 & \text{otherwise} \end{cases}, \quad (11)$$

where the threshold is a specific value of Z_j . An alternative behavioral response is that people have a continuous response to neighborhood composition; in other words they are sensitive to even small changes in composition regardless of the actual level of the compositional variable. That is, utility is a continuous specification of neighborhood composition—for example, $U_{ij} = \beta Z_j$. In addition, a number of intermediate functional form specifications allow for indifference over some intervals of neighborhood composition with a threshold response at key points. These functional form assumptions about how people respond to neighborhoods have implications for neighborhood turnover and segregation dynamics. Bruch and Mare (2006, 2009) show how the shape of choice functions affects segregation dynamics.

3.7. Models for Ranked Data

The discrete choice models discussed thus far assume that the analyst observes only the chosen alternative and has no information on the relative utilities of unchosen alternatives. Stated preference data, however, may provide information on full or partial ranking of alternatives, albeit for a hypothetical choice set (Allison and Christakis 1994).⁹ Ties occur in the data when respondents assign multiple items the same rank, and incomplete rankings occur when respondents leave certain items un-ranked. In this case, we observe

groups of items that are ranked together, providing a partial ranking. The rank-ordered logit accommodates tied rankings (Allison and Christakis 1994:206–8). The likelihood function is an extension of the simpler discrete choice likelihood Equation (5), except that Y_{ij} is a rank rather than a 0/1 indicator for the chosen alternative, and the model includes an additional term δ_{ijk} , which equals 1 if the ranking of the k th choice is greater than or equal to the ranking of the j th choice, and is zero otherwise. That is,

$$L = \prod_{i=1}^N \prod_{j=1}^J \sum_{k \in C_{(j)}} \frac{y_{ij} \exp(\beta Z_j + \gamma Z_j X_i)}{\delta_{ijk} \exp(\beta Z_k + \gamma Z_k X_i)}. \quad (12)$$

In the case where one alternative is ranked “first” and all others are tied for “last,” the rank-ordered logit model simplifies to the discrete choice model for a single choice.

4. COMPLICATIONS FOR ACTUAL CHOICE DATA

In this section we discuss features of residential choice data that require modifications of standard discrete choice models. These include the aggregation of alternatives, violations of the independence from irrelevant alternatives assumption, unfeasibly large choice sets, choice-based sampling, and the treatment of a respondent’s current place of residence. We discuss how each of these problems can be handled within the choice model.

4.1. Aggregation of Alternatives

In actual residential choice, individuals select among housing units, apartments, or even rooms. Typically, however, we observe choices of aggregate units such as Census tracts. When the units that individuals actually choose are not the ones that we observe, it is necessary to modify the choice model to take into account the differential size and variability of the aggregate units (Ben-Akiva and Lerman 1985, Ch. 9). Denote by L the actual choice set (e.g., housing units). $P_i(l)$ is the probability that the i th decision-maker chooses the l th housing unit (where $l \in L$). The L housing units are partitioned into J nonoverlapping aggregates (e.g., Census tracts denoted as C_j) such that the total number of units in the j th aggregate, $M_j = \sum_{l \in C_j} \text{housing unit}_l$. The probability of choosing the j th tract is equal to the sum of the probabilities that the respondent chooses each of the tract’s constituent housing units. Thus, the probability that the chooser selects a housing unit located in the j th parcel is $P_i(j) = \sum_{l \in C_j} P_i(l)$, and the utility associated with the j th aggregate is the average utility of all its housing units:

$$\bar{U}_{ij} = \frac{1}{M_j} \sum_{l \in C_j} U_{il}. \quad (13)$$

An implication of this result is that, all else being equal, aggregate utilities and choice probabilities vary with the size of the aggregate units. Census tracts with more housing units will, *ceteris paribus*, be chosen more often than those with fewer units. Further, within tracts, individual dwelling units may be heterogeneous in their desirability. Thus, the estimated effects of other measured characteristics of tracts may be distorted by their correlations with tract size and variability. To take these complications into account we modify the general choice model in Equation (4) as follows:

$$P_j(i) = \frac{\exp(\bar{U}_{ij} + \mu_1 \ln M_j + \mu_2 \ln B_j)}{\sum_{k=1}^K \exp(\bar{U}_{ik} + \mu_1 \ln M_k + \mu_2 \ln B_k)}, \quad (14)$$

where \bar{U}_{ij} is the average utility of the housing units within the j th Census tract, M_j is the number of housing units in the j th Census tract, B_j measures the variation in the utilities of housing units within the j th Census tract, and μ_1, μ_2 are positive scaling coefficients (Ben-Akiva and Lerman 1993). Estimates of the M_j are typically available from Census data and thus can be straightforwardly included as regressors in the discrete choice model. However, we rarely have complete descriptions of the distribution of utilities of individual housing units and thus do not know the B_j .¹⁰

4.2. Large Number of Potential Destinations

When the residential choice set is all neighborhoods or housing units in a city or other large area, the number of observations can be very large in a discrete choice model, making it computationally burdensome to compute choice probabilities for every individual-alternative observation. For example, a discrete choice model for 1,000 individuals (and their location decisions) in a metropolitan area of 2,000 Census tracts has $1,000 \times 2,000 = 2,000,000$ individual-alternative combinations (if each tract is in the choice set of every sampled individual). Such a large data set makes computation difficult. However, we can obtain consistent estimates of the discrete choice model by sampling from the individual-destination observations within each respondent (McFadden 1978; Ben-Akiva and Lerman 1985). This procedure can be accomplished without significant loss of information, if we use all information on actually chosen alternatives and a random subsample of unchosen alternatives. This is analogous to the procedure of subsampling the risk sets in survival analysis (e.g., Breslow et al. 1983) or subsampling controls in case-control designs (Jewell 2004). If we subsample unchosen alternatives, it is possible to estimate a modified version of the model shown in Equation (4), which is

$$p_{ij}(Z_j, X_i, C_{(i)}) = \frac{\exp(\beta Z_j + \gamma Z_j X_i - \ln q_{ij})}{\sum_{k \in C_{(i)}} \exp(\beta Z_k + \gamma Z_k X_i - \ln q_{ik})}, \quad (15)$$

where q_{ij} denotes the known probability of sampling the j th destination for the i th respondent. We sample according to the following rules:

1. If the alternative is chosen, sample with $q_{ij} = 1.0$.
2. If the alternative is not chosen, sample with $q_{ij} < 1.0$.

For example, if we sample the unchosen alternatives with probability 0.05, this procedure yields a sample of $1,000 + (1,999 \times 1,000) \times 0.05 = 100,950$, a more manageable number of alternative-individual observations. This model can be estimated using standard maximum likelihood approaches for the discrete choice model, subject to the constraint that the coefficient on q_{ij} is 1.0. In practice, there are no firm guidelines for selecting a value of q_{ij} . The value will depend on both the sample size and also the size of the choice set. However, the computational burden of estimating the choice model is linear in both the number of observations and the number of alternatives. Thus if we have sufficient observations, it is more fruitful to analyze a sample of many observations with a small number of sampled alternatives rather than fewer observations with a large number of alternatives (Ben-Akiva and Lerman 1985:263). In practice, we can do sensitivity analyses to determine how alternative subsampling probabilities affect the estimated coefficients and standard errors. For example, we can vary the subsampling fraction and pick the smallest fraction that does not result in marked loss of precision of estimates.

4.3. Choice-based Sampling

Many surveys employ a form of stratified sampling that overrepresents some kinds of neighborhoods and underrepresents others. For example, surveys may oversample poor neighborhoods within a city or be drawn from schools or school districts with atypical minority or socioeconomic representation. Whereas this stratification scheme may be exogenous for some analytic purposes, it results in endogenous stratification for the study of neighborhood choice. Neighborhood stratified samples, therefore, are *choice-based* (Manski and Lerman 1977), in that the sampling procedure is confounded with the residential choices of the respondents. Without correction for sample design, estimates of parameters in discrete choice models are not, in general, consistent. If choice-based sampling probabilities are known, however, we can obtain consistent estimates of the model parameters using sampling weights. Manski and Lerman (1977) introduce an estimator in which each observed residential choice is weighted by its representation in the population as a whole. We define a function for each respondent,

$$WT_i = V_i / H_i, \quad (16)$$

where V_i denotes the population shares and H_i denotes the sample shares for that respondent's type. These weights enter the likelihood function for the model as

$$L = \prod_{i=1}^N \prod_{j=1}^J W T_i \times \frac{y_{ij} \exp(\beta Z_j + \gamma Z_j X_i)}{\sum_{k \in C(i)} \exp(\beta Z_k + \gamma Z_k X_i)}.$$

In practice, the correction weights for choice-based sampling can be estimated using the “importance weights” option in statistical estimation packages. For example, consider a sample of households where the proportion of respondents in high-poverty neighborhoods ($\geq 30\%$ of households below the poverty line) and low-poverty neighborhoods ($< 30\%$ of households below the poverty line) are each 0.5, whereas the population proportions of households in high- and low-poverty neighborhoods are 0.3 and 0.7, respectively. In this case, the Manski-Lerman weights are 0.3/0.5 for respondents in high poverty tracts and 0.7/0.5 for respondents in low poverty tracts.

4.4. Nuances of Behavior

4.4.1. Treatment of Own Neighborhood. In most populations the most common choice that an individual makes is his or her *own* residential location—that is, the choice not to move. This tendency to stay put may be due to the costs of moving as well as familiarity and comfort with the current location. Nonmoves are informative about residential choice because it is likely that the chances of opting for one’s own neighborhood do in fact depend on the measured characteristics of the neighborhood. Models of residential choice, however, should take into account the possibility that the weights that individuals place on neighborhood characteristics may be different for their own neighborhoods than for other potential destinations. We can represent the differential treatment people give to their own housing units or neighborhoods in the choice model by including a dichotomous variable, D_{ij} , that equals 1 if the housing unit or neighborhood under consideration is the respondent’s current residence and 0 otherwise. D_{ij} can enter into the model alone, which allows for a tendency not to move, or in interactions with characteristics of individuals or neighborhoods, which implies the differential own neighborhood by individuals with varying characteristics or differential evaluation of characteristics of own neighborhood. We illustrate how D_{ij} is used Section 7.

4.4.2. Neighborhood Change Versus Neighborhood Levels. Mobility history data can also show the extent to which people respond to neighborhood *change*, above and beyond their response to static compositional levels. Expectations regarding future changes in population composition and housing prices are important factors that may be based on recent changes in these conditions and may affect individuals’ mobility decisions. An expectation of continuing trends may create a self-fulfilling prophecy, where neighborhoods that are believed to improve or decline may in fact change in these directions because people act on these beliefs. These ideas are easily incorporated into the discrete choice model by including variables that represent

changes in neighborhood characteristics (that is, recent change in the Z_j), provided such data are available.

4.4.3 The Effect of Experience. Individuals' preferences may change as a result of their prior residential experiences, and this may affect their residential choices. When panel data on residential mobility or retrospective residential histories are available, the analyst observes multiple choices made by each decision maker and variation within as well as between individuals in exposure to different kinds of neighborhoods. If the unobserved component of utility is uncorrelated within people over time, we can treat each period as independent and analyze the longitudinal observations in the same way as cross-sectional data. In models estimated from these data, including covariates from other time periods can capture dynamic aspects of behavior. For example, a measure of the race-ethnic composition of individuals' previous neighborhoods, possibly interacted with the current neighborhood's race-ethnic composition, may reveal how past exposure to integrated or segregated neighborhoods can affect later decisions. However, the assumption that the unobserved component of utility is uncorrelated over time within people may not hold because some unobserved factors that affect choices persist over time. Moreover, if observable factors evolve over time, then unobserved factors may also be changing in a nonrandom way. For further discussion of how to separate enduring unobserved factors that affect choices from "habit formation" and other forms of inertia or persistence in discrete choice models, see Abbring (2010), Carro (2007), and Heckman and Navarro (2007).

5. COMPLICATIONS FOR STATED PREFERENCE DATA

In this section, we discuss potential issues for the analysis of stated preference data. With stated preference data, some of the complications created by mobility histories are avoidable, although other problems may arise. Typically the choice set observed in stated preference data is relatively small (e.g., five neighborhood vignettes in the MCSUI data), so choice-based sampling does not occur and the units of analysis are well defined. Although survey data on stated residential preferences typically do not offer respondents the option of choosing their own neighborhood, in principle, there is no obstacle to incorporating such measures in vignette designs. If the vignette data contain a choice that represents the respondent's current residence, we can explore whether the characteristics of one's own neighborhood have different effects from those of other potential destinations. Similarly, if the preference data are from a panel, it is possible to estimate models that allow for preferences to evolve over time.

However, discrete choice models based on stated preferences may, like those based on actual choices, be subject to unmeasured individual and location specific heterogeneity. Although randomized designs in stated preference studies eliminate correlation between unmeasured individual characteristics and exposure to neighborhood types, these designs cannot rule out *interactions* between unobserved individual characteristics and measured neighborhood characteristics. Moreover, whereas some characteristics of neighborhoods are observed by design, respondents may impute

additional dimensions of neighborhood composition based on the characteristics shown in the vignette. For example, if vignette neighborhoods vary in their ethnic composition, respondents may make assumptions about other aspects of neighborhood quality (such as safety and schools) that are correlated with ethnicity (Harris 1999). This leads to the same specification error that occurs when there is unobserved heterogeneity across neighborhoods in the actual move data. Whereas it is relatively straightforward to incorporate individual-level heterogeneity into stated preference models (e.g., by adding additional covariates or incorporating random coefficients using a mixed logit approach), allowing for unobserved heterogeneity in hypothetical alternatives is not possible. A potential solution is multidimensional vignettes (Emerson, et al. 2001), although respondents may find it difficult to respond to hypothetical multidimensional choices.

A problem specific to stated preference data is ambiguity in how respondents interpret vignette questions. The MCSUI asked three questions: (1) Would you move into/out of a neighborhood? (2) What is the relative attractiveness of each neighborhood? and (3) What is your ideal neighborhood? Responses to “Would you move into this neighborhood?” may yield different results from those to “Would you move out of this neighborhood?” Because people may evaluate their own neighborhood differently from other potential destinations, these two questions may not elicit the same stated preferences. Beyond this, the three questions may be measuring distinct aspects of preferences. The “would move in/out” provides a measure of the desirable neighborhoods above some acceptability threshold; the “ranked attractiveness” question provides a full ranking of neighborhood desirability; and the “ideal neighborhood” question measures the most desirable neighborhood in a multiethnic context. However, relative “attractiveness” of neighborhoods may not dictate the relative likelihoods that respondents would in fact choose those neighborhoods. The ideal neighborhood question allows the respondent to create a neighborhood rather than respond to prespecified proportions in a given ethnic group. If IIA holds, we can compare these preference data to those from different choice sets (including the two-race neighborhoods used in the other MCSUI vignettes). However, it is not clear whether a respondent’s “ideal” neighborhood is also his or her “most attractive” neighborhood. These problems do not reduce the value of the MCSUI and similar data for understanding racial preferences, but they imply that we must be careful in interpreting the results from each question type.

6. PRICES AND MARKETS

Residential choices are made in the context of housing markets and are thus constrained by limits to information, prices, incomes, and other institutional barriers. Actual move data are not a true measure of residential preference because they reflect the combined effects of preferences and constraints. If the analyst knows what informational constraints limit the choices of specific households, such constraints can be accommodated via restrictions to the choice set. Typically, however, we do not know what options an individual considers, although it may be possible to document the

different housing search strategies used by different race-ethnic groups or survey respondents' willingness to search for housing in specific areas (Krysan 2008; Krysan and Bader 2007). Affordability constraints can be incorporated into the choice model using measures of housing costs and the individual's economic resources. By itself, however, this approach assumes that housing prices are exogenous characteristics of dwelling units or neighborhoods. From the standpoint of modeling the marginal effect of neighborhood or housing characteristics, this assumption may be valid. Because prices are sensitive to housing demand, however, they are unlikely to be exogenous in the aggregate. The endogeneity of prices must be taken into account when we attempt to extrapolate individual behavior to aggregate population change.

6.1. Housing Markets and Housing Prices

Although housing prices affect choice behavior, the estimated effects of prices may be contaminated by factors omitted from the model that affect neighborhood desirability and thus also affect demand for housing in an area and housing prices. Estimating discrete choice models that include housing costs without taking into account this problem of unmeasured sources of desirability will result in inconsistent parameter estimates. In linear models, a possible solution is to use instrumental variables to eliminate correlation between the error term and covariates. However, discrete choice models are more complicated because of the nonlinearity of the model and possible interactions between the characteristics of individuals and the characteristics of their potential choices. To address these problems, Berry (1994) and Berry, Levinsohn, and Pakes (1995) estimate a series of alternative-specific constants that capture average demand for different alternatives (based on both observed and unobserved characteristics) and incorporate them into a conditional logit or mixed logit model. When applied to neighborhood choice data, the alternative-specific constants absorb the unobserved component of neighborhood desirability. This removes the simultaneity problem that arises out of correlation between prices and unobserved features of neighborhoods in models of individual choice.

This approach decomposes unobserved determinants of neighborhood choice into (1) the average utility that individuals derive from unobserved neighborhood characteristics (ξ_j) and (2) random individual deviations in the utility (ε_{ij}). The utility function can be written

$$U_{ij} = \beta Z_j + \gamma Z_i X_i - \alpha p_j + \varepsilon_{ij} + \xi_j, \quad (18)$$

where p_j denotes the average house price in the j th neighborhood. The negative coefficient indicates that neighborhood utility varies inversely with housing prices, all else being equal. The endogeneity problem is that prices depend on both observed and unobserved attributes of neighborhoods that affect desirability and thus demand. In other words, prices are a function of ξ_j . The solution is to introduce a constant for each neighborhood that captures its average utility (based on both observed and

unobserved characteristics). This moves ξ_j out of the error term and into this alternative specific constant. Rearranging terms in (18), we have

$$U_{ij} = [\beta Z_j - \alpha p_j + \xi_j] + \gamma Z_j X_i + \varepsilon_{ij},$$

where the term in brackets does not vary over individuals. If we denote the alternative specific constants as $\delta_j = \beta Z_j - \alpha p_j + \xi_j$ then

$$U_{ij} = \delta_j + \gamma Z_j X_i + \varepsilon_{ij}. \quad (20)$$

This choice model no longer has an endogeneity problem because ξ_j are subsumed into the alternative specific constants, which can be estimated along with the other parameters of the model. (We present this solution for the standard conditional logit model, but this strategy can also be applied to other models, including the mixed logit model). This model provides estimates of the alternative specific coefficient and the remaining parameters for choice behavior. However, the parameters associated with the utility for a given neighborhood that is common to all individuals remain subsumed in the δ_j . Fortunately, because these parameters enter the definition of the alternative specific constants linearly, they can be treated as outcomes in a regression model where the dependent variable is the alternative specific constant and the explanatory variables are characteristics of the neighborhood, including price. Here ξ_j is endogenous, but there are well-developed IV procedures for handling endogeneity in a linear model. The practical problem with this approach is that when the number of alternatives is large it is not feasible to estimate the alternative specific constants. Berry et al. (1995) provide an algorithm for estimating these parameters when there is a large number of alternatives.

Bayer and colleagues (Bayer and McMillan, 2008; Bayer, McMillan, and Rueben 2004) use this method in their analyses of residential choice and segregation dynamics. To obtain consistent estimates of the relationship between housing costs and mobility behavior, they divide their discrete choice utility function into a house-specific fixed effect, δ_j , and individual-specific interaction component, λ_{ij} , such that $U_{ij} = \delta_j + \lambda_{ij} + \varepsilon_{ij}$. They estimate model parameters using an iterative two-step procedure. In step 1, they estimate the parameters in λ_{ij} and the average utilities δ_j using a discrete choice model. In step 2, they instrument for prices to recover the parameters in δ_j . They then use a measure of the relative scarcity of a given housing unit or neighborhood in the housing market as the instrument. Neighborhoods that are unique or occur less frequently—for example, a perfectly racially mixed area that contains new housing stock—command higher prices assuming there is some demand.

7. EXAMPLES

7.1. Stated Residential Preferences in MCSUI Data

We illustrate the analysis of stated preference data using the MCSUI data for Los Angeles. For illustrative purposes, we analyze only the “ranked attractiveness” and

Table 2. Percentage of Neighborhoods Ranked First or Second, by Respondent's Ethnicity and Percentage Other in Neighborhood

	Other Group in Neighborhood			
	Black	White	Asian	Hispanic
Black Respondents				
Proportion Other-Group				
0	—	22.1	25.3	30.6
29	—	30.5	33.8	32.9
50	—	18.0	13.6	14.2
86	—	18.3	16.8	14.8
100	—	11.1	10.5	7.4
N=	—	356	374	380
Asian Respondents				
Proportion Other-Group				
0	47.0	14.8	—	42.0
29	43.6	31.6	—	38.6
50	7.0	15.8	—	11.3
86	1.9	23.6	—	6.7
100	0.6	14.3	—	1.5
N=	356	343	—	344
Hispanic Respondents				
Proportion Other-Group				
0	41.9	21.1	32.6	—
29	41.5	29.1	35.2	—
50	7.7	17.4	12.9	—
86	5.8	18.6	11.9	—
100	3.1	14.0	7.4	—
N=	307	341	338	

“would move in” data. The ranked-attractiveness data were collected only for non-white respondents. Table 2 shows the percentage of neighborhoods that were ranked first or second by black, Asian, and Hispanic respondents who were asked about neighbors of different race-ethnicities. Among black respondents asked about white, Asian, or Hispanic neighbors, the most attractive neighborhoods were those with a minority of other-group neighbors. However, a nontrivial proportion of black respondents identified the entirely other-group neighborhood (e.g., 100% white) as the most attractive neighborhood. Asian respondents were also most likely to rank neighborhoods with a minority of other-group neighbors as most attractive, although they found Hispanic and black neighbors less attractive than white neighbors. Similarly, Hispanic respondents found white neighbors more attractive than black or Asian neighbors but were most likely to rank neighbors with a strong Hispanic presence most attractive.

Table 3 shows the percentage of white, black, Hispanic, and Asian respondents willing to move into a neighborhood based on its neighborhood proportion other (where the other-group may be white, black, Asian, or Hispanic). The first column of

Table 3. Percentage Willing to Move into Neighborhood, by Respondent's Ethnicity and Percentage Other in Neighborhood

		Other Group in Neighborhood			
		Black	White	Asian	Hispanic
White Respondents					
Proportion Other-Group					
0	82.4	—		89.6	86.5
29	78.9	—		88.0	87.5
50	72.4	—		85.7	82.3
86	52.1	—		77.2	68.1
100	37.9	—		64.2	54.9
N=	261	—		307	288
Black Respondents					
Proportion Other-Group					
0	—		77.4	81.7	81.6
29	—		99.2	98.9	99.2
50	—		98.9	98.4	98.7
86	—		87.6	88.8	88.2
100	—		37.9	42.6	38.8
N=	—		354	376	374
Asian Respondents					
Proportion Other-Group					
0	97.5		85.3	—	95.3
29	96.3		100.0	—	99.2
50	88.1		100.0	—	98.7
86	74.3		97.70	—	88.2
100	23.7		72.1	—	38.8
N=	354		341	—	374
Hispanic Respondents					
Proportion Other-Group					
0	91.2		76.3	85.2	—
29	97.1		98.0	97.9	—
50	87.0		98.0	95.9	—
86	66.8		88.3	82.0	—
100	21.8		53.4	44.4	—
N=	307		341	338	

the table, which shows how white, Asian, and Hispanic respondents evaluate black neighbors, indicates that all groups avoid majority black neighborhoods. These descriptive tables show the distribution of responses over categories of neighborhood proportion other, but they do not provide a succinct way of showing the relationship between neighborhood preferences and neighborhood characteristics.

7.1.1. Models. We analyze the “ranked attractiveness” data by treating the five responses (one for each vignette neighborhood) as a full ranking of the alternatives.

Table 4. Coefficients for Neighborhood Proportion Other Group on Residential Preferences, Neighborhood Attractiveness and Would Move In Questions, Nonlinear Continuous Functions

Preferences of Blacks

	β	$SE(\beta)$	$ Z(\beta) $	β	$SE(\beta)$	$ Z(\beta) $
	Attractiveness Rankings			Would Move In		
neigh	1.202	0.190	6.33	2.412	0.400	6.02
neigh ²	-3.312	0.199	16.63	-2.962	0.452	6.55
I(Hispanic other) \times neigh	-0.739	0.189	3.91	-0.100	0.447	0.22
I(Hispanic other) \times neigh ²	0.290	0.298	0.97	0.142	0.485	0.29
I(Asian other) \times neigh	-0.528	0.313	1.69	-0.030	0.404	0.08
I(Asian other) \times neigh ²	0.492	0.251	1.96	-0.011	0.459	0.02
Log-likelihood		-11997			-13375	
N		15298			15556	

Preferences of Hispanics

	β	$SE(\beta)$	$ Z(\beta) $	β	$SE(\beta)$	$ Z(\beta) $
	Attractiveness Rankings			Would Move In		
neigh	0.891	0.329	2.71	2.550	0.460	5.54
neigh ²	-2.211	0.538	4.11	-2.752	0.358	7.68
I(black other) \times neigh	-1.420	1.219	1.16	-0.704	0.151	4.66
I(black other) \times neigh ²	-1.268	0.249	5.16	-0.169	0.295	0.57
I(Asian other) \times neigh	-0.738	0.572	1.29	-0.577	0.122	4.72
I(Asian other) \times neigh ²	-0.377	0.239	1.57	0.257	0.143	
Log-likelihood		-6230			-6949	
N		8197			8282	

Preferences of Asians

	β	$SE(\beta)$	$ Z(\beta) $	β	$SE(\beta)$	$ Z(\beta) $
	Attractiveness Rankings			Would Move In		
neigh	0.745	0.146	5.11	1.426	0.069	20.62
neigh ²	-2.123	0.171	12.40	-1.543	0.070	22.09
I(Hispanic other) \times neigh	-0.373	0.238	1.57	-0.103	0.056	1.85
I(Hispanic other) \times neigh ²	-3.810	0.255	14.95	-0.722	0.030	24.03
I(black other) \times neigh	-1.306	0.137	9.52	-0.077	0.072	1.08
I(black other) \times neigh ²	-1.306	0.838	7.30	-1.047	0.078	13.49
Log-likelihood		-2999			-4555	
N		5285			5355	

(continued)

In contrast, we treat the five responses to the “would you move in/out” question as a partial ranking of the alternative vignette neighborhoods, and use these rankings to estimate rank-ordered logit models with ties. In Table 1 each respondent has five

Table 4. (continued)

Preferences of Whites						
	β	$SE(\beta)$	$ Z(\beta) $	P	$SE(\beta)$	$ Z(\beta) $
	Attractiveness Rankings			Would Move In		
neigh				-0.250	0.226	1.10
neigh2				-1.800	0.159	12.05
l(Asian other) \times neigh				0.270	0.150	1.80
l(Asian other) \times neigh ²				-0.127	0.549	0.23
l(black other) \times neigh				-1.000	0.388	2.57
l(black other) \times neigh ²				-0.265	0.297	0.89
Log-likelihood					-10285	
N					14414	

lines of data, one for each neighborhood ethnic composition vignette and the respondent's rank of the vignette. The vignette rank is the dependent variable and is modeled as a function of the percent other-group in the neighborhood.¹¹ Separate parameters are estimated for each combination of a respondent's own race and the race of the other group in the vignette neighborhood. The nonlinear continuous model adequately describes residential preferences for these simple data. The coefficients from these models are shown in Table 4.

The predicted probabilities from the models for two of the ethnic groups, blacks and Hispanics, are presented in Figures 2 and 3. Figure 2(a) shows the probability that black respondents rank a vignette neighborhood most attractive. Separate panels are shown for black-white, black-Hispanic, and black-Asian neighborhoods. Black respondents tend to rank as most attractive those neighborhoods where their own ethnic group is heavily represented most. However, when asked which neighborhoods they would be willing to move into, blacks display a strong preference for integrated neighborhoods. Blacks are also slightly more partial to white neighbors than Hispanic or black neighbors; they respond to all three groups in a similar way for both the neighborhood attractiveness and "would move in" questions. Figure 3 shows the corresponding response profiles for Hispanics. Like blacks, Hispanics tend to find neighborhoods where their own group is heavily represented more attractive. However, unlike blacks, Hispanics tend to respond to mixed neighborhoods differently depending on the ethnicity of the other group. Hispanics find black neighbors least attractive. Hispanics are most likely to move into diverse neighborhoods.

7.1.2. Unobserved Heterogeneity. Within race-ethnic groups, individuals vary in their residential preferences and their expressed tolerance of other groups. To allow for unobserved heterogeneity within race-ethnic groups, we estimate a set of latent class models allowing for a distribution of responses to neighborhood composition within each ethnic group. This is a specific instance of the mixed logit model discussed above, where $W_{ij} = Z_j$ and the mixing distribution $f(\mu)$ is discrete, with μ taking a finite set of values $\mu_m \in \{\mu_1, \dots, \mu_M\}$ each with probability π_m . Here the

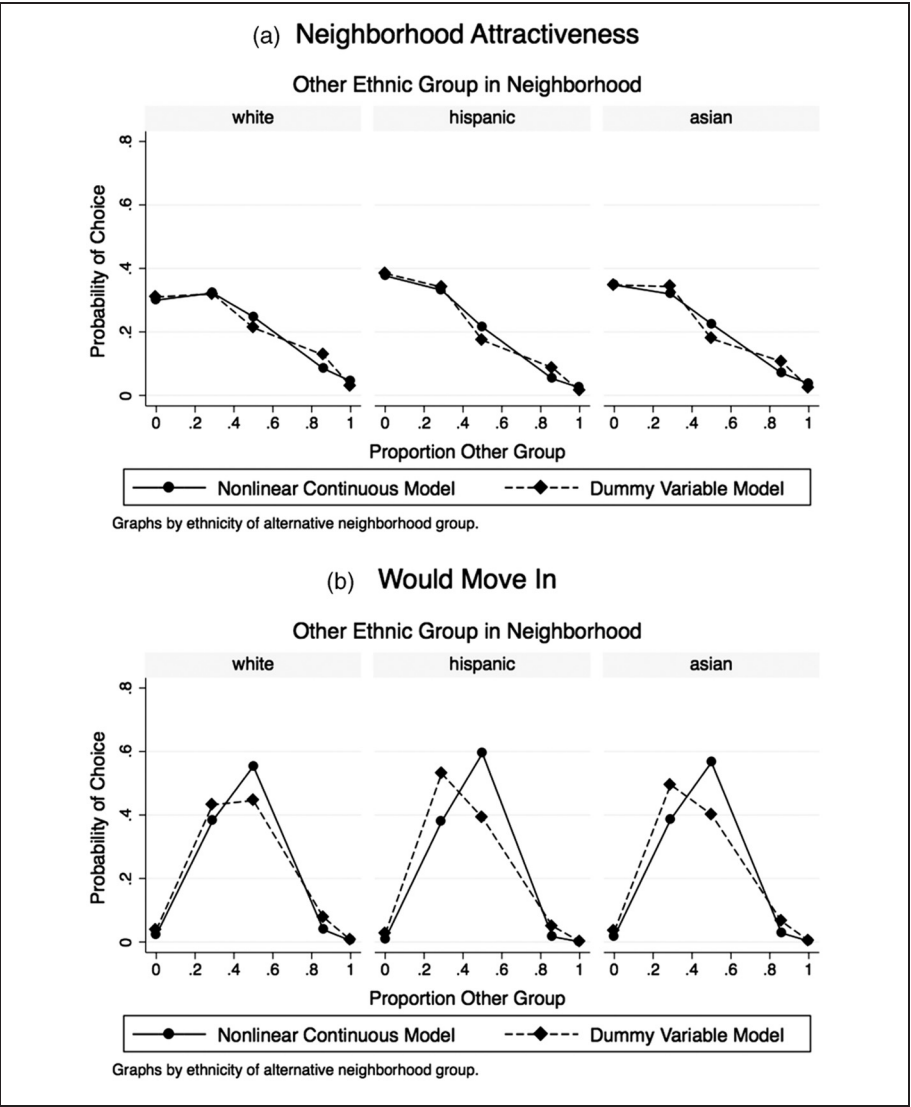


Figure 2. Predicted probabilities for black respondents, nonparametric (dummy variable) and nonlinear continuous models

mixed logit becomes a latent class model where subscript m denotes a particular class. The choice probability is then

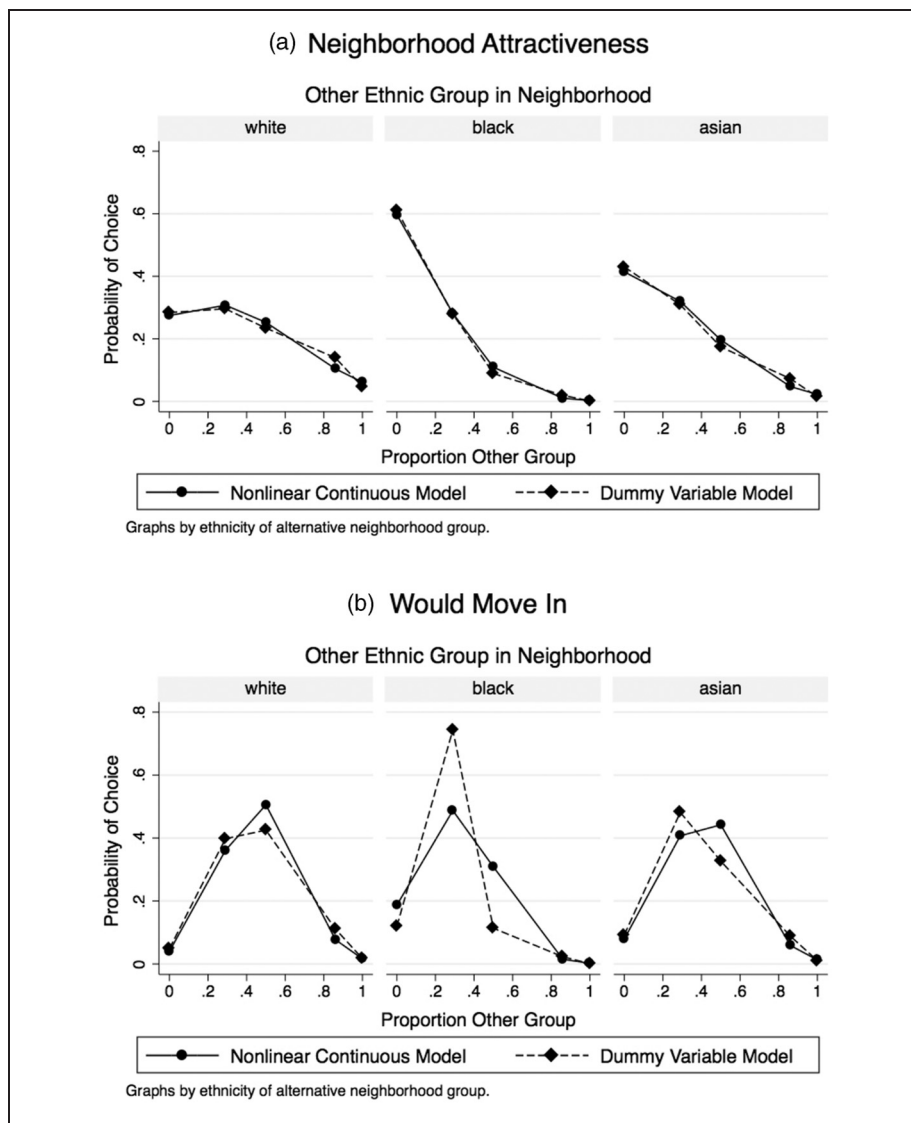


Figure 3. Predicted probabilities for Hispanics, nonparametric (dummy variable) and nonlinear continuous models

$$p_{ij} = \sum_{m=1}^M \pi_m \frac{\exp(\beta Z_j + \mu_m Z_j)}{\sum_{k \in C_{(i)}} \exp(\beta Z_k + \mu_m Z_k)}. \quad (21)$$

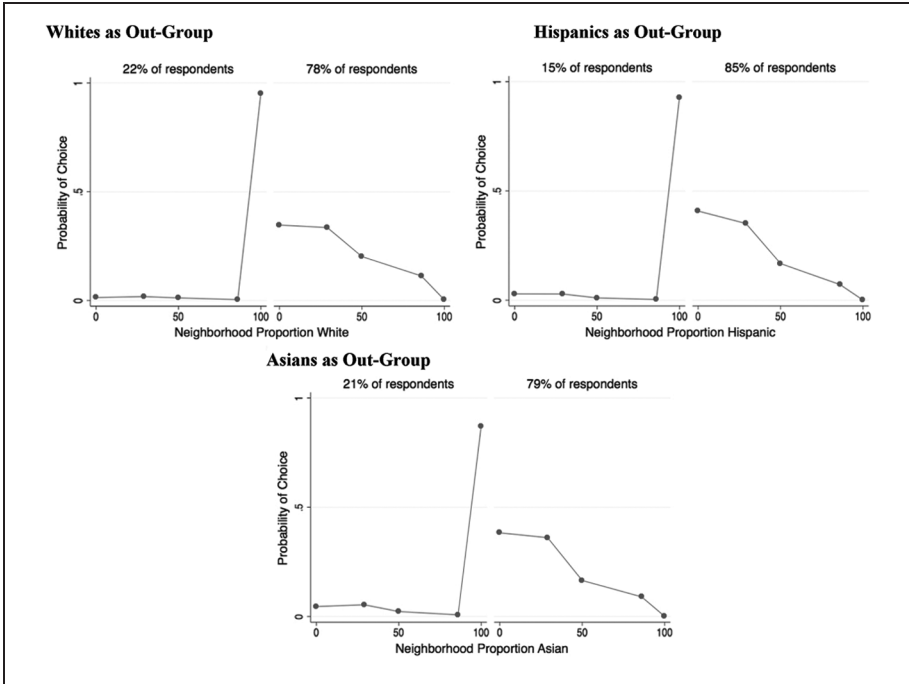


Figure 4. Predicted probabilities for blacks, unmeasured heterogeneity models, 2 groups

In our example below, we use the ranked-attractiveness data to estimate separate models by respondents' race and by the race of their vignette neighbors. We estimate a nonparametric model with dummy variables for each vignette neighborhood (omitted category is the 100% own-group neighborhood). Here Z_j is a set of dummy variables that identify vignette neighborhoods, so that $\bar{Z}_j = \{Z_0, Z_{29}, Z_{50}, Z_{86}\}$ and μ_m is the vector $\mu_m = \{\mu_m^0, \mu_m^{29}, \mu_m^{50}, \mu_m^{86}\}$. The utility for a member of the m th latent class is $U_{im} = \sum_j \beta_j Z_j + \mu_m^j Z_j$, where $j \in \{0, 29, 50, 86\}$ and the estimated effect of each individual Z_j for group $m = \beta_j Z_j + \mu_m^j Z_j$. Separate coefficients are estimated for each own-race/other-race combination.¹²

The results from estimating these models are shown for blacks and Hispanics in Figures 4 and 5, respectively. There is a clear pattern of response. For most respondents, the attractiveness of the neighborhood declines with the proportional representation of one's own race/ethnic group. However, among Hispanic and black respondents who were asked about white neighbors, roughly one quarter indicate that the most attractive neighborhood is the one that is 100% white. Similarly, among blacks and Hispanics who were asked about living among Asians, 19% of Hispanics and 21% of blacks in the sample identify the all-Asian neighborhood as most

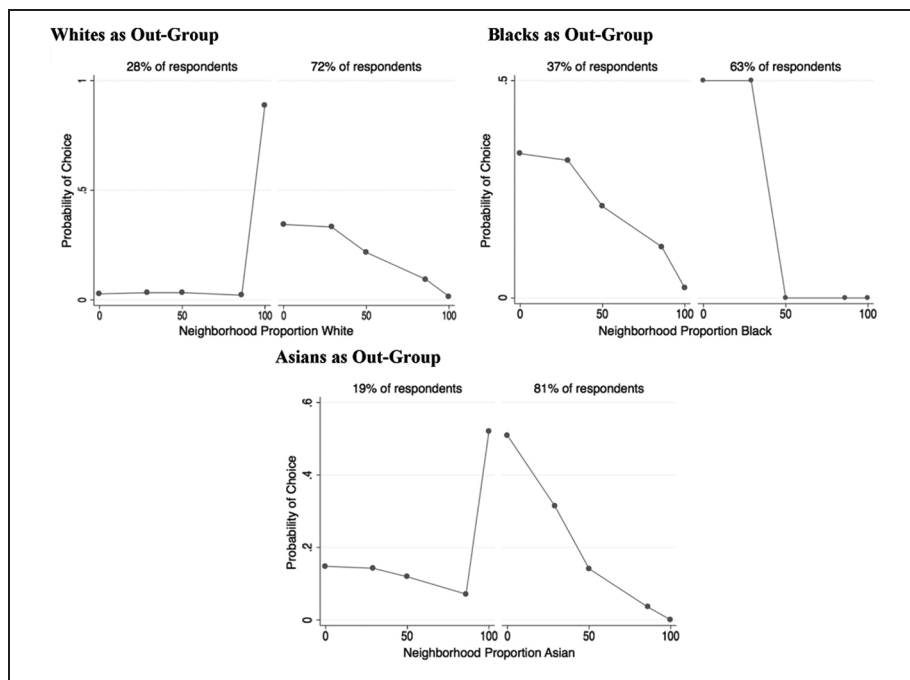


Figure 5. Predicted probabilities for Hispanics, unmeasured heterogeneity models

attractive. These results are consistent with those reported by other analysts of the same data (e.g., Charles 2000).

7.2. Actual Mobility Histories in the LA FANS Data

We illustrate how to analyze actual move data using the LA FANS Wave 1 data, which is a stratified sample of approximately 2,700 households in 65 Census tracts in Los Angeles County. The residential mobility history for each respondent was collected via an event history calendar for the 24 months preceding the survey date. Seventy percent of LA FANS respondents did not move during the two-year period prior to the interview, whereas 20% moved exactly once. Previous addresses in Los Angeles County are geolinked to the correspondent Census tract. However, we omit the small percentage (6.5%) of moves that occurred outside of Los Angeles County. We measure mobility in terms of annual moves, and observe up to two moves per respondent. Figure 6 shows one hypothetical mobility history for an LA FANS respondent. Because we examine annual mobility, multiple moves that occur within a single year are counted as a single move. Table 5 summarizes the information available for the analysis of residential mobility using the LA FANS data.¹³ The 2,332 respondents provide information on 4,508 annual residential mobility

Table 5. Summary of Observations in LA FANS and Race-Ethnic and Income Composition of L.A. County

Census Tracts (1990 Census): 1627					
Respondents in LA FANS Data: 2332					
Mobility Decisions	Total	White	Black	Hispanic	Asian
Year 1	2,178	600	227	1,186	162
Year 2	2,330	630	239	1,279	179
Total	4,508	1,230	466	2,465	341
Race-Ethnic Composition					
LA FANS	100.00	27.29	10.38	54.77	7.56
2000 Census	100.00	31.10	10.90	44.60	13.10
Moves Between Tracts					
Year 1	174	46	30	89	9
Year 2	238	43	37	147	11
Total	412	89	67	236	20
Person-Year-Options (Total)					
Year 1	3,538,825	976,300	369,329	1,929,622	263,574
Year 2	3,789,289	1,026,637	388,853	2,082,566	291,233
Total	7,328,114	2,002,937	758,182	4,012,188	554,807
Person-Year-Options (Sampling the Choice Set)					
Year 1	37,149	10,246	3,889	20,251	2,763
Year 2	39,797	10,753	4,100	21,890	3,054
Total	76,946	20,999	7,989	42,141	5,817

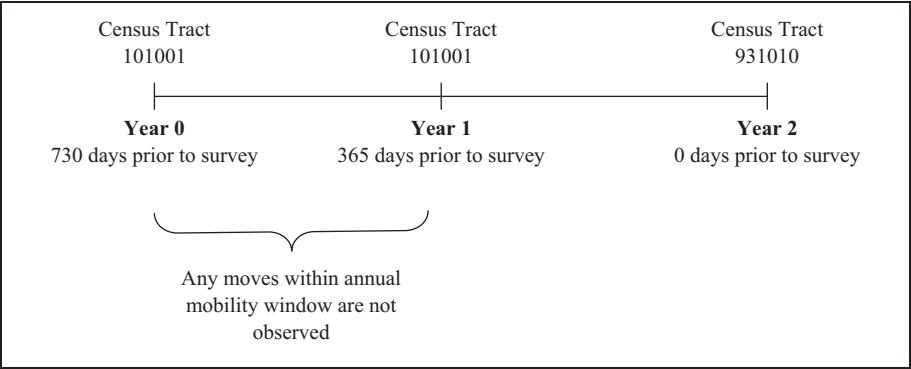


Figure 6. Example of one mobility history from the LA FANS

decisions.¹⁴ As indicated by the comparison with the 2000 Census data for Los Angeles County, our data overrepresent Hispanics and underrepresent non-Hispanic whites and Asians. Despite the relatively large number of mobility decisions faced by LA FANS respondents, they report only 412 annual between-tract moves during

Table 6. Adjustment for Choice-based Sample

Sampling Stratum	Sample Fraction (H_i)	Population Fraction (W_i)	Manski-Lerman Weight
Year 1			
40+ % poverty	0.281	0.090	0.320
20–39% poverty	0.319	0.301	0.944
0–19% poverty	0.401	0.600	1.495
Year 2			
40+ % poverty	0.296	0.090	0.304
20–39% poverty	0.307	0.301	0.980
0–19% poverty	0.397	0.600	1.510

the two-year mobility window, and 105 within-tract moves. For the purposes of this analysis, we consider moves to occur only if a respondent changes Census tracts during the annual mobility period.

7.2.1. Choice-based Sampling. The LA FANS is a stratified sample that overrepresents neighborhoods where at least 40 percent of households have incomes below the poverty line. For the purpose of estimating models of neighborhood choice, LA FANS is a choice-based sample. Our models include Manski-Lerman weights (see Equation 16) to correct for the differential representation neighborhoods in the data. A further complication is that the data come from retrospective mobility histories. Thus, whereas LA FANS is a choice-based sample at the time of the survey, prior to that respondents could live anywhere *conditional on living in one of the sampled tracts when the data were collected*. Thus the sample is purely choice-based at the time of the survey (Year 2, as shown in Figure 6), but influenced in a complex way by the choice-based sample in the periods prior to the survey date. Thus, we create two sets of Manski-Lerman weights: one using the distribution of choices at the time the LA FANS sample was drawn (in Year 2 of the mobility window), and one using the distribution of choices one year prior (Year 1 of the mobility window).

Table 6 illustrates the construction of Manski-Lerman weights in the LA FANS. The first column (H_i) shows the distribution of respondents across the sampling stratum in each of the two years, whereas the second column (W_i) shows the distribution of the population across sampling stratum. The LA FANS overrepresents high-poverty neighborhoods in both years. The chosen neighborhoods of respondents were 28% high-poverty in Year 1 and 30% high-poverty in Year 2 (when the data were collected). In contrast, only 9% of Los Angeles County neighborhoods were high-poverty during this period. The sample distribution more accurately represents the population one year prior to the survey date because individuals could, in principle, live in any Los Angeles neighborhood during this period rather than only in one of the 65 sampled neighborhoods. The Manski-Lerman weights, which are the ratio of the population fractions to the sampled fractions in each stratum, are shown in

column 3. The weights correct for over- and underrepresentativeness of sampled neighborhoods. The weights enter our discrete choice models using the “importance weights” option in Stata.

7.2.2. Large Number of Choices. Table 5 shows the distribution of mobility decisions over years and race-ethnicity of respondents. The 1,627 occupied Census tracts in Los Angeles (based on the 1990 Census) are potential destinations in each of 4,508 sample mobility decisions, resulting in an effective sample size of $1,627 \times 4,508 = 7,334,516$ person-year options, far too many observations for a tractable analysis. Thus, we sample from the alternatives within each respondent’s choice set with probability 1.0 for chosen alternatives and 0.05 for unchosen alternatives. This produces the smaller number of person-year-options shown in the bottom panel of Table 5. The models include the correction factor, $-\ln(q_{ij})$, for each alternative in each respondent’s choice set, where q_{ij} is the probability that the alternative is sampled, taking a value of $-\ln(1.0) = 0$ if the alternative was chosen by the respondent and $-\ln(0.05) = 3$ if the alternative was not chosen.

7.2.3. Definition of the Choice Set and Aggregation of Choices. When people choose where to live, they select a specific housing unit within a neighborhood. However, our observations consist of moves within and between Census tracts, rather than actual dwelling units. Thus, we add a term to our models, $\ln M_j$, where M_j is the number of housing units in the j th Census tract, to take into account between-tract variation in the number of potential destinations. In measuring within-tract mobility opportunities this way, we assume that the fraction of dwelling units that are in fact available to the respondent is invariant across tracts. With more detailed data on housing vacancies, it may be possible to relax this assumption. We do not know the variation in housing desirability within each tract, and thus estimate a discrete choice model similar to Equation (14), but omitting the term B_j .

7.2.4 Models of Residential Choice. We estimate conditional logit models that incorporate the effects of individuals’ personal characteristics and the characteristics of neighborhoods to which they might move, assuming that the choice set of each individual is all Census tracts in Los Angeles County. We allow for the possibility that respondents evaluate their current location differently from other potential destinations, by including a dummy variable D_{ij} , which equals 1 when destination j is the neighborhood currently occupied by respondent i , and 0 otherwise. The model, which can be written as

$$p_{ij}(Z_j, X_i) = \frac{\exp(\beta Z_j + \gamma Z_j X_i + \nu D_{ij} + \phi Z_j D_{ij} - \ln q_{ij} + \ln M_j)}{\sum_{k \in C(i)} \exp(\beta Z_k + \gamma Z_k X_i + \nu D_{ik} + \phi Z_k D_{ik} - \ln q_{ik} + \ln M_k)}, \quad (22)$$

incorporates terms for sampling the choice set, $-\ln q_{ij}$, for the number of households in a Census tract, M_j , for the “cost of moving” from one’s current location, and for the possibility that respondents evaluate their own neighborhood’s quality differently than they evaluate others. The model can be used to explore a number of

Table 7. Effects of Respondent and Tract Characteristics on Residential Choice, LA FANS data

Variable	Model 1.1 (Race Effects Only)		Model 1.2 (Mover-Stayer Only)		Model 1.3 (Race and Mover-Stayer Effects)	
	Beta	z(B)	Beta	z(B)	Beta	z(B)
D_{ij}			17.738	12.530	16.772	11.62
%black	0.469	0.62			-0.154	0.11
%black ²	-20.396	11.75			-13.215	5.19
black \times %black	13.607	9.76			3.301	1.29
black \times %black ²	-9.986	3.48			-0.250	0.06
%Hispanic	-14.224	18.97			-12.631	7.67
%Hispanic ²	5.434	9.34			6.071	4.83
Hispanic \times %Hispanic	5.950	8.90			4.416	3.07
Hispanic \times %Hispanic ²	-1.587	2.65			-1.994	1.35
%Asian	-4.338	6.00			-2.343	1.42
%Asian ²	-7.526	7.65			-9.022	3.54
Asian \times %Asian	17.529	11.62			3.986	1.28
Asian \times %Asian ²	-24.803	9.05			-3.267	0.52
%white ²	-10.323	19.16			-7.755	7.63
white \times %white ²	4.159	17.17			2.050	4.58
ln(#Number of households in tract)	1.172	30.62	1.021	9.080	0.952	8.29
$D_{ij} \times \ln(\# \text{ Number of}$ households in tract)			-0.920	5.720	-0.834	5.09
N	76746		76746		76746	
Log-likelihood	-30238		-4031		-3873	

Note: Models include Manski-Lerman weights, and the offset term (q_{ij}) for sampling the choice set

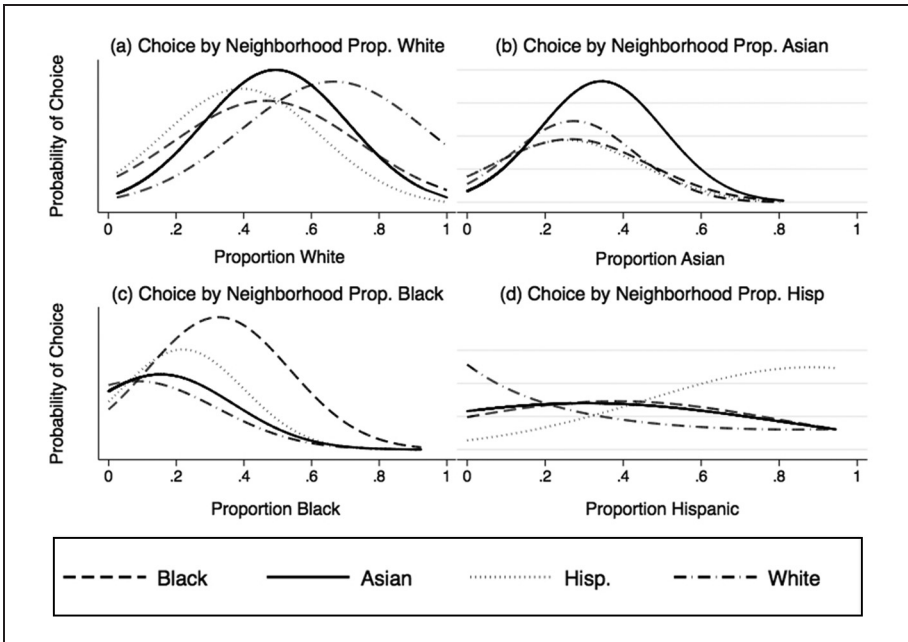


Figure 7. Predicted probabilities for whites, blacks, Asians, and Hispanics, LA FANS movers, by ethnic proportions

possible behavioral aspects of residential choice. For example, an interaction between neighborhood proportion black and neighborhood proportion Hispanic could represent the idea that Hispanics provide a “buffer” between blacks and whites. Table 7 presents coefficient estimates for a somewhat simpler specification in which each ethnic group responds uniquely to its own group and individuals evaluate their own neighborhoods differently from other potential destinations. The marginal probabilities from the full model (1.3) are shown in Figure 7.

8. LINKING RESIDENTIAL MOBILITY DECISIONS TO NEIGHBORHOOD CHANGE

Residential choice models predict the probabilities that individuals with varying characteristics select a neighborhood or housing unit, conditional on features of that alternative and of other potential destinations. Taken alone, these probabilities are ambiguous in their implications for aggregate neighborhood change because the choice probabilities estimated from these models describe the behavior of the marginal individual rather than the expected flows of population subgroups. At the aggregate level it is necessary to recognize that the mobility behaviors of all individuals are interdependent; that is, individuals respond to the composition of their local areas and all potential destinations, but their responses change that composition. When

behavior is interdependent, there is feedback from the aggregate to the individual level and no simple relationship between the choices of individuals and the residential patterns that result. To understand the implications of residential choice for neighborhood change, we need to connect individual-level probabilities with the distribution and size of the relevant population groups. We discuss three strategies for making this connection: (1) interactive Markov models, (2) general equilibrium models with price effects, and (3) agent-based models. Each method allows residential choice to change the attributes of neighborhoods, which affects subsequent mobility decisions. Markov and general equilibrium models are variants of macrosimulation approaches, representing mobility as expected rates of transition among neighborhoods or aggregate market adjustments respectively, whereas agent based models are microsimulations, in which individual mobility decisions are realizations of probabilistic choice.¹⁵ Markov and agent-based models are dynamic models that are useful for estimating the changes in population distribution across neighborhoods that result from underlying regimes of individual residential preferences. These changes can be pathways to an equilibrium residential distribution or between nonequilibrium states. General equilibrium models are useful for exploring the variation in equilibrium population distributions across neighborhoods associated with variation in exogenous conditions.

All three approaches assume a population of individuals distributed across a neighborhood environment and a set of rules governing mobility behavior. Individuals may be drawn from a hypothetical or a realistic population. An example of a hypothetical population involves two groups, each constituting 50% of the population, whereas a realistic population may be specified by Census data—for example, all households in Los Angeles County. The environment may be a highly stylized landscape (such as a 10 by 10 grid, where each cell on the grid represents a potential destination) or a realistic city (such as all Census tracts in Los Angeles County). The key features of the landscape are characteristics endogenous to the mobility process, such as neighborhood race-ethnic and economic composition. Fixed features, such as elevation, the location of highways and commercial areas, and air quality, may also be included. However, only neighborhood characteristics that can be represented as aggregates of individual characteristics and that affect individual decisions have a dynamic component. Neighborhood boundaries may be objectively defined, as in the case of Census tracts where all inhabitants of the same tract have the same neighborhood boundaries. Alternatively, in the case of agent-based models, neighborhoods can be defined such that each household has its own unique neighborhood. In all cases, individuals have rules for evaluating neighborhoods. In the cases we discuss below, this rule is operationalized through a discrete choice model. In all these models, the composition of neighborhoods is an endogenous outcome of the model. Each move between times t and $t + 1$ changes the opportunity structure for all individuals who contemplate a move between $t + 1$ and $t + 2$. Thus, all models incorporate not only the aggregate implications of individual preferences but also the feedback effects of aggregate change on the mobility behavior of individuals.

8.1. Interactive Markov Models

Markov models link a set of individual- or group-specific residential mobility probabilities to expected patterns of neighborhood turnover. A Markov model has a finite set of J states, $S = \{s_1, s_2, \dots, s_J\}$. The states can be specific neighborhoods (for example, Census tracts in a city) or neighborhood types (for example, poor versus nonpoor neighborhoods). The expected distribution of the population across the J states at time t is

$$m[t] = [m_1^1(t), \dots, m_J^1(t), m_1^2(t), \dots, m_J^2(t), \dots, m_1^G(t), \dots, m_J^G(t)], \quad (23)$$

where superscript $g = 1, 2, \dots, G$ indexes group membership (e.g., race-ethnic groups). We also specify a GJ by GJ matrix P of conditional probabilities that a member of group g moves to state j at time $t + 1$ conditional on being in state i at time t . Markov models assume that the distribution of the population at time $t + 1$ depends only on characteristics and locations of the population at time t (and no prior time periods). The population distribution at time $t + 1$ is then

$$m[t + 1] = Pm[t]. \quad (24)$$

This is equivalent to the operation of summing over transition probabilities within destinations

$$m[t + 1] = \sum_j \sum_g m[t]_{gj} \times P_t^{gj} \quad (25)$$

where $m[t]_{gj}$ denotes the size of population group g in state j . Markov models usually assume time-invariant probabilities (\mathbf{P}) of moving between states. However, if individuals both react to and transform features of their neighborhoods through their mobility behavior, then their behavior follows an *interactive* Markov model (IM) (Conlisk 1976), where the elements of P depend on the population distribution at time t :

$$m[t + 1] = P(m[t])m[t]. \quad (26)$$

Here $m[t]$ represents the distribution of blacks and whites across neighborhoods, and the probability of moving into a given neighborhood is a function of its ethnic composition. In this model, preferences for neighborhood characteristics are fixed, but the attractiveness of specific neighborhoods changes as a result of their changing characteristics.

To illustrate the interactive Markov model, we consider a simple city with two neighborhoods and a population of 10 blacks and 10 whites. At time 0 the population is completely segregated; all blacks are in one state, and all whites are in the other. Next, we compute the population trajectory for whites and blacks using their respective preference functions. For example, if people evaluate their neighborhoods according to a simplified version of Equation (4), where the probability that the i th

person selects the j th neighborhood is $e^{Z_j} / \sum_k e^{Z_k}$, where Z_j is neighborhood proportion own-group, then

$$\begin{aligned}
 m[1] &= P_0 \times m[0] \\
 &= \begin{bmatrix} \frac{e^{m[0][1]}}{e^{m[0][1]} + e^{m[0][2]}} & \frac{e^{m[0][1]}}{e^{m[0][1]} + e^{m[0][2]}} & 0 & 0 \\ \frac{e^{m[0][2]}}{e^{m[0][1]} + e^{m[0][2]}} & \frac{e^{m[0][2]}}{e^{m[0][1]} + e^{m[0][2]}} & 0 & 0 \\ 0 & 0 & \frac{e^{m[0][3]}}{e^{m[0][1]} + e^{m[0][4]}} & \frac{e^{m[0][3]}}{e^{m[0][1]} + e^{m[0][4]}} \\ 0 & 0 & \frac{e^{m[0][3]}}{e^{m[0][1]} + e^{m[0][4]}} & \frac{e^{m[0][3]}}{e^{m[0][1]} + e^{m[0][4]}} \end{bmatrix} \times \begin{bmatrix} 10 \\ 0 \\ 0 \\ 10 \end{bmatrix} \\
 &= \begin{bmatrix} \frac{e^1}{e^1 + e^0} & \frac{e^1}{e^1 + e^0} & 0 & 0 \\ \frac{e^0}{e^1 + e^0} & \frac{e^0}{e^1 + e^0} & 0 & 0 \\ 0 & 0 & \frac{e^0}{e^1 + e^0} & \frac{e^0}{e^1 + e^0} \\ 0 & 0 & \frac{e^1}{e^1 + e^0} & \frac{e^1}{e^1 + e^0} \end{bmatrix} \times \begin{bmatrix} 10 \\ 0 \\ 0 \\ 10 \end{bmatrix} \\
 &= \begin{bmatrix} 7.31 \\ 2.69 \\ 2.69 \\ 7.31 \end{bmatrix}
 \end{aligned}$$

At step 2,

$$\begin{aligned}
 m[2] &= p_1 \times m[1] \\
 &= \begin{bmatrix} \frac{e^{0.731}}{e^{0.731} + e^{0.269}} & \frac{e^{0.731}}{e^{0.731} + e^{0.269}} & 0 & 0 \\ \frac{e^{0.731}}{e^{0.731} + e^{0.269}} & \frac{e^{0.731}}{e^{0.731} + e^{0.269}} & 0 & 0 \\ 0 & 0 & \frac{e^{0.269}}{e^{0.731} + e^{0.269}} & \frac{e^{0.269}}{e^{0.731} + e^{0.269}} \\ 0 & 0 & \frac{e^{0.731}}{e^{0.731} + e^{0.269}} & \frac{e^{0.731}}{e^{0.731} + e^{0.269}} \end{bmatrix} \times \begin{bmatrix} 7.31 \\ 2.69 \\ 2.69 \\ 7.31 \end{bmatrix} \\
 &= \begin{bmatrix} 0.613 & 0.613 & 0 & 0 \\ 0.387 & 0.387 & 0 & 0 \\ 0 & 0 & 0.387 & 0.387 \\ 0 & 0 & 0.613 & 0.613 \end{bmatrix} \times \begin{bmatrix} 7.31 \\ 2.69 \\ 2.69 \\ 7.31 \end{bmatrix} = \begin{bmatrix} 6.14 \\ 3.87 \\ 3.87 \\ 6.14 \end{bmatrix} \quad (28)
 \end{aligned}$$

The process can continue until the neighborhoods reach equilibrium—that is, where $\Pi m[t+1] = P_t(m[t])$ and $m[\infty] = P_\infty = \prod P_t(m[t])$. Given an estimated discrete choice function that can generate the P_t , it is possible to compute the expected pattern of residential segregation under the mobility regime summarized in mobility matrices P_t using the standard measures of residential segregation (Mare and Bruch 2003). Tuljapurkar, Bruch, and Mare (2010) provide a mathematical analysis of Markov models for segregation and neighborhood change. In principle, an interactive Markov models for mobility between individual *neighborhoods* can be represented as a fixed rate Markov model of mobility between neighborhood *types* (e.g., Hermanns 2002).

8.2. General Equilibrium Models with Price Effects

Another strategy for studying neighborhood dynamics is using general equilibrium (GE) models with price effects. Bayer and colleagues (Bayer and McMillan 2008; Bayer, McMillan, and Rueben 2004) use GE models to examine the relationship between residential choice behavior and neighborhood outcomes. The analysis consists of two parts: (1) estimating a discrete choice model and (2) simulating the expected distribution of individuals in each neighborhood implied by the choice model. GE models assume that observed neighborhoods are in equilibrium, such that each individual had made an optimal choice given the choices of all other individuals. The models can be used to show how a new equilibrium distribution of neighborhoods results from some change in initial conditions or behavior (e.g., assuming that people are indifferent to the racial composition of their neighborhoods or assigning all ethnic groups equal income distributions). The first step is assuming or estimating a discrete choice model for the effects of housing prices, neighborhood race-ethnic composition, and other factors (as discussed in Section 6). Given this model, it is possible to simulate the impact of counterfactual conditions. For example, choice-model coefficients associated with neighborhood race-ethnic composition may be set to zero, to represent a city in which people make race-blind residential decisions and, using this modified choice model, it is possible to compute a new equilibrium distribution of neighborhoods.

In the first stage, predicted probabilities are computed representing the likelihood that an individual with a given demographic profile chooses a neighborhood of a given demographic composition. These probabilities are summed over neighborhoods to generate the demographic composition of neighborhoods in the next time period. Residential choice probabilities are recomputed to take into account changing neighborhoods, and the procedure repeats. More formally, the demographic composition of neighborhoods at time $t + 1$ is $Z_j^{t+1} = \sum_i P_{ij}$, where P_{ij} is the probability that the i th individual chooses the j th neighborhood. The process continues until a new equilibrium is reached, where $Z_j^{t+1} = Z_j^t \forall j$. As the composition of neighborhoods changes, their desirability, reflected in housing prices, changes as well. The establishment of a new equilibrium requires an update of housing prices so that the market clears. Market clearing prices are set such that, given the valuation of neighborhood characteristics by different types of individuals and a population, the expected number of people in each neighborhood matches the number of available dwellings. Housing prices are computed using an adaptation of an algorithm developed by Berry et al. (1995), that is,

$$\hat{p}_j^{t+1} = \hat{p}_j^t + \ln \left(\frac{s_j}{\hat{s}_j(\hat{p}_j^t)} \right), \quad (29)$$

where s_j and $\hat{s}_j(\hat{p}_j^t)$ are the actual and expected number of people in the j th neighborhood and p_j is a measure of housing prices in the j th neighborhood. To summarize, the new equilibrium population distribution over neighborhoods is computed in the

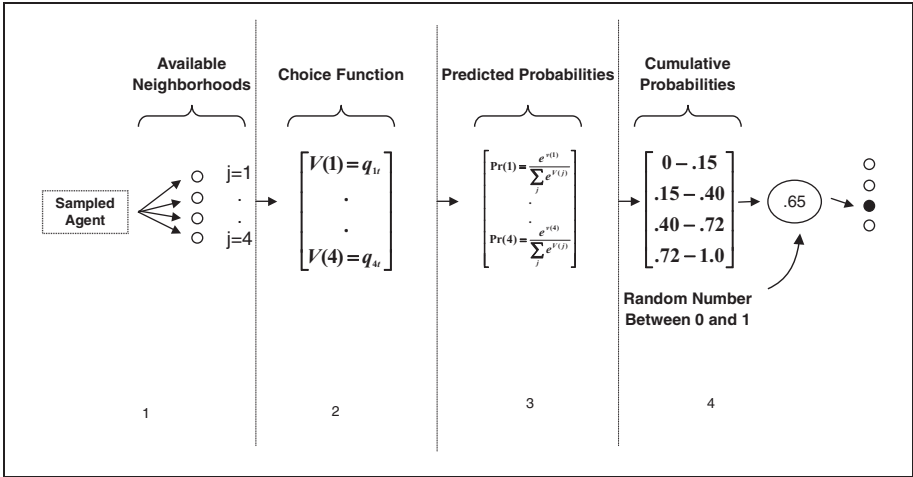


Figure 8. Determining chosen destination for agent using discrete choice framework.
Source: From Elizabeth Bruch and Robert Mare. 2006. "Neighborhood Choice and Neighborhood Change." *American Journal of Sociology* 112:667–709.

following steps: (1) compute residence probabilities associated with neighborhoods at time t ; (2) sum over individuals within neighborhoods to get new values for Z_j^t ; (3) compute new market clearing prices; (4) repeat 1 through 3 until convergence.

8.3. Agent-based Models

Agent-based models are a third approach to linking individual mobility to neighborhood dynamics (Macy and Willer 2002; Bonabeau 2002). Such models are microsimulations in which hypothetical individuals make choices based on either assumed behavioral rules or a statistical model of behavior. Agent-based models explicitly represent the feedback between individuals' behavior and aggregate processes (e.g., residential mobility and neighborhood change, mate preferences and marriage market dynamics, decisions to smoke or drink and high school norms around these behavior), and they can allow for detailed geography and individual heterogeneity. Schelling's (1971, 2006) model of residential tipping is an example of an agent-based model of a social process. Related models have been used to study norms regarding age at first marriage (Todd, Billari, and Simao 2005), income inequality and racial residential segregation (Bruch 2011), and other phenomena.

Agent-based models contain a population of actors who are assigned behaviors appropriate to the substantive application. An agent-based model of residential mobility assumes rules about how agents evaluate the desirability of neighborhoods and decide when and where to move. These rules can be simple heuristics or a more complex model such as that provided by the coefficients of a discrete choice model.

If agents' behavior is grounded in a discrete choice model, they use the values of neighborhood characteristics in their simulated world (as well as their own attributes) in combination with the model parameters to generate transition probabilities for moving among neighborhoods. The agent translates these transition probabilities into a (multinomial) distribution for the probabilities of selecting each neighborhood and "samples" a neighborhood via a draw from this distribution. In practice, the neighborhood-specific probabilities are cumulated, and the agent picks a neighborhood by drawing a number between 0 and 1 and choosing the neighborhood associated with the interval that contains that number. Figure 8 illustrates this process. See Bruch and Mare (2006, 2009) for a more detailed description of how to incorporate discrete choice models of residential mobility into agent-based models of neighborhood dynamics.

There are a number of software packages available for doing agent-based modeling, including Mason, Swarn, and Ascape. Netlogo (<http://ccl.northwestern.edu/netlogo/>) and Repast (<http://repast.sourceforge.net/>) are two of the most widely used options; both are free. Netlogo is a good option for researchers new to object-oriented programming. It is aimed at a less technical audience, has extensive documentation and user support, and contains a number of basic social science models in its model library. Repast requires some knowledge of Java, but can be used to run more complex models with larger populations. The software also allows for parallel processing. Both programs allow the user to import Geographic Information Systems (GIS) data to simulate mobility using realistic geography.

8.4. Comparing Approaches for Micro and Macro Linkage

Each of the three approaches takes a different tack to looking at segregation processes. Both Interactive Markov (IM) models and General Equilibrium (GE) models focus on the aggregate distributions of populations across neighborhoods, whether estimated in practice from aggregate or individual-level data. As they have been applied to residential choice studies, however, GE models are used for comparing equilibria under alternate assumptions, whereas IM models are used to examine the dynamics of residential mobility and neighborhood change (contrast Mare and Bruch 2003 using IM models with papers by Bayer and colleagues). Like interactive Markov models, agent-based models also focus on population dynamics, but they are built up from the actions of simulated individuals. However, there are two key differences between agent-based models and the GE and IM models: (1) Agent-based models have an explicit notion of "vacancies," where agents can move into an area only if there is an available slot; and (2) individuals in agent-based models make realized—not probabilistic—decisions. These differences may lead to substantively different segregation dynamics (for the same population and behavioral model).

9. FUTURE DIRECTIONS IN RESIDENTIAL MOBILITY RESEARCH

There are several promising directions for future research into residential mobility patterns and neighborhood change. One is the development of two-sided matching models for the coordination of housing seekers with housing providers. In the discrete choice models discussed in this paper, we treat the choice to move into a unit separate from the choice of the previous occupant to move out of a unit, without explicitly representing housing vacancies, whether transitory or longer term. We can imagine viewing the flow of people in and out of housing units as a coordinated set of simultaneous moves. This may be more analogous to real-world mobility behavior in that most housing units are not vacant for long periods of time, as residents move out and then in again in a matter of hours or days. The use of data on moves, housing stock (including vacancies), housing searches, prices, and real estate transactions would greatly enhance knowledge on residential mobility and neighborhood change.

A second research direction involves finding better tools for aggregate model evaluation. Agent-based models, interactive Markov models, and general equilibrium models are all tools for linking residential mobility behavior described by discrete choice models with aggregate patterns of neighborhood change. However, further work is needed to develop tools for evaluating the inferences from these models from sample or fragmentary population data (Berk 2008).

Despite their limitations, the discrete-choice modeling approach can reveal a great deal about how people choose where to live and the implications of these choice patterns for neighborhood change. The techniques discussed in this paper also potentially have relevance for a wider range of applications, including choice of marriage partner and aggregate patterns of assortative mating, the choice of occupation or academic field and the aggregate composition of the people holding that occupation or degree (e.g., England et al. 2007), and the choice of friends in high school and the composition of peer networks.

Appendix. MCSUI Residential Preference Questions

1. Full Ranking on “Attractiveness”

Now I would like you to imagine that you have been looking for a house and have found a nice house you can afford. This house could be located in several different types of neighborhoods as shown on these cards. Some of the neighborhoods have more white families and others have more black families. Would you look through the cards and rearrange them so that the neighborhood that is most attractive to you is on top, the next most attractive second, and so on down the line with the least attractive neighborhood on the bottom.

2. Partial Ranking on “Move Into”

Are there any of the neighborhoods you would not want to move into?
Would you show me all the ones you would not move into?

3. Single Choice “Ideal Neighborhood”

Now I’d like you to imagine an ideal neighborhood that had the ethnic and racial mix you personally would feel most comfortable in. Here is a blank neighborhood card like those we have been using. Using the letters A for Asian, B for Black, H for Hispanic, and W for White, please put a letter in each of these houses to represent where you would most like to live. Please be sure to fill in all the houses.

Acknowledgments

The paper benefited greatly from the comments provided by three anonymous reviewers.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The authors acknowledge funding from the Robert Wood Johnson Foundation, the National Science Foundation, and the John D. and Catherine T. MacArthur Foundation.

Notes

1. We focus on one aspect of human migration—namely residential mobility in urban settings and its implications for neighborhood change. However, the methods described in this paper can in principle also be applied to the analysis of international or other long-distance migration.
2. The statistical models for discrete choice that are discussed in this paper are all variants of conditional (multinomial) logit models, including generalized versions such as the mixed multinomial logit model. The mixed logit model, as discussed below, allows for very flexible treatment of various types of unmeasured heterogeneity. McFadden and Train (2000) demonstrate that the choice probabilities from any discrete choice model, including the multinomial probit model, can be estimated, with suitable choice of the mixing distribution for unmeasured heterogeneity, from a mixed multinomial logit model.
3. Most standard statistical software packages can be used to estimate the basic discrete choice models discussed in this section (that is, those that do not include unmeasured heterogeneity), either as a conditional logit, fixed effects logit, or multinomial logit model.
4. A number of papers from land use and transportation research use related frameworks to look at bundles of choices (e.g., choice of residential location, transportation, and workplace). See Waddell (1996) and Pinjari et al. (2007) for examples of this line of work.

5. However, Cheng and Long (2007) show that these tests often fail to reject IIA, and advocate either not using these models unless confident that the choice set alternatives are distinct and complete, or using a model that incorporates unobserved heterogeneity such as the nested or mixed logit described below.
6. Nested logit models can be estimated in most standard statistical software packages, including Stata, SAS, R and the Limdep package NLOGIT.
7. The standard nested logit assumes a simple hierarchical classification of alternatives within nests. Wen and Koppelman (2001) define a generalized nested logit that allows for more flexible substitution patterns.
8. Within the transportation and land-use literature, there have been a few applications of the nested logit model to residential choice. Lee and Waddell (2010) and Kim, Pagliara, and Preston (2005) estimate nested logit models that treat survey respondents' current housing unit as one nest, and place all other dimensions in a second nest. This allows for unobservable features of one's own neighborhood to be treated separately from other neighborhoods, but as we show later the treatment of one's current location can be handled parametrically within the standard discrete choice model. Quigley (1985) defines a three-stage structure where housing units are nested within neighborhoods, and neighborhoods are nested within towns.
9. Ranked preference data arise from situations in which individuals are asked to pick their most preferred option among all available (for example, among a set of neighborhood vignettes), their (next) most preferred option among those that remain, and so on until all options have been ranked. This is a generalization of the standard discrete choice problem. Allison and Christakis (1994) provide further details about this choice model.
10. When available housing vacancies are exactly proportional to Census tract size (that is, where each tract has the same vacancy rate and every vacant unit is available to every individual, M_j enters the choice equation with a coefficient $\mu_1 = 1$ and equation 14 can be estimated treating M_j as an offset). This is formally analogous to the offset term used by Zeng and Xie (2008) to represent opportunity constraints in friendship choice. However, the empirical separability of constraints from preferences in the Zheng-Xie models is only possible if the opportunity choice set is fully known. In the case of residential mobility, there are many restrictions on opportunities (e.g., affordability constraints, racial steering on the part of real-estate agents) that are not observed by the analyst. Insofar as one has information about potential opportunity constraints in the choice process, it may be more appropriate to simply include these attributes of choices as parameters in the model.
11. We estimate the models using the `rlogit` command in Stata.
12. We estimate these models using the GLLAMM module within Stata. Because of small sample sizes, we were unable to obtain estimates for Asians who were asked about white neighbors.
13. The LA FANS restricted data provided geocoded addresses mapped to 1990 tract boundaries. Thus the 2000 data shown in this table are adjusted to 1990 tract boundaries.
14. Respondents who failed to provide valid information about their location 12 months prior to the interview data are omitted from the sample. Respondents who provided valid information about their location 12 months prior to the interview date but failed to provide valid information about their location 24 months prior to their interview contribute one observation to the data: a mobility decision from the second year.
15. Although a full discussion of the practical use of dynamic models to connect individuals' choices to population processes is beyond the scope of this paper, we provide an overview of what methods are available (and how one might incorporate empirically grounded

choice behavior using the discrete choice models detailed above). More detailed technical information about the implementation of these models and the inferences that can be drawn from them may be found in the works cited in this section.

References

- Abbring, J. 2010. "Identification of Dynamic Discrete Choice Models." *Annual Review of Economics* 2:367–94.
- Allison, P. D. and N. A. Christakis. P. P. 1994. "Logit Models for Sets of Ranked Items." Pp. 199–228 in *Sociological Methodology* vol. 24, edited by Peter V. Mardson, Cambridge, MA: Blackwell Publishing.
- Basolo, V. and M. T. Nguyen. 2005. "Does Mobility Matter? The Neighborhood Conditions of Housing Voucher Holders by Race and Ethnicity." *Housing Policy Debate* 16:297–324.
- Bayer, P. and R. McMillan. 2008. "Distinguishing Racial Preferences in the Housing Market: Theory and Evidence." Pp. 225–44 in *Hedonic Methods in Housing Markets*, edited by Andrea Baranzini, Jose Ramirez, Caroline Schaerer and Philippe Thalmann. New York: Springer.
- Bayer, P., R. McMillan, and K. S. Rueben. 2004. "What Drives Racial Segregation? New Evidence Using Census Microdata." *Journal of Urban Economics* 3:514–35.
- Bayer, P. and C. Timmons. 2007. "Estimating Equilibrium Models for Sorting Across Destinations." *Economic Journal* 117:353–74.
- Ben-Akiva, M. and S. R. Lerman. 1985. *Discrete Choice Analysis: Theory and Application to Travel Demand*. Cambridge, MA: MIT Press.
- Ben-Akiva, M. and S. R. Lerman. 1993. *Discrete Choice Analysis: Theory and Application to Travel Demand*. Cambridge, MA: MIT Press.
- Berk, R. 2008. "How Can You Tell if the Simulations in Computational Criminology Are Any Good?" *Journal of Experimental Criminology* 4:289–308.
- Berry, S. T. 1994. "Estimating Discrete-Choice Models of Product Differentiation." *RAND Journal of Economics* 25:242–62.
- Berry, S. T., J. Levinsohn and A. Pakes. 1995. "Automobile Prices in Market Equilibrium." *Econometrica* 63:841–90.
- Bhat, C. and J. Guo. 2004. "A Mixed Spatially Correlated Logit Model: Formulation and Application to Residential Choice Modeling." *Transportation Research: Part B: Methodological* 38:147–68.
- Bobo, L., J. Johnson, M. Oliver, R. Farley, B. Bluestone, I. Browne, S. Danziger, G. Green, H. Holzer, M. Krysan, M. Massagli, and C. Z. Charles. 2000. *Multi-city Study of Urban Inequality, 1992-1994: [Atlanta, Boston, Detroit and Los Angeles] [Household Survey Data]* [Computer file]. 3rd ICPSR version. Interuniversity Consortium for Political and Social Research [distributor].
- Bonabeau, E. 2002. "Agent-based Modeling: Methods and Techniques for Simulating Human Systems." *Proceedings of the National Academy of Sciences* 99 Suppl. 3:7280–87.
- Breslow, N, J. H. Lubin, P. Marek, and B. Langholtz. 1983. "Multiplicative Models and Cohort Analysis." *Journal of the American Statistical Association* 78:1–12.
- Brock, W. and S. Durlauf. 2001. "Interactions Based Models." Pp. 3297–380 in *The Handbook of Econometrics*, edited by J. Heckman and E. Learner. Oxford, UK: Elsevier.
- Bruch, E. E. 2011. "Residential Sorting by Race and Income." Unpublished manuscript.

- Bruch, E. E. and R. D. Mare. 2006. "Neighborhood Choice and Neighborhood Change." *American Journal of Sociology* 112:667–709.
- Bruch, E. E. and R. D. Mare. 2009. "Preferences and Pathways to Segregation: Reply to Van de Rijt, Siegel and Macy." *American Journal of Sociology* 114:1181–98.
- Carro, J. M. 2007. "Estimating Dynamic Panel Data Discrete Choice Models with Fixed Effects." *Journal of Econometrics* 140:503–28.
- Chamberlain, G. 1980. "Analysis of Covariance with Qualitative Data." *Review of Economic Studies* 47: 225–38.
- Charles, C. Z. 2000. "Neighborhood Racial Composition Preferences: Evidence from a Multiethnic Metropolis." *Social Problems* 47:379–407.
- Charles, C. Z. 2005. "Can We Live Together? Racial Preferences and Neighborhood Outcomes." Pp. 45–80 in *The Geography of Opportunity: Race and Housing Choice in Metropolitan America*, edited by X. de Souza Briggs. Washington, DC: Brook- ings.
- Cheng, S. and J. S. Long. 2007. "Testing for IIA in the Multinomial Logit Model." *Sociological Methods and Research* 35:583–600.
- Clark, W. A. V. 2005. "Intervening in the Residential Mobility Process: Neighborhood Outcomes for Low-income Populations." *Proceedings of the National Academy of Sciences (PNAS)* 43 15307–312.
- Conlisk, J. 1976. "Interactive Markov Chains." *Journal of Mathematical Sociology* 4:157–85.
- Crowder, K. D. and S. J. South. 2008. "Spatial Dynamics of White Flight: The Effects of Local and Extralocal Racial Conditions on Neighborhood Out-migration." *American Sociological Review* 73:792–812.
- Crowder, K., S. J. South, and E. Chavez. 2006. "Wealth, Race and Inter-neighborhood Migration." *American Sociological Review* 71:72–94.
- Emerson, M. O., K. J. Chai and G. Yancey. 2001. "Does Race Matter in Residential Segregation? Exploring the Preferences of White Americans." *American Sociological Review* 66:922–35.
- England, P., P. Allison, S. Li, N. Mark, J. Thompson, M. Budig and H. Sun. 2007. "Why Are Some Academic Fields Tipping toward Female? The Sex Composition of U. S. Fields of Doctoral Degree Receipt, 1971-2002." *Sociology of Education* 80:23–42.
- Farley, R. and W. H. Frey. 1994. "Changes in the Segregation of Whites from Blacks during the 1980s: Small Steps toward a More Integrated Society." *American Sociological Review*. 59:23–45.
- Farley, R., C. Steeh, T. Jackson, M. Krysan and K. Reeves. 1993. "Continued Racial Residential Segregation in Detroit: Chocolate City, Vanilla Suburbs Revisited." *Journal of Housing Research* 4:1–38.
- Farley, R., C. Steeh, M. Krysan and T. Jackson. 1994. "Stereotypes and Segregation: Neighborhoods in the Detroit Area." *American Journal of Sociology* 100:750–80.
- Fussell, E., N. Sastry, and M. VanLandingham. 2010. "Race, Socioeconomic Status and Return Migration to New Orleans after Hurricane Katrina." *Population and Environment* 31:20–42.
- Gabriel, S. and S. Rosenthal. 1989. "Household Location and Race: Estimates of a Multinomial Logit Model." *Review of Economics and Statistics* 71:240–49.
- Groen, J. A. and A. E. Polivka. 2010. "Going Home after Hurricane Katrina: Determinants of Return Migration and Changes in Affected Areas." *Demography* 47:821–44.

- Harris, D. R. 1999. "Property Values Drop When Blacks Move in, Because...: Racial and Socioeconomic Determinants of Neighborhood Desirability." *American Sociological Review* 64:461–79.
- Hausman, K. and D. McFadden. 1984. "Specification Tests for the Multinomial Logit Model." *Econometrica* 52:1219–40.
- Heckman, J. and S. Navarro. 2007. "Dynamic Discrete Choice and Dynamic Treatment Effects." *Journal of Econometrics* 136:341–96.
- Hermanns, H. 2002. *Interactive Markov Chains: The Quest for Quantified Quality*. New York: Springer.
- Hoshino, T. 2011. "Estimation and Analysis of Preference Heterogeneity in Residential Choice Behaviour." *Urban Studies* 48:363–82.
- Jackson, M. I. and R. D. Mare. 2007. "Cross-Sectional and Longitudinal Measurements of Neighborhood Experience and Their Effects on Children." *Social Science Research* 36:590–610.
- Jewell, N. P. 2004. *Statistics for Epidemiology*. Boca Raton, FL: Chapman and Hall/CRC.
- Kim, J. F. Pagliara and J. Preston. 2005. "The Intention to Move and Residential Choice Behavior." *Urban Studies* 42:1621–35.
- Kingsley, G. T. and J. Johnson. 2003. "Patterns of Section 8 Relocation in the H. P. VI Program." *Journal of Urban Affairs* 25:427–7.
- Krysan, M. 2008. "Does Race Matter in the Search for Housing? An Exploratory Study of Search Strategies, Experiences and Locations." *Social Science Research* 37:581–603.
- Krysan, M. and M. Bader. 2007. "Perceiving the Metropolis: Seeing the City through a Prism of Race." *Social Forces* 86:699–733.
- Krysan, M., M. Couper, R. Farley and T. Forman. 2009. "Does Race Matter in Neighborhood Preferences? Results from a Video Experiment." *American Journal of Sociology* 115:527–59.
- Krysan, M. and R. Farley. 2002. "Residential Preferences of Blacks: Do They Explain Persistent Segregation." *Social Forces* 80:937–80.
- Lee, B. and P. Waddell. 2010. "Residential Mobility and Location Choice: A Nested Logit Model with Sampling of Alternatives." *Transportation* 37:587–601.
- Lerman, S. 1975. "A Disaggregate Behavioral Model of Urban Mobility Decisions." Center for Transportation Studies Report no. 75-5. Cambridge, MA: Massachusetts Institute of Technology.
- Logan, J. A. 1996. "Opportunity and Choice in Socially Structured Labor Markets." *American Journal of Sociology* 102:114–60.
- Logan, J. A. 1998. "Estimating Two-sided Logit Models." Pp. 139-73 in *Sociological Methodology*, vol 28, edited by Adrian E. Raftery. Boston, MA: Blackwell Publishers.
- Logan, I., P. Hoff and M. Newton. 2008. "Two-sided Estimation of Mate Preferences for Similarities in Age, Education and Religion." *Journal of the American Statistical Association* 103:559–69.
- Louviere, I., D. Hensher and J. Swait. 2000. *Stated Choice Methods: Analysis and Application*. Cambridge, England: Cambridge University Press.
- Macy, M. and R. Wilier. 2002. "From Factors to Actors: Computational Sociology and Agent-Based Modeling." *Annual Review of Sociology* 28:143–66.
- Maddala, G. S. 1983. *Limited-dependent and Qualitative Variables in Econometrics*. Cambridge, England: Cambridge University Press.

- Manski, C. and S. R. Lerman. 1977. "The Estimation of Choice Probabilities from Choice Based Samples." *Econometrica* 45:1977–88.
- Mare, R. and E. Bruch. 2003. "Spatial Inequality, Neighborhood Mobility and Residential Segregation." Working Paper PWP-CCPR-2003-002. Los Angeles. California Center for Population Research.
- McFadden, D. 1978. "Modeling the Choice of Residential Location." *Spatial Interaction Theory and Planning Models* 25:75–96.
- McFadden, D. and K. Train. 2000. "Mixed MNL Models for Discrete Response." *Journal of Applied Econometrics* 15:447–70.
- McFadden, D. K. Train and W. Tye. 1981. "An Application of Diagnostic Tests for the Independence from Irrelevant Alternatives Property of the Multinomial Logit Model." *Transportation Research Board Record* 637: 39–46.
- Pager, D. and H. Shepherd. 2008. "The Sociology of Discrimination: Racial Discrimination in Employment, Housing, Credit and Consumer Markets." *Annual Review of Sociology* 34:181–209.
- Pinjari, A., R. Pendyala, C. Bhat and P. Waddell. 2007. "Modeling Residential Sorting Effects to Understand the Impact of Built Environment on Commute Mode Choice." *Transportation* 34:557–73.
- Quigley, J. 1985. "Consumer Choice of Dwelling, Neighborhood and Public Services." *Regional Science and Urban Economics* 15:41–63.
- Quillian, L. 1995. "Prejudice as a Response to Perceived Group Threat: Population Composition and Anti-Immigrant and Racial Prejudice in Europe." *American Sociological Review* 60:586–611.
- Quillian, L. 1999. "Migration Patterns and the Growth of High-Poverty Neighborhoods, 1970–1990." *American Journal of Sociology* 105:1–37.
- Rosenbaum, E. and S. Friedman. 2001. "Differences in the Locational Attainment of Immigrant and Native-born Households with Children in New York City." *Demography* 38:337–48.
- Roth, A. and M. Sotomayor. 1990. *Two-sided Matching: A Study in Game Theoretic Modeling and Analysis*. Cambridge, England: Cambridge University Press.
- Sastry, N, B. Ghosh-Dastidar, I Adams and A. Pabley. 2006. "The Design of a Multilevel Survey of Children, Families and Communities: The Los Angeles Family and Neighborhood Survey." *Social Science Research* 35:100024
- Schelling, T. C. 1969. "Models of Segregation." *American Economic Review* 59:488–93.
- Schelling, T. C. 1971. "Dynamic Models of Segregation." *Journal of Mathematical Sociology* 1:143–86.
- Schelling, T. C. 1978. *Micromotives and Macrobehaviour*. New York: Norton.
- Schelling, T. C. 2006. "Some Fun, Thirty-five Years Ago." *Handbook of Computational Economics* 2:1639–44.
- Sermons, M. and F. Koppelman. 2001. "Representing the Differences between Female and Male Commuting Behavior in Residential Location Models." *Journal of Transport Geography* 9:101–10.
- Skrondal, A. and Rabe-Hesketh, S. 2004. *Generalized Latent Variable Modeling: Multilevel, Longitudinal and Structural Equation Models*. Boca Raton, FL: Chapman and Hall/CRC.
- Small, Kenneth and Cheng Hsiao. 1985. "Multinomial Logit Specification Tests." *International Economic Review* 26:619–27.

- South, S. I. and K. D. Crowder. 1997. "Escaping Distressed Neighborhoods: Individual, Community and Metropolitan Influences." *American Journal of Sociology* 4: 1040–1084.
- South, S. I. and K. D. Crowder. 1998. "Housing Discrimination and Residential Mobility: Impacts for Blacks and Whites." *Population Research and Policy Review* 17:369–87.
- Steglich, C., T. Snijders, and M. Pearson. 2010. "Dynamic Networks and Behavior: Separating Selection from Influence." Pp. 329–93 in *Sociological Methodology* vol 40, edited by Tim Futing Liao. Washington, DC: American Sociological Association.
- Todd, P., F. Billari, and J. Simao. 2005. "Aggregate Age-at-Marriage Patterns from individual Mate-search Heuristics." *Demography* 42: 559–74.
- Train, K. 2003. *Discrete Choice Methods with Simulation*. Cambridge, England: Cambridge University Press.
- Tuljapurkar, S., E. Bruch, and R. Mare. 2010 "Neighborhoods and Individual Preferences: A Markovian Model." Unpublished manuscript.
- Waddell, P. 1996. "Accessibility and Residential Location: The Interaction of Workplace, Residential Mobility, Tenure and Location Choices." Presented at the Lincoln Land Institute T. E. Conference, Cambridge, MA.
- Wen, C. H. and F. S. Koppelman. 2001. "The Generalized Nested Logit Model." *Transportation Research Part B: Methodological* 35:627–11.
- Zeng, Z. and Y. Xie. 2008. "A Preference-Opportunity-Choice Framework with Applications to Intergroup Friendship." *American Journal of Sociology* 114:615–48.

Bios

Elizabeth E. Bruch is an Assistant Professor in Sociology and Complex Systems at the University of Michigan, and an Affiliate of the Population Studies Center at the Institute for Social Research. Her work spans a broad array of population phenomena in which the actions of individuals and other units (such as couples or neighborhoods) are dynamically interdependent. She is currently working on problems related to income inequality and income segregation, statistical modeling of individuals' decision-making strategies, and the relationship between mate preferences, opportunity structure, and patterns of assortative mating.

Robert D. Mare is a Distinguished Professor of Sociology at the University of California-Los Angeles and Faculty Affiliate of the California Center for Population Research. His research focuses on residential mobility and segregation, marriage markets and assortative mating, the interdependence of demographic processes and social mobility, and models for multigenerational effects. With Robert Sampson he is supervising a coordinated re-interview study of residential mobility and mixed income neighborhoods based on the Los Angeles Family and Neighborhood Survey and the Project on Human Development in Chicago Neighborhoods. He was President of the Population Association of America in 2010 and is a Member of the National Academy of Sciences and Fellow of the American Academy of Arts and Sciences.