Modeling the Choice of Residential Location

Daniel McFadden, Department of Economics, Massachusetts Institute of Technology, Cambridge

The problem of translating the theory of economic choice behavior into concrete models suitable for analyzing housing location is discussed. The analysis is based on the premise that the classical, economically rational consumer will choose a residential location by weighing the attributes of each available alternative and by selecting the alternative that maximizes utility. The assumption of independence in the commonly used multinomial logit model of choice is relaxed to permit a structure of perceived similarities among alternatives. In this analysis, choice is described by a multinomial logit model for aggregates of similar alternatives. Also discussed are methods for controlling the size of data collection and estimation tasks by sampling alternatives from the full set of alternatives.

The classical, economically rational consumer will choose a residential location by weighing the attributes of each available alternative—accessibility to work place, shopping, and schools; quality of neighborhood life and availability of public services; costs, including price, taxes, and travel costs; and dwelling characteristics, such as age, number of rooms, type of appliances—and by choosing the alternative that maximizes utility.

This paper considers the problem of translating the theory of economic choice behavior into concrete models suitable for the empirical analysis of housing location. We are concerned particularly with two problems in the modeling of individual, or disaggregate, choice among residential locations. First, there may be a structure of perceived similarities among alternatives that invalidates the commonly used joint multinomial logit model of choice. We treat individual dwelling units as the basic alternatives among which choice is made. Each unit will have a list of attributes, observed and unobserved, to which the individual responds. We assume that the space of attributes, including unobserved attributes, is sufficiently rich so that each physical dwelling unit is represented by a unique point in attribute space. Of course, the individual may perceive two dwellings that are similar in some attributes as quite similar overall; it is the impact of such perceptions on choice that I wish to model.

I shall introduce a family of probabilistic choice models, of which the joint multinomial logit model is a special case, that has the property of aggregating dwelling units perceived as similar. The weight given to an aggregate of alternatives in the choice process will depend on the degree of perceived similarity.

At one extreme, the elements of the aggregate will be perceived as independent, and choice will be described by a multinomial logit model with individual dwellings as alternatives. At the other extreme, all dwellings with the same observed attributes will be perceived as virtually the same, and choice will be described by a multinomial logit model with dwelling types, which are distinguished by observed attributes, as the alternatives. The family of models introduced here permits empirical estimation of the degree of perceived similarity and tests of the two extreme cases mentioned above.

The second problem treated in this paper is that of estimation of individual choice models when the number of elemental alternatives is impractically large. The section on limiting the number of alternatives establishes that, if choice among a set of alternatives is described by a multinomial logit model, then the model can be estimated by sampling from the full set of alternatives, with appropriate adjustment in the estimation mechanism. Thus, estimation can be carried out with limited data collection and computation.

The solutions I give to the two problems above will be applied to empirical studies of housing location by Quigley (1) and Lerman (2).

THEORY OF HOUSING LOCATION CHOICE

Assume the classical model of the rational, utility-maximizing consumer. Suppose the consumer faces a residential location decision, with a choice of communities indexed $c = 1, \ldots, C$ and dwellings indexed $n = 1, \ldots, N_c$ in community c. The consumer will have a utility U_{cn} for alternative cn, which is a function of the attributes of this alternative, including accessibility, quality of public services, neighborhood and dwelling characteristics, etc., as well as a function of the consumer's characteristics, such as age, family size, and income. The consumer will choose the alternative that maximizes his utility.

Not all attributes of alternatives will be observed. The unobserved variables will have some probability distribution in the population, conditioned on the value of the observed variables. If the observer knows the form of the utility function and the probability distribution of unobserved variables, then probabilistic statements can be made about the expected distribution of choices:

$$P_{cn} = \text{Prob}[U_{cn} > U_{bm} \text{ for } bm \neq cn] \tag{1}$$

where P_{cn} denotes the probability of choice cn and the probability on the right side is defined with respect to the distribution of unobserved variables. The econometric approach to this problem is to specify, as a maintained hypothesis, a class of utility forms and distributions from which one member can be statistically identified.

Consider the decomposition $U_{cn} = V_{cn} + \epsilon_{cn}$ of utility into a term V_{cn} that is a function specified up to a finite vector of unknown parameters, of observed variables, and a term ϵ_{cn} summarizing the contribution of unobserved variables. Hereafter, V_{cn} will be called the strict utility of cn. Let $\underset{\sim}{\epsilon}$ denote the vector $(\epsilon_{11}, \ldots, \epsilon_{1N_1}, \ldots, \epsilon_{C1}, \ldots, \epsilon_{CN_C})$ and let $F(\underset{\sim}{\epsilon})$ denote the cumulative distribution function of $\underset{\sim}{\epsilon}$. Then Equation 1 can be written

$$P_{cn} = \int_{\epsilon_{cn}=-\infty}^{+\infty} F_{cn}(\langle V_{cn} + \epsilon_{cn} - V_{dm}\rangle) d\epsilon_{cn} \tag{2}$$

where F_{cn} denotes the derivative of F with respect to its cn argument, and $\langle V_{cn} + \epsilon_{cn} - V_{dm}\rangle$ denotes a vector with components indexed by dm. An econometric model of choice is specified by choosing a parametric form for

V_{dn} and a parametric distribution F.

MULTINOMIAL LOGIT MODEL

An empirically important specialization of Equation 2 is the multinomial logit model,

$$P_{cn} = \exp(V_{cn}) \Big/ \sum_{b=1}^{C} \sum_{m=1}^{N_b} \exp(V_{bm}) \tag{3}$$

obtained by assuming the ϵ_{cn} to be independently and identically distributed with the extreme value distribution,

$$\text{Prob}(\epsilon_{cn} < \epsilon) = \exp(-e^{-\epsilon}) \tag{4}$$

This model was proposed as a theory of psychological choice behavior by Luce (3). Its econometric analysis has been investigated by McFadden (4, 5) and Nerlove and Press (6). A particular structural feature of this model, termed independence from irrelevant alternatives by Luce, is that the relative odds for any two alternatives are independent of the attributes, or even the availability, of any other alternative. This property is extremely useful in simplifying econometric estimation and forecasting (7) but can be shown to be implausible for choice problems where it is unreasonable to assume that the ϵ_{cn} are statistically independent (8, 9).

For later analysis, it will be useful to rewrite the joint choice Equation 3 in terms of a conditional choice probability $P_{n|c}$ for dwelling, given community, and a marginal choice probability P_c for community. The strict utility V_{cn} can often be expressed in an additively separable, linear-in-parameters form

$$V_{cn} = \beta' x_{cn} + \alpha' y_c \tag{5}$$

where x_{cn} is a vector of observed attributes that vary with both community and dwelling (e.g., work-place accessibility), y_c is a vector of observed attributes that vary only with community (e.g., availability of community recreation facilities), and α and β are vectors of unknown parameters. Hereafter, we assume the structure of Equation 5. From Equations 3 and 5, one obtains the formulas

$$P_{n|c} = \exp(V_{cn}) \Big/ \sum_{m=1}^{N_c} \exp(V_{cm}) = \exp(\beta' x_{cn}) \Big/ \sum_{m=1}^{N} \exp(\beta' x_{cm}) \tag{6}$$

$$\begin{aligned} P_c &= \sum_{n=1}^{N_c} \exp(V_{cn}) \Big/ \sum_{b=1}^{C} \sum_{m=1}^{N_b} \exp(V_{bm}) \\ &= \exp(\alpha' y_c) \left[\sum_{n=1}^{N_c} \exp(\beta' x_{cn})\right] \Big/ \sum_{b=1}^{C} \exp(\alpha' y_b) \left[\sum_{m=1}^{N_b} \exp(\beta' x_{bm})\right] \end{aligned} \tag{7}$$

Define an inclusive value

$$I_c = \log\left[\sum_{n=1}^{N_c} \exp(\beta' x_{cn})\right] \tag{8}$$

Then, Equations 6 and 7 can be rewritten

$$P_{n|c} = \exp(\beta' x_{cn})/\exp(I_c) \tag{9}$$

$$P_c = \exp(\alpha' y_c + I_c) \Big/ \sum_{b=1}^{C} \exp(\alpha' y_b + I_b) \tag{10}$$

One method of estimating the joint model (Equation 3) is to first estimate the parameters β from the conditional choice model (Equation 6). Next define I_c using the log of the denominator of the estimated equation. Finally, estimate the parameters α from the marginal probability model (Equation 10), given I_c. This sequential approach to estimation economizes on the number of alternatives and the number of parameters considered at each stage of estimation, with some loss of efficiency relative to direct estimation of the joint model (Equation 3).

NESTED LOGIT MODEL

An empirical generalization of the multinomial logit model in the form of Equations 9 and 10 is obtained by allowing the inclusive value I_c in the latter to have a coefficient other than one:

$$P_c = \exp[\alpha' y_c + (1-\sigma) I_c] \Big/ \sum_{b=1}^{C} \exp[\alpha' y_b + (1-\sigma) I_b] \tag{11}$$

where $(1 - \sigma)$ is a parameter. The model represented by Equations 9 and 11, termed the "nested logit model," was first used with the estimation procedure described above, but with an unsatisfactory definition of inclusive value (9). Ben-Akiva has suggested the correct definition (Equation 8) of inclusive value and explored the implications of fitting the joint model or various nested models. Amemiya (10) corrects an error in the formula used in the earlier studies to compute the standard errors of estimates in the last stage of the sequential estimation procedure [see also McFadden (11)].

GENERALIZED EXTREME VALUE MODEL

I shall now introduce a family of choice models, derived from stochastic utility maximization, that includes multinomial and nested logit. This family allows a general pattern of dependence among the unobserved attributes of alternatives and yields an analytically tractable closed form for the choice probabilities. The following result characterizes the family.

Suppose $G(y_1, \ldots, y_J)$ is a nonnegative, homogeneous-of-degree-one function of $(y_1, \ldots, y_J) \geq 0$. Suppose $G \to \infty$ if $y_i \to \infty$ for each i, and for k distinct components $i_1, \ldots, i_k$, $\partial^k G/\partial y_1 \ldots y_{i_k}$ is nonnegative if k is odd and nonpositive if k is even. Then

$$P_i = \{\exp(V_i) G_i[\exp(V_1), \ldots, \exp(V_J)]\}/G[\exp(V_1), \ldots, \exp(V_J)] \tag{12}$$

defines a probabilistic choice model from alternatives $i = 1, \ldots, J$, which is consistent with utility maximization. Further, expected maximum utility, defined by

$$\bar{U} = \int_{\underset{\sim}{\epsilon}=-\infty}^{+\infty} \max_i (V_i + \epsilon_i) f(\underset{\sim}{\epsilon})\, d\underset{\sim}{\epsilon} \tag{13}$$

(with f the density for F), satisfies

$$\bar{U} = \log G[\exp(V_1), \ldots, \exp(V_J)] + \gamma \tag{14}$$

where $\gamma = 0.57721$ is Euler's constant, and

$$P_i = \partial \bar{U}/\partial V_i \tag{15}$$

I have proved this result (11).

The special case $G(y_1, \ldots, y_J) = \sum_{j=1}^{J} y_j$ yields the multinomial logit model. An example of a more general

G function satisfying the hypotheses of the theorem is

$$G(y) = \sum_{m=1}^{M} a_m \left[\sum_{i\epsilon B_m} y_i^{1/(1-\sigma_m)}\right]^{1-\sigma_m} \tag{16}$$

where $B_m \subset \{1, \ldots, J\}$, $\bigcup_{m=1}^{M} B_m = \{1, \ldots, J\}$, $a_m > 0$, and $0 \le \sigma_m < 1$.

For the bivariate case with a single class m, Equation 16 reduces to

$$G(y) = \left[y_1^{1/(1-\sigma)} + y_2^{1/(1-\sigma)}\right]^{1-\sigma} \tag{17}$$

The bivariate extreme value distribution based on this form has been studied by Oliveira (12, 13), who shows that σ is the product-moment correlation between the two variates. In the general case of Equation 16, σ_m can be interpreted as an index of the similarity of the unobserved attributes B_m. However, the relation between the σ_m and product-moment correlations between the alternatives is more complex.

The choice probabilities for Equation 16 satisfy

$$P_i = \sum_{m=1}^{M} P(i|B_m)P(B_m) \tag{18}$$

where

$$P(i|B_m) = \exp[V_i/(1-\sigma_m)] \Big/ \sum_{j\epsilon B_m} \exp[V_j/(1-\sigma_m)] \tag{19}$$

if $i \epsilon B_m$, and

$$P(i|B_m) = 0 \tag{20}$$

if $i \notin B_m$, with $P(i|B_m)$ denoting the conditional probability, and

$$P(B_m) = a_m \left\{\sum_{j\epsilon B_m} \exp[V_j/(1-\sigma_m)]\right\}^{1-\sigma_m} \div \sum_{n=1}^{M} a_n \left\{\sum_{k\epsilon B_n} \exp[V_k/(1-\sigma_n)]\right\}^{1-\sigma_n} \tag{21}$$

Choice probabilities of the form of Equation 18 were apparently first derived, for the case of three alternatives and $B_1 = \{1\}$, $B_2 = \{2, 3\}$, by Scott Cardell. For the case of disjoint B_m, the form of Equation 18 was treated independently by Daly and Zachary (14), Williams (15), and Ben-Akiva and Lerman (16). The demonstration by Daly and Zachary that Equation 18 is consistent with random utility maximization is noteworthy in that it permits generalization of the generalized extreme value model and provides a powerful tool for testing the consistency of choice models.

Consider an example of Equation 16,

$$G(y_1, y_2, y_3) = y_1 + \left[y_2^{1/(1-\sigma)} + y_3^{1/(1-\sigma)}\right]^{1-\sigma} \tag{22}$$

where alternative 1 represents a dwelling in one community, and alternatives 2 and 3 represent dwellings of a similar type in a second community. Let V_i be the strict utility of alternative i. The choice probabilities when the three alternatives are offered are, from Equation 18,

$$P(1|1,2,3) = \exp(V_1)/\left(\{\exp(V_1) + \exp[V_2/(1-\sigma)] + \exp[V_3/(1-\sigma)]\}^{1-\sigma}\right) \tag{23}$$

$$P(2|1,2,3) = \exp[V_2/(1-\sigma)]\{\exp[V_2/(1-\sigma)] + \exp[V_3/(1-\sigma)]\}^{-\sigma} \div \left(\exp(V_1) + \{\exp[V_2/(1-\sigma)] + \exp[V_3/(1-\sigma)]\}^{1-\sigma}\right) \tag{24}$$

where $P(i|A)$ denotes the probability that i is chosen from the alternatives A. If only alternatives 1 and 2 are available, then the choice probability (obtained from Equation 23 by setting $V_3 = -\infty$) has the binomial form

$$P(1|1,2) = \exp(V_1)/[\exp(V_1) + \exp(V_2)] \tag{25}$$

If only alternatives 2 and 3 are available, the choice probability again has a binomial logit form,

$$P(2|2,3) = \exp[V_2/(1-\sigma)]/\{\exp[V_2/(1-\sigma)] + \exp[V_3/(1-\sigma)]\} \tag{26}$$

Examining the choice probabilities of Equations 23 and 24 when all three alternatives are available, the value $\sigma = 0$ gives multinomial logit probabilities, while the limiting value $\sigma \to 1$ gives the probabilities

$$P(1|1,2,3) = \exp(V_1)/\{\exp(V_1) + \max[\exp(V_2), \exp(V_3)]\} \tag{27}$$

$$\begin{aligned} P(2|1,2,3) &= \exp(V_2)/[\exp(V_2) + \exp(V_3)] \quad \text{if } V_2 > V_3 \\ &= \tfrac{1}{2}\exp(V_2)/[\exp(V_2) + \exp(V_3)] \quad \text{if } V_2 = V_3 \\ &= 0 \quad \text{if } V_2 < V_3 \end{aligned} \tag{28}$$

In this extreme case, the consumer will treat two alternatives with identical strict utilities $V_2 = V_3$ as a single alternative in comparisons with alternative 1.

RELATION BETWEEN THE NESTED LOGIT AND THE GENERALIZED EXTREME VALUE MODEL

The choice probabilities in Equation 18 can be specialized to the nested logit model given by Equations 9 and 11, as we shall now show. This result establishes that nested logit models are consistent with stochastic utility maximization and that the coefficient of inclusive value provides an estimate of the similarity of the unobserved terms in the first level of the nested model. Hence, it is possible to estimate some generalized extreme value choice models using nested logit models and inclusive values. Further, the generalized extreme value choice models provide a generalization of nested logit models and could be estimated directly to test for the presence and form of a nested (or tree) structure for similarities.

To obtain the nested logit model Equations 9 and 11 from Equation 18: replace the alternative index i with the double index cn for community c and dwelling n; replace m by c; assume the sets B_c have the form $B_c = \{c1, \ldots, cN_c\}$; and assume the similarity coefficients have a common value σ. Then Equation 18 becomes

$$P_{cn} = \exp[V_{cn}/(1-\sigma)]\left\{\sum_{m=1}^{N_c} \exp[V_{cm}/(1-\sigma)]\right\}^{-\sigma} \div \left(\sum_{b=1}^{C}\left\{\sum_{m=1}^{N_b} \exp[V_{bm}/(1-\sigma)]\right\}^{1-\sigma}\right) \tag{29}$$

implying that

$$P_c = \sum_{n=1}^{N_c} P_{cn} = \left\{\sum_{m=1}^{N_c} \exp[V_{cm}/(1-\sigma)]\right\}^{1-\sigma} \div \left(\sum_{b=1}^{C}\left\{\sum_{m=1}^{N_b} \exp[V_{bm}/(1-\sigma)]\right\}^{1-\sigma}\right) \tag{30}$$

and that

$$P_{n|c} = P_{cn}/P_c = \exp[V_{cn}/(1-\sigma)] \Big/ \sum_{m=1}^{N_c} \exp[V_{cm}/(1-\sigma)] \tag{31}$$

Recalling that $V_{cn} = \beta' x_{cn} + \alpha' y_c$, these formulas can be written

$$P_c = \exp[\alpha' y_c + (1-\sigma) I_c] \Big/ \sum_{b=1}^{C} \exp[\alpha' y_b + (1-\sigma) I_b] \tag{32}$$

$$P_{n|c} = \exp[\beta' x_{cn}/(1-\sigma)] \Big/ \sum_{m=1}^{N_c} \exp[\beta' x_{cm}/(1-\sigma)]$$
$$= \exp[\beta' x_{cn}/(1-\sigma)]/\exp(I_c) \tag{33}$$

$$I_c = \log \sum_{m=1}^{N_c} \exp[\beta' x_{cm}/(1-\sigma)] \tag{34}$$

Hence, the nested logit model is a specialization of the generalized extreme value model, with the coefficient $1 - \sigma$ of inclusive value an index of the degree of independence of random terms for alternative dwellings in the same community.

This argument can be extended to trees of any depth. A sufficient condition for a nested logit model to be consistent with stochastic utility maximization is that the coefficient of each inclusive value lie in the unit interval.

LIMITING THE NUMBER OF ALTERNATIVES CONSIDERED

Consider application of the joint multinomial logit model Equation 3 to the demand for housing, with alternatives indexed by community and by dwelling within the community. Ideally, the functional form of the model is appropriate for describing choice among the full set of alternatives available to consumers, and it is practical in terms of data collection and statistical analysis to study decision behavior at this level.

In practice, the number of available alternatives at the most disaggregate level often imposes infeasible data-processing requirements and strains the plausibility of the independence from irrelevant alternatives property of the multinomial logit functional form, as in the example of similar dwellings in the same community that are likely to have similar unobserved attributes.

Consider first the problem where enumeration of all alternatives is impractical but where data on selected disaggregate alternatives can be observed. If the multinomial logit functional form is valid, we shall establish the result that consistent estimates of the parameters of the strict utility function can be obtained from a fixed or random sample of alternatives from the full choice set.

Let C denote the full choice set. We shall assume it does not vary over the sample; however, this is inessential and can easily be generalized. Let $P(i|C, z, \theta^*)$ denote the true selection probabilities, where θ is a vector of parameters, and z is a vector of explanatory variables. We assume the choice probabilities satisfy the independence from irrelevant alternatives assumption:

$$i \in D \subseteq C \rightarrow P(i|C,z,\theta) = P(i|D,z,\theta) \sum_{j \in D} P(j|C,z,\theta) \tag{35}$$

which characterizes the multinomial logit model.

Now suppose for each case that a subset D is drawn from the set C according to a probability distribution $\pi(D|i, z)$, which may but need not be conditioned on the observed choice i. The observed choice may be either in or out of the set D. Examples of π distributions are (a) choose a fixed subset D of C independent of the observed choice, (b) choose a random subset D of C containing the observed choice, and (c) choose a subset D of C consisting of the observed choice i and one or more other alternatives, selected randomly.

We give two examples of distributions of type (c):

1. (c-1): Suppose D is always selected to be a two-element set containing i and one other alternative selected at random. If J is the number of alternatives in C, then

$$\pi(D|i,z) = 1/(J-1) \qquad \text{if } D = [i,j] \text{ and } j \neq i \tag{36}$$

or zero otherwise.

2. (c-2): Suppose C is partitioned into sets $\{C_1, \ldots, C_M\}$, with J_m elements in C_m, and suppose D is formed by choosing i (from the partition set C_n) and one randomly selected alternative from each remaining partition set. Then

$$\pi(D|i,z) = J_n / \prod_{m=1}^{M} J_m \qquad \text{if } i \in D,\ M = \#(D) \tag{37}$$

and $D \cap C_m \neq \phi$ for m = 1, ..., M, or zero otherwise.

The π distributions of the types (a), (b), and (c-1) and (c-2) all satisfy the following basic property, which guarantees that, if an alternative j appears in an assigned set D, then it has the logical possibility of being an observed choice from the set D, in the sense that the assignment mechanism could assign the set D if a choice of j is observed.

Positive Conditioning Property

If $j \in D \subseteq C$ and $\pi(D|i, z) > 0$, then $\pi(D|j, z) > 0$.

The π distributions (a), (b), and (c-1) but not (c-2) satisfy a stronger condition.

Uniform Conditioning Property

If $i, j \in D \subset C$, then $\pi(D|i, z) = \pi(D|j, z)$.

Consider a sample n = 1, ..., N, with the alternative chosen on case n denoted i_n, and D_n denoting the choice set assigned to this case from the distribution $\pi(D|i_n, z_n)$. Observations with an observed choice not in the assigned set of alternatives are assumed to be excluded from the sample. Write the multinomial logit model in the form

$$P(i|C,z,\theta) = \exp[V_i(z,\theta)] \Big/ \sum_{j \in C} \exp[V_j(z,\theta)] \tag{38}$$

where $V_i(z, \theta)$ is the strict utility of alternative i.

If $\pi(D|i, z)$ satisfies the positive conditioning property and the choice model is multinomial logit, then maximization of the modified likelihood function

$$L_N = (1/N) \sum_{n=1}^{N} \log \Big\{ \exp[V_{i_n}(z_n,\theta) + \log \pi(D_n|i_n,z_n)]$$
$$\div \sum_{j \in D} \exp[V_j(z_n,\theta) + \log \pi(D_n|j,z_n)] \Big\} \tag{39}$$

yields, under normal regularity conditions, consistent estimates of θ^*. When $\pi(D \mid i, z)$ satisfies the uniform conditioning property, then Equation 39 reduces to the standard likelihood function,

$$L_N = (1/N) \sum_{n=1}^{N} \log \left\{ \exp[(V_{i_n}(z,\theta)] \Big/ \sum_{j \epsilon D} \exp[V_j(z_n,\theta)] \right\} \tag{40}$$

A proof is given by McFadden (17).

In conclusion, analysis of housing location can be carried out with a limited number of alternatives, which facilitates data collection and processing, provided the choice process is described by the multinomial logit model. If a mechanism such as (c-2) is used to select alternatives, the likelihood function should be modified to the form of Equation 39 to obtain consistent estimates of all parameters. If a non-modified likelihood function is used, estimation can still be carried out satisfactorily provided the effect of the selection mechanism for alternatives is absorbed by class-specific parameters. Caution is required in this case in verifying that the configuration of class-specific variables in the model is adequate to accommodate the selection mechanism effects, and in interpreting the estimates of class-specific parameters.

AGGREGATION OF ALTERNATIVES AND THE TREATMENT OF SIMILARITIES

The preceding section has shown that, when the multinomial logit functional form is valid, estimation can be carried out by using randomly selected "representative" alternatives from each "class" of elemental alternatives, where the classes are defined by the analyst. Community and dwelling type were classification criteria mentioned in the earlier examples. Analysis of choice among classes by identifying them with "representative" members can be viewed as a method of aggregation of alternatives.

We shall now consider alternative methods of aggregation that can be employed when the multinomial logit form fails because of dependence between unobserved attributes of different alternatives within a class.

Again consider a consumer faced with a choice of housing locations in $c = 1, \ldots, C$ communities, with $n = 1, \ldots, N_c$ dwellings in community c, all of which have common unobserved community attributes. This introduces a dependence that conflicts with the assumptions of the joint multinomial logit model. To represent this dependence we shall assume that the choice probabilities have the nested logit structure of Equations 32-34, with σ a measure of the degree to which dwellings within a class c are perceived as similar. When $\sigma = 0$, Equation 32 reduces to the multinomial logit model, and in the limit when $\sigma = 1$, it reduces to

$$P_c = \exp(\alpha' y_c + \max \beta' x_{cn}) \Big/ \sum_{b=1}^{C} \exp(\alpha' y_b + \max \beta' x_{bn}) \tag{41}$$

An analysis of housing demand by Quigley (1) using Pittsburgh data employs a model of the form of Equation 41. In Quigley's model, the nesting of community and housing type is reversed, with c denoting housing type, and n denoting specific dwelling, identified by community and location. Quigley assumes a sufficient structure on location choice so that the term $\max_n \beta' x_{cn}$ can be computed prior to parameter estimation. Then Equation 41 can be treated as an ordinary multinomial logit model.

In an analysis of neighborhood choice using Washington, D.C., data, Lerman (2) estimates a model of the form

$$P_c = \exp[\alpha' y_c + X_c^* + (1-\sigma)\log N_c] \div \sum_{b=1}^{C} \exp[\alpha' y_b + X_b^* + (1-\sigma)\log N_b] \tag{42}$$

where c indexes census tracts and X_c^* is an "average" of the utility terms $\beta' x_{cn}$ of the dwellings in tract c. He notes that $\log N_c$ is

> the measure of tract size required to correct for the fact that a census tract is actually a group of housing units. Other conditions being equal, a very large tract (i.e., one with a large number of housing units) would have a higher probability of being selected than a very small one, since the number of disaggregate opportunities is greater in the former than the latter. If all units of a particular type in a given zone are relatively homogeneous and the [joint multinomial] logit model applies to each individual unit, then the appropriate term to correct for tract size is the natural logarithm of the number of units [with] a coefficient of one.

Noting the model (Equation 41) as a second extreme case, Lerman concludes that "if the assumptions of the [joint multinomial] logit model are violated, the coefficient may differ from one." Lerman estimates the coefficient of $\log N_c$ to be $1 - \sigma = 0.492$, with a standard error of 0.094. Hence, σ satisfies the hypotheses of theorem 1 and is significantly different from both zero and one.

In the nested logit model (Equations 32 and 34), the inclusive value can be rewritten

$$I_c = [X_c^*/(1-\sigma)] + \log N_c + \log 1/N_c \sum_{m=1}^{N_c} \times \exp[(\beta' x_{cm} - X_c^*)/(1-\sigma)] \tag{43}$$

If a tract c is homogeneous in terms of observed variables so that $\beta' x_{cn} = X_c^*$, then the last term in Equation 43 vanishes, and the choice probability for the nested logit model (Equation 32) is exactly the Lerman model (Equation 42). This establishes the consistency of the Lerman model with stochastic utility maximization and supports his conclusion that the coefficient of $\log N_c$ indexes the degree of independence of the alternatives within a tract. The same argument can be used to interpret Quigley's model, with $X_c^* = \max \beta' x_{cn}$.

When X_c^* is the mean of $\beta' x_{cn}$, and not all $\beta' x_{cn} = X_c^*$, the convexity of the exponential implies

$$1/N_c \sum_{m=1}^{N_c} \exp[(\beta' x_{cm} - X_c^*)/(1-\sigma)] > 1 \tag{44}$$

and hence $I_c \geq [X_c^*/(1-\sigma)] + \log N_c$, with the difference of the two sides of the inequality depending on the variance of $\beta' x_{cn}$. One limiting case of Equation 43 that is of interest occurs when the number of dwellings within a tract is large and the x_{cn} behave as if they were independently identically normally distributed with mean X_c^*. Let ω_c^2 denote the variance of $\beta' x_{cn}$. If $N_c = r_c N$, with r_c fixed and $N \to \infty$, then

$$P_c \to \frac{\exp\{[\alpha' y_c + \beta' x_c^* + (1-\sigma)\log r_c + \frac{1}{2}\omega_c^2]/(1-\sigma)\}}{\sum_{b=1}^{C} \exp\{[\alpha' y_b + \beta' x_b + (1-\sigma)\log r_b + \frac{1}{2}\omega_b^2]/(1-\sigma)\}} \tag{45}$$

When the disaggregate data x_{cn} are not observed, but their distribution can be approximated or estimated, and ω_c is known, then Equation 45 can be used with standard multinomial logit estimation programs to provide estimates of σ and β. If r_o is unobserved, then it can be estimated when ω_o is known; when y_o contains a tract-specific dummy variable, however, the tract-specific coefficient and r_o are unidentified. This suggests one interpretation of tract-specific coefficients as indicating in part the number of equivalent disaggregate alternatives contained in the tract.

When ω_o is not known, but is known to have the structure $\omega_o^2 = \beta' \Omega_o \beta$, and the variables x_{on} are multivariate normal with covariance matrix Ω_o, direct estimation of β, σ, and α is possible. A modification of standard multinomial logit programs to handle nonlinear constraints on β would be required for full maximum likelihood estimation. Alternately, consistent estimators could be obtained by writing out the terms in the quadratic form $\beta' \Omega_o \beta$ as independent parameters and ignoring constraints.

CONCLUSION

This paper has considered the problem of modeling disaggregate choice of housing location when the number of disaggregate alternatives is impractically large and when the presence of a structure of similarities between alternatives invalidates the commonly used joint multinomial logit choice model. Theorems on sampling from the full set of alternatives and on generalizations of the multinomial logit model structure to accommodate similarities provide methods for circumventing these problems. Studies of housing demand by Quigley (1) and Lerman (2) motivate the analysis and illustrate its applicability.

ACKNOWLEDGMENTS

This research was motivated by, and has benefitted from, discussions with Moshe Ben-Akiva, Steven Lerman, Charles Manski, and William Tye, and comments by Anthony E. Smith and Folke Snickars. I am indebted to the National Science Foundation for research support. This paper abstracts a more complete research report.

REFERENCES

1. J. M. Quigley. Housing Demand in the Short-Run: An Analysis of Polytomous Choice. Explorations in Economic Research, Vol. 3, No. 1, Winter 1976, pp. 76-102.
2. S. R. Lerman. Location, Housing, Automobile Ownership, and Mode to Work: A Joint Choice Model. TRB, Transportation Research Record 610, 1977, pp. 6-11.
3. R. D. Luce. Individual Choice Behavior. Wiley, New York, 1959.
4. D. McFadden. Conditional Logit Analysis of Qualitative Choice Behavior. In Frontiers in Econometrics (P. Zarembka, ed.), Academic Press, New York, 1973.
5. D. McFadden. Quantal Choice Analysis: A Survey. Annals of Economic and Social Measurement, Vol. 5, No. 4, 1976, pp. 363-390.
6. M. Nerlove and J. Press. Univariate and Multivariate Log-Linear and Logistic Models. RAND, Rept. No. R-1306-EDA/NIH, 1973.
7. D. McFadden, W. Tye, and K. Train. Diagnostic Tests for the Independence From Irrelevant Alternatives Property of the Multinomial Logit Model. Paper presented at the 57th Annual Meeting, TRB, 1978.
8. G. Debreu. Review of R. Luce, Individual Choice Behavior. American Economic Review, Vol. 50, 1960, pp. 186-188.
9. T. Domencich and D. McFadden. Urban Travel Demand: A Behavioral Analysis. North-Holland, Amsterdam, 1975.
10. T. Amemiya. Specification and Estimation of a Multinomial Logit Model. Institute of Mathematical Studies in the Social Sciences, Stanford Univ., Stanford, CA, Technical Rept. No. 211, 1976.
11. D. McFadden. Econometric Models of Probabilistic Choice. In Econometric Analysis of Discrete Data (C. Manski and D. McFadden, eds.), MIT Press, Cambridge, MA, 1979.
12. J. T. de Oliveira. Extremal Distributions. Revista de Faculdada du Ciencia, Lisboa, Serie A, Vol. 7, 1958, pp. 215-227.
13. J. T. de Oliveira. La Representation des distributions extrémales bivariés. Bulletin of the Intional Statistical Institute, Vol. 33, 1961, pp. 477-480.
14. A. Daly and S. Zachary. Improved Multiple Choice Models. Planning and Transport Research and Computation (International), London, 1976.
15. H. C. L. Williams. On the Formation of Travel Demand Models and Economic Evaluation Measures of User Benefit. Environment and Planning, Vol. A.9, 1977, pp. 285-344.
16. M. Ben-Akiva and S. Lerman. Disaggregate Travel and Mobility Choice Models and Measures of Accessibility. Paper presented at the 3rd International Conference on Behavioral Travel Modeling, Tanenda, Australia, 1977.
17. D. McFadden. Modelling the Choice of Residential Location. In Spatial Interaction Theory and Planning Models (A. Karlqvist, L. Lundqvist, F. Snikars, and J. Weibull, eds.), North-Holland, Amsterdam, 1978.

Publication of this paper sponsored by Committee on Passenger Travel Demand Forecasting and Committee on Traveler Behavior and Values.