12-1-1977

Modelling the Choice of Residential Location

Daniel McFadden

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS

AT YALE UNIVERSITY

Box 2125, Yale Station
New Haven, Connecticut 06520

COWLES FOUNDATION DISCUSSION PAPER NO. 477

MODELLING THE CHOICE OF RESIDENTIAL LOCATION

Daniel McFadden

December 8, 1977

MODELLING THE CHOICE OF RESIDENTIAL LOCATION

by

Daniel McFadden*
University of California, Berkeley and Yale University

1 INTRODUCTION

The classical economically rational consumer will choose a residential location by weighing the attributes of each available alternative — accessibility of workplace, shopping, and schools; quality of neighborhood life and the availability of public services; costs, including housing price, taxes, and travel costs; dwelling characteristics, such as age, number of rooms, type of appliances; and so forth — and picking the alternative which maximizes utility. Housing prices and the supply of new dwelling units will adjust to reconcile consumer tastes with the existing housing stock at each point in time. Theoretical models of urban location often posit a population of consumers with identical tastes, and a housing market in which prices adjust frictionlessly to an equilibrium in which the consumer is indifferent

*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, Grant No. SOC75-22657, for research support.

among all housing alternatives. Then, housing price is the carrier of all information on consumer tastes for public services, accessibility, and dwelling characteristics. This observation can be taken as the basis for empirical analysis of the housing attributes entering the determination of price; see, for example, Pollakowski (1974).

In reality, consumers vary substantially in their tastes for housing, and may also display bounded rationality, with the result that a distribution of responses would result from presentation of the same apparent alternatives to each consumer in a population. Further, housing markets may be slow to adjust to equilibrium — the importance of middle-men and speculators in the market is an indication that disequilibria occur, making arbitrage a profitable activity. Then, there is something to be learned about consumer tastes and behavior from the study of consumer housing location decisions.

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 modelling of individual, or disaggregate, choice among residential locations. First, there may be a structure of perceived similarities between alternatives which invalidates the commonly used joint multinomial logit model of choice. We treat individual dwelling units as the elemental alternatives among which choice is made. Each dwelling unit will have a list of attributes, observed and unobserved, to which the individual is responsive. We assume 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 which are similar in some attributes as quite similar overall; it is the impact of such perceptions on choice that we wish to model. Sections 3 to 6 introduce a family of probabilistic choice models, of which the joint multinomial logit model is a special case, with the property that dwelling units which are perceived as similar are aggregated. 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, 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. Section 7 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 given in this paper to the two problems above are applied in Section 8 to empirical studies of housing location by

Quigley (1976) and Lerman (1977). The results are shown to permit a reinterpretation of the empirical conclusions, and suggest ways of generalizing the empirical analysis.

2 A 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, income, and so on. The consumer will choose the alternative which maximizes his utility.

Not all attributes of alternatives of consumer characteristics 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 probabilitistic statements can be made about the expected distribution of choices; namely,

$$(1) \qquad P_{cn} = \text{Prob } [U_{cn} > U_{bm} \text{ for } bm \neq cn] \quad ,$$

where P_{cn} denotes the probability of choice cn and the right-handside probability is defined with respect to the distribution of unobserved variables. Conversely, observed distributions of choices can be used to draw inferences on the form of utility and the distribution of unobserved variables. It should be noted that observations can be inconsistent with the existence of any stochastic utility function satisfying (1), and that if a stochastic utility function exists satisfying (1), it is normally not unique [see McFadden and Richter (1970); McFadden (1975b)]. 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 $U_{cn} = V_{cn} + \varepsilon_{cn}$ of utility into a term V_{cn} which 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 <u>strict utility</u> of cn . Let $\underset{\sim}{\varepsilon}$ denote the vector $(\varepsilon_{11},\ldots,\varepsilon_{1N_1},\ldots,\varepsilon_{C1},\ldots,\varepsilon_{CN_C})$, and let $F(\underset{\sim}{\varepsilon})$ denote the cumulative distribution function of $\underset{\sim}{\varepsilon}$. Then (1) can be written

$$(2) \qquad P_{cn} = \int_{\varepsilon_{cn}=-\infty}^{+\infty} F_{cn}(\langle V_{cn} + \varepsilon_{cn} - V_{dm}\rangle)d\varepsilon_{cn} \quad ,$$

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

Although the class of models (2) was formulated starting from the theory of the rational economic consumer, it should be noted that this specification of choice probabilities is considerably more general. In particular, unobserved random variables may enter the determination of utility for each consumer, as well as between consumers; this is known in psychology as the random utility model.

3 THE MULTINOMIAL LOGIT MODEL

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

$$P_{cn} = e^{V_{cn}} \Big/ \sum_{b=1}^{C} \sum_{m=1}^{N_b} e^{V_{bm}} \quad , \tag{3}$$

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

$$\text{Prob}\ [\varepsilon_{cn} \leq \varepsilon] = \exp\ (-e^{-\varepsilon}) \quad . \tag{4}$$

This model was proposed as a theory of psychological choice behavior by Luce (1959). Its econometric analysis has been investigated by McFadden (1973, 1976) and Nerlove-Press (1973). A particular structural feature of this model, termed by Luce independence from irrelevant alternatives, 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 (see McFadden-Tye-Train (1977); McFadden (1977)), but can be shown to be implausible for choice problems where it is unreasonable to assume the ε_{cn} are statistically independent (see Debreu (1960); Domencich-McFadden (1975)).

For later analysis, it will be useful to re-write the joint choice model (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

$$(5) \qquad V_{cn} = \beta' x_{cn} + \alpha' y_c$$

where x_{cn} is a vector of observed attributes which vary with both community and dwelling (e.g., workplace accessibility), y_c is a vector of observed attributes which 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 (5). From (3) and (5), one obtains the formulae:

$$(6) \qquad P_{n|c} = e^{V_{cn}} \Big/ \sum_{m=1}^{N_c} e^{V_{cm}} = e^{\beta' x_{cn}} \Big/ \sum_{m=1}^{N_c} e^{\beta' x_{cm}} \quad ,$$

$$(7) \qquad P_c = \sum_{n=1}^{N_c} e^{V_{cn}} \Big/ \sum_{b=1}^{C} \sum_{m=1}^{N_b} e^{V_{bm}} = e^{\alpha' y_c} \left(\sum_{n=1}^{N_c} e^{\beta' x_{cn}} \right) \Big/ \sum_{b=1}^{C} e^{\alpha' y_b} \left(\sum_{m=1}^{N_b} e^{\beta' x_{mb}} \right) \quad .$$

Define an <u>inclusive value</u>

$$(8) \qquad I_c = \log \left(\sum_{n=1}^{N_c} e^{\beta' x_{cn}} \right) \quad .$$

Then, (6) and (7) can be re-written

(9) $$P_{n|c} = e^{\beta' c_{cn}} / e^{I_c} ,$$

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

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

4 THE NESTED LOGIT MODEL

An empirical generalization of the multinomial logit model in the form (9) — (10) is obtained by allowing the inclusive value I_c in (10) to have a coefficient other than one; i.e.,

(11) $$P_c = e^{\alpha' y_c + (1-\sigma) I_c} / \sum_{b=1}^{C} e^{\alpha' y_b + (1-\sigma) I_b} ,$$

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, by Domencich and McFadden (1975). Ben-Akiva (1974) suggested the correct definition (8) of inclusive value, and explored the implications of fitting the joint model or various nested models. Amemiya (1976) 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.

The next section of this paper permits us to establish conditions under which the nested logit model can be derived from a theory of stochastic utility maximization, in a manner analogous to the argument for the multinomial logit model.

5 THE GENERALIZED EXTREME VALUE MODEL

We introduce a family of choice models derived from stochastic utility maximization which 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:

THEOREM 1. *Suppose* $G(y_1,\ldots,y_J)$ *is a nonnegative, homogeneous-of-degree-one function of* $(y_1,\ldots,y_J) > 0$. *Suppose* $\lim_{y_i \to +\infty} G(y_1,\ldots,y_J) = +\infty$ *for* $i = 1,\ldots,J$. *Suppose for any distinct* $(i_1,\ldots,i_k)$ *from* $\{1,\ldots,J\}$, $\partial^k G/\partial y_{i_1} \ldots \partial y_{i_k}$ *is nonnegative if* k *is odd and non-positive if* k *is even. Then,*

$$P_i = e^{V_i} G_i(e^{V_1},\ldots,e^{V_J}) \;/\; G(e^{V_1},\ldots,e^{V_J}) \tag{12}$$

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

Proof. Consider the function

$$F(\varepsilon_1,\ldots,\varepsilon_J) = \exp\{-G(e^{-\varepsilon_1},\ldots,e^{-\varepsilon_J})\} \quad .$$

We shall first prove that this is a multivariate extreme value distribution. If $\varepsilon_i \to -\infty$, then $G \to +\infty$, implying $F \to 0$. If $(\varepsilon_1,\ldots,\varepsilon_J) \to +\infty$, then $G \to 0$, implying $F \to 1$. Define, recursively, $Q_1 = G_1$ and $Q_k = Q_{k-1}G_k - \partial Q_{k-1}/\partial y_k$. Then, Q_k is a sum of signed terms, with each term a product of cross derivatives of G of various orders. Suppose each signed term in Q_{k-1} is nonnegative. Then $Q_{k-1}G_k$ is nonnegative. Further, each term in $\partial Q_{k-1}/\partial y_k$ is non-positive, since one of the derivatives within each term has increased in order, changing from even to odd or vice versa, with a hypothesized change in sign. Hence, each term in Q_k is nonnegative. By induction, Q_k is nonnegative for $k = 1,\ldots,J$.

Differentiating F , $\partial F/\partial\varepsilon_1 = e^{-\varepsilon_1}Q_1F$. Suppose $\partial^{k-1}F/\partial\varepsilon_1,\ldots,\partial\varepsilon_{k-1} = e^{-\varepsilon_1} \ldots e^{-\varepsilon_{k-1}}Q_{k-1}F$. Then, $\partial^k F/\partial\varepsilon_1,\ldots,\partial\varepsilon_k = e^{-\varepsilon_1} \ldots e^{-\varepsilon_k}\{Q_{k-1}G_kF - F\partial Q_{k-1}/\partial y_k\} = e^{-\varepsilon_1} \ldots e^{-\varepsilon_k}Q_kF$. By induction, $\partial^J F/\partial\varepsilon_1,\ldots,\partial\varepsilon_J = e^{-\varepsilon_1} \ldots e^{-\varepsilon_J}Q_JF \geqq 0$. Hence, F is a cumulative distribution function. When $\varepsilon_J = +\infty$ for $j \neq i$, $F = \exp\{-a_ie^{-\varepsilon_i}\}$, where $a_i = G(0,\ldots,0,\underset{\substack{i\text{-th}\\ \text{place}}}{1},0,\ldots,0)$. This is the univariate extreme value distribution. Hence, F is a multivariate extreme value distribution.

Suppose a population has utilities $u_i = V_i + \varepsilon_i$, where $(\varepsilon_1,\ldots,\varepsilon_J)$ is distributed F . Then, the probability that the first alternative is selected satisfies

$$P_1 = \int_{\varepsilon=-\infty}^{+\infty} F_1(\varepsilon, V_1 - V_2 + \varepsilon, \ldots, V_1 - V_J + \varepsilon)d\varepsilon$$

$$= \int_{\varepsilon=-\infty}^{+\infty} e^{-\varepsilon} G_1(e^{-\varepsilon-V_1+V_2},\ldots,e^{-\varepsilon-V_1+V_J}) \exp\left\{-G\left(e^{-\varepsilon},e^{-\varepsilon-V_1+V_2},\ldots,e^{-\varepsilon-V_1+V_J}\right)\right\} d\varepsilon$$

$$= \int_{\varepsilon=-\infty}^{+\infty} e^{-\varepsilon} G_1(e^{V_1},e^{V_2},\ldots,e^{V_J}) \exp\left\{-e^{-\varepsilon}e^{-V_1} G\left(e^{V_1},e^{V_2},\ldots,e^{V_J}\right)\right\} d\varepsilon$$

$$= e^{V_1} G_1(e^{V_1},e^{V_2},\ldots,e^{V_J}) / G(e^{V_1},\ldots,e^{V_J}) \quad ,$$

where the third equality uses the homogeneity of degree one of G , and consequent homogeneity of degree zero of G_1 . Since this argument can be applied to any alternative, the theorem is proved. Q.E.D.

Corollary. Under the hypotheses of Theorem 1, expected maximum utility, defined by

$$(13) \qquad \overline{U} = \int_{\underset{\sim}{\varepsilon}=-\infty}^{+\infty} \max_i \, (V_i + \varepsilon_i) f(\underset{\sim}{\varepsilon}) d\underset{\sim}{\varepsilon}$$

(with f the density for F), satisfies

(14) $$\bar{U} = \log G(e^{V_1},\ldots,e^{V_J}) + \gamma ,$$

where $\gamma = .57721\ 56649\ldots$ is Euler's constant, and

(15) $$P_i = \partial\bar{U}/\partial V_i \quad .$$

Proof. The extreme value distribution $\exp(-ae^{-\varepsilon})$, with $a > 0$ a constant, has mean $\log a + \gamma$. The integral (13) can be partitioned into regions where each alternative has maximum utility, yielding

(16) $$\bar{U} = \sum_i \int_{\varepsilon_i=-\infty}^{+\infty} (V_i + \varepsilon_i) F_i(\langle V_i + \varepsilon_i - V_j \rangle) d\varepsilon_i \quad .$$

Let $a = G(e^{V_1},\ldots,e^{V_J})$. Then,

$$F_i(\langle V_i + \varepsilon_i - V_j\rangle) = \exp(-G(\langle e^{-V_i-\varepsilon_i+V_j}\rangle))G_i(\langle e^{-V_i-\varepsilon_i+V_j}\rangle)e^{-\varepsilon_i}$$

$$= \exp(-ae^{-V_i-\varepsilon_i})G_i(\langle e^{V_j}\rangle)e^{-\varepsilon_i} \quad .$$

Making the transformation $V_i + \varepsilon_i \to w$, (16) becomes

$$\bar{U} = \sum_i \int_{w=-\infty}^{+\infty} w \exp(-ae^{-w})e^{V_i}G_i(\langle e^{V_j}\rangle)e^{-w}dw$$

$$= \int_{w=-\infty}^{+\infty} wae^{-w} \exp(-ae^{-w})dw$$

$$= \log a + \gamma \quad ,$$

where the second equality uses Euler's law, $\sum_i e^{V_i}G_i(\langle e^{V_j}\rangle) = G(\langle e^{V_j}\rangle)$.

It is immediate from (12) and (14) that $P_i = \partial\bar{U}/\partial V_i$. Q.E.D.

The special case $G(y_1,\ldots,y_J) = \sum_{j=1}^{J} y_j$ yields the MNL model. An example of a more general G function satisfying the hypotheses of the theorem is

$$(17) \qquad G(y) = \sum_{m=1}^{M} a_m \left(\sum_{i\epsilon B_m} y_i^{\frac{1}{1-\sigma_m}} \right)^{1-\sigma_m} ,$$

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

For the bivariate case with a single class m , (17) reduces to

$$G(y) = \left(y_1^{\frac{1}{1-\sigma}} + y_2^{\frac{1}{1-\sigma}} \right)^{1-\sigma} .$$

The bivariate extreme value distribution based on this form has been studied by Oliveira (1958, 1961), who shows that σ is the product-moment correlation between the two variates. In the general case (17), σ_m can be interpreted as an index of the similarity of the unobserved attributes in B_m . However, the relation between the σ_m and product-moment correlations between the alternatives is more complex.

The choice probabilities for the function (17) satisfy

$$(18) \qquad P_i = \sum_{m \ni i\epsilon B_m} e^{\frac{V_i}{1-\sigma_m}} a_m \left(\sum_{j\epsilon B_m} e^{\frac{V_j}{1-\sigma_m}} \right)^{-\sigma_m} \Big/ \sum_{n=1}^{M} a_n \left(\sum_{k\epsilon B_n} e^{\frac{V_k}{1-\sigma_n}} \right)^{1-\sigma_n}$$

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

where

$$P(i|B_m) = e^{\frac{V_i}{1-\sigma_m}} \Big/ \sum_{j\epsilon B_m} e^{\frac{V_j}{1-\sigma_m}} \qquad \text{if } i \in B_m ,$$

$$0 \qquad \text{if } i \notin B_m ,$$

with $P(i|B_m)$ denoting conditional probability, and

$$P(B_m) = a_m \left\{ \sum_{j\epsilon B_m} e^{\frac{V_j}{1-\sigma_m}} \right\}^{1-\sigma_m} \Big/ \sum_{n=1}^{M} a_n \left\{ \sum_{k\epsilon B_n} e^{\frac{V_k}{1-\sigma_n}} \right\}^{1-\sigma_n} .$$

Functions of the form in (18) can also be nested to yield a wider class satisfying the theorem hypotheses. For example, the function

$$\text{(19)} \qquad G = \sum_{q=1}^{Q} a_q \left\{ \sum_{m\epsilon D_q} \left\{ \sum_{j\epsilon B_m} Y_j^{\frac{1}{1-\sigma_m}} \right\}^{\frac{1-\sigma_m}{1-\delta_q}} \right\}^{1-\delta_q} ,$$

where B_m is defined as in (17) and $D_q \subseteq \{1,\ldots,M\}$, satisfies the hypotheses provided $1 > \sigma_m \geq \delta_q \geq 0$ for $m \in D_q$.

Choice probabilities of the form (18) were apparently first derived, for the case of three alternatives and $B_1 = \{1\}$, $B_2 = \{2,3\}$, by Scott Cardell (1975). For the case of disjoint B_m , the form (18) was treated, independently, by Daly and Zachary (1976), Williams (1977), and Lerman and Ben-Akiva (1977). The demonstration by Daly and Zachary that (18) is consistent with random utility maximization is noteworthy in that it permits generalization of the GEV model and provides a powerful tool for testing the consistency of choice models. The corollary to Theorem 1 is adapted from their analysis.

Consider an example of (17),

$$G(y_1,y_2,y_3) = y_1 + \left(y_2^{\frac{1}{1-\sigma}} + y_3^{\frac{1}{1-\sigma}}\right)^{1-\sigma} \quad ,$$

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 (18),

$$(20) \qquad P(1|1,2,3) = e^{V_1} \Big/ \left\{ e^{V_1} + \left(e^{\frac{V_2}{1-\sigma}} + e^{\frac{V_3}{1-\sigma}}\right)^{1-\sigma} \right\} \quad ;$$

$$(21) \qquad P(2|1,2,3) = e^{\frac{V_2}{1-\sigma}} \left(e^{\frac{V_2}{1-\sigma}} + e^{\frac{V_3}{1-\sigma}}\right)^{-\sigma} \Big/ \left\{ e^{V_1} + \left(e^{\frac{V_2}{1-\sigma}} + e^{\frac{V_3}{1-\sigma}}\right)^{1-\sigma} \right\} \quad ;$$

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 (20) by setting $V_3 = -\infty$) has the binomial form

$$(22) \qquad P(1|1,2) = e^{V_1} \Big/ \left\{ e^{V_1} + e^{V_2} \right\} \quad .$$

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

$$\text{(23)} \qquad P(2|2,3) = e^{\frac{V_2}{1-\sigma}} \Big/ \left\{ e^{\frac{V_2}{1-\sigma}} + e^{\frac{V_3}{1-\sigma}} \right\} \quad .$$

Examining the choice probabilities (20) and (21) when all three alternatives are available, the value $\sigma = 0$ gives multinomial logit probabilities, while the limiting value $\sigma \to 1$ gives the probabilities

$$\text{(24)} \qquad P(1|1,2,3) = e^{V_1} \Big/ \left\{ e^{V_1} + \max\left\{ e^{V_2}, e^{V_3} \right\} \right\} \quad ;$$

$$\text{(25)} \qquad P(2|1,2,3) = \begin{cases} e^{V_2} \Big/ \left\{ e^{V_1} + e^{V_2} \right\} & , \text{ if } V_2 > V_3 \;; \\ \frac{1}{2} e^{V_2} \Big/ \left\{ e^{V_1} + e^{V_2} \right\} & , \text{ if } V_2 = V_3 \;; \\ 0 & , \text{ if } V_2 < V_3 \;. \end{cases}$$

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 .

6 RELATION OF THE NESTED LOGIT AND GENERALIZED EXTREME VALUE MODELS

The choice probabilities (18) can be specialized to the nested logit model given by (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 (9) and (11) from (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, (18) becomes

$$(26) \qquad P_{cn} = e^{\frac{V_{cn}}{1-\sigma}} \left(\sum_{m=1}^{N_c} e^{\frac{V_{cm}}{1-\sigma}} \right)^{-\sigma} \Big/ \left\{ \sum_{b=1}^{C} \left(\sum_{m=1}^{N_b} e^{\frac{V_{bm}}{1-\sigma}} \right)^{1-\sigma} \right\} ,$$

implying

$$(27)\qquad P_c = \sum_{n=1}^{N_c} P_{cn} = \left(\sum_{m=1}^{N_c} e^{\frac{V_{cm}}{1-\sigma}}\right)^{1-\sigma} \Bigg/ \left\{\sum_{b=1}^{C}\left(\sum_{m=1}^{N_b} e^{\frac{V_{bm}}{1-\sigma}}\right)^{1-\sigma}\right\} ,$$

and

$$(28)\qquad P_{n|c} = P_{cn}/P_c = e^{\frac{V_{cn}}{1-\sigma}} \Big/ \sum_{m=1}^{N_c} e^{\frac{V_{cm}}{1-\sigma}} .$$

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

$$(29)\qquad P_c = e^{\alpha' y_c + (1-\sigma) I_c} \Big/ \sum_{b=1}^{C} e^{\alpha' y_b + (1-\sigma) I_b} ;$$

$$(30)\qquad P_{n|c} = e^{\frac{\beta' x_{cn}}{1-\sigma}} \Big/ \sum_{m=1}^{N_c} e^{\frac{\beta' x_{cm}}{1-\sigma}} = e^{\frac{\beta' x_{cn}}{1-\sigma}} \Big/ e^{I_c} ;$$

$$(31)\qquad I_c = \log \sum_{m=1}^{N_c} e^{\frac{\beta' x_{cm}}{1-\sigma}} .$$

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.

Applying the argument above to the choice probabilities corresponding to (19), with a triple index of alternatives and each level of nesting corresponding to one of these indices, leads to a nested logit model with two levels of conditional probabilities and inclusive values This argument for three-level decision trees can be extended to trees of any depth. The condition for (19) to satisfy the hypotheses of Theorem 1 implies that <u>a sufficient condition for a nested logit model</u>

to be consistent with stochastic utility maximization is that the coefficient of each inclusive value lie in the unit interval.

Application of the Daly-Zachary method shows that these restrictions on the coefficients of inclusive values are also necessary for consistency with stochastic utility maximization; see McFadden (1977b).

7 LIMITING THE NUMBER OF ALTERNATIVES CONSIDERED

Consider application of the joint multinomial logit model (3) to the demand for housing, with alternatives indexed by community and by dwelling within the community. Ideally, the functional form of the model (3) 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, which are likely to have similar unobserved attributes.

Consider first the problem where enumeration of all alternatives is impractical, but 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,

$$(32) \qquad i \in D \subseteq C \Longrightarrow P(i|C,z,\theta) = P(i|D,z,\theta) \sum_{j \in D} P(j|C,z,\theta) \quad ,$$

which characterizes the multinomial logit model.

Now suppose for each case, 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 , independent of 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 several examples of distributions of type (c):

(c-1) Suppose D is comprised of i plus a sample of alternative from the set $C \setminus \{i\}$, obtained by considering each element of this set independently, and including it with probability p . Then, the probability of D will depend solely on the number of elements $K = \#(D)$ it contains, and is given by the binomial formula

$$(33)\qquad \pi(D|i,z) = \begin{cases} p^{K-1}(1-p)^{J-K} & \text{if } i \in D \text{ and } K = \#(D) , \\ 0 & \text{if } i \notin D , \end{cases}$$

where J is the number of alternatives in C . For example, the probability that D will be a specified two-alternative set containing i as one alternative is $p(1 - p)^{J-2}$.

(c-2) 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

$$(34)\qquad \pi(D|i,z) = \begin{cases} \frac{1}{J-1} & \text{if } D = \{i,j\} \text{ and } j \neq i , \\ 0 & \text{otherwise} . \end{cases}$$

(c-3) Suppose C has four elements, and

$$(35)\qquad \pi(\{1,4\}|4) = \pi(\{1,4\}|1) = \pi(\{2,3\}|2) = \pi(\{2,3\}|3) = 1 ,$$
$$\text{and } \pi(D|i) = 0 \text{ otherwise} .$$

(c-4) 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,

$$(36)\qquad \pi(D|i,z) = \begin{cases} J_n \Big/ \prod_{m=1}^{M} J_m & \text{if } i \in D , M = \#(D) , \text{ and} \\ & D \cap C_m \neq \emptyset \text{ for } m = 1,\ldots,M , \\ 0 & \text{otherwise} . \end{cases}$$

The π distributions of the type (a), (b), and (c-1) to (c-4) 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:

<u>Positive conditioning property</u>: <u>If</u> $j \in D \subseteq C$ <u>and</u> $\pi(D|i,z) > 0$, <u>then</u> $\pi(D|j,z) > 0$.

The π distributions (a), (b), and (c-1) to (c-3), but not (c-4), satisfy a stronger condition:

<u>Uniform conditioning property</u>: <u>If</u> $i,j \in D \subseteq C$, <u>then</u> $\pi(D|i,z) = \pi(D|j,z)$.

A distribution with the uniform conditioning property can be written $\pi(D|i,z) = \phi(D,z)X_D(i)$, where $X_D(i)$ equals one for $i \in D$, and zero otherwise.

Consider a sample $n = 1,\ldots,N$, with the alternative chosen on case n denoted by 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

$$(37) \qquad P(i|C,z,\theta) = e^{V_k(z,\theta)} \Big/ \sum_{j\in C} e^{V_j(z,\theta)} ,$$

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

THEOREM 2. If $\pi(D|i,z)$ satisfies the positive conditioning property and the choice model is multinomial logit, then maximization of the modified likelihood function

$$(38) \qquad L_N = \frac{1}{N}\sum_{n=1}^{N} \log\left\{e^{V_{i_n}(z_n,\theta)+\log \pi(D_n|i_n,z_n)} \Big/ \sum_{j\in C} e^{V_j(z_n,\theta)+\log \pi(D_n|j,z_n)}\right\}$$

yields, under normal regularity conditions, consistent estimates of θ^* . When $\pi(D|i,z)$ satisfies the uniform conditioning property, then (38) reduces to the standard likelihood function,

$$(39) \qquad L_N = \frac{1}{N}\sum_{n=1}^{N} \log\left\{e^{V_{i_n}(z,\theta)} \Big/ \sum_{j\in C} e^{V_j(z_n,\theta)}\right\} .$$

Proof. Consider the likelihood function (38), and consider its probability limit $\text{plim } L_N = L$, where

$$(40) \qquad L = \int_z \left\{\sum_{i\in C}\sum_{D\subseteq C} P(i|C,z,\theta^*)\pi(D|i,z) \log \left[\frac{e^{V_i(z,\theta)+\log \pi(D|i,z)}}{\sum_{j\in C} e^{V_j(z,\theta)+\log \pi(D|j,z)}}\right]\right\} p(z)dz$$

with $p(z)$ the frequency distribution of z . Then,

$$(41) \qquad L = \int_z \left\{\sum_{D\subseteq C}\left\{\frac{\sum_{j\in C} e^{V_j(z,\theta^*)}\pi(D|j,z)}{\sum_{j\in C} e^{V_j(z,\theta^*)}}\right\} \sum_{i\in C} \frac{e^{V_i(z,\theta^*)}\pi(D|i,z)}{\sum_{j\in C} e^{V_j(z,\theta^*)}\pi(D|j,z)}\right.$$

(continued)

$$\cdot \log \left[\frac{e^{V_i(z,\theta)} \pi(D|i,z)}{\sum_{j\in C} e^{V_j(z,\theta^*)} \pi(D|j,z)} \right] \Bigg\} p(z)dz \quad .$$

The parameters θ enter this expression in a term of the form

(42) $$\sum_{i\in C} \phi_i(\theta^*) \log \phi_i(\theta) \quad \text{with} \quad \sum_{i\in C} \phi_i(\theta) = 1 \quad .$$

One can show that (42) has a maximum at $\theta = \theta^*$. Then, L has a maximum at $\theta = \theta^*$. Under normal regularity conditions, this maximum is unique, and it can be shown (Manski-McFadden (1977)) that the maxima of L_N converge in probability to the maximum of L . This establishes that maximization of (38) yields consistent estimators.

If the uniform conditioning property holds, then the terms $\log \pi(D|i,z)$ in (38) cancel out, and (39) results. Q.E.D.

In a study of choice of community and dwelling type using data for the San Francisco peninsula, Friedman (1975) used the mechanism (c-4) to limit the number of alternatives in the analysis, partitioning the choice set by community and major dwelling type. Friedman's strict utility function has the form $V_{ctd}(z,\theta) = \alpha_{ct} + z_{itd}\beta$, where c indexes community, t indexes dwelling type, d indexes dwelling within class ct , z_{ctd} is a sub-vector of z specifying attributes of alternative ctd , α_{ct} is a class-ct-specific parameter, β is a parameter vector, and θ denotes the collection of parameters β and α_{ct} . The modified likelihood function (38), using (36) and cancelling terms, is

(43)

$$L_N = \frac{1}{N} \sum_{n=1}^{N} \log \left\{ e^{\alpha_{c_n t_n} + z_{c_n t_n d_n} \beta + \log J_{c_n t_n}} \Big/ \sum_{c',t',d'} e^{\alpha_{c't'} + z_{c't'd'} \beta + \log J_{c't'}} \right\} ,$$

where J_{ct} is the number of dwellings in class ct. However, Friedman estimates the parameters by maximizing the likelihood function (39), which differs from (43) by omission of the terms $\log J_{ct}$. Comparison of these functions shows that Friedman's procedure provides consistent estimates of the parameter vector β, and that his estimates of the class-specific parameters converge in probability to the sum of the true class-specific parameters, α^*_{ct}, and the log of the number of dwellings available in the ct class.

In conclusion, analysis of housing location can be carried out with a limited number of alternatives, facilitating data collection and processing, provided the choice process is described by the multinomial logit model. If a mechanism such as (c-4) is used to select alternatives, the likelihood function should be modified to the form (38) 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.

8 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 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 which 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 the dwellings in a community c have common unobserved community attributes, introducing a dependence which 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, from (29) — (31),

$$(44) \qquad P_c = e^{\alpha' y_c + (1-\sigma) I_c} \Big/ \sum_{b=1}^{C} e^{\alpha' y_b + (1-\sigma) I_b} ,$$

$$(45) \qquad I_c = \log \sum_{m=1}^{N_c} \frac{\beta' x_{cm}}{e^{1-\sigma}} .$$

As was shown in Section 6, this model is derivable from the generalized extreme value model of stochastic utility maximization, with σ a measure of the degree to which dwellings within a class c are perceived as similar. When $\sigma = 0$, (44) reduces to the multinomial logit model, and in the limit when $\sigma = 1$, (44) reduces to

$$(46) \qquad P_c = e^{\alpha' y_c + \max_n \beta' x_{cn}} \Big/ \sum_{b=1}^{C} e^{\alpha' y_b + \max_n \beta' x_{bn}} \quad .$$

An analysis of housing demand by Quigley (1976) using Pittsburgh data employs a model of the form of (46). (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, (46) can be treated as an ordinary multinomial logit model.

Two features of the Quigley model require note. First, Quigley defines $\max_n \beta' x_{cn}$ to be the mean of the largest five percent of the $\beta' x_{cn}$'s , rather than the largest value of $\beta' x_{cn}$. Second, Quigley includes the number of units N_c in each class as an explanatory variable, justifying it as a "proxy for the information available to consumers about the location and prices of alternative housing types." As the discussion below shall make clear, a possible alternative explanation of the significant positive coefficients of N_c in Quigley's model is that the true model is of the form (44) with $\sigma < 1$.

In an analysis of neighborhood choice using Washington, DC data, Lerman (1977) estimates a model of the form

$$P_c = e^{\alpha' y_c + X_c^* + (1-\sigma)\log N_c} \Big/ \sum_{b=1}^{C} e^{\alpha' y_b + X_b^* + (1-\sigma)\log N_b} \quad , \tag{47}$$

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 (46) as a second extreme case, with $X_c^* = \max_n \beta' x_{cn}$, 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 $1 - \sigma = .492$, 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 (44) and (45), the inclusive value can be re-written

$$I_c = \frac{X_c^*}{1-\sigma} + \log N_c + \log \frac{1}{N_c} \sum_{m=1}^{N_c} e^{\frac{\beta' x_{cm} - X_c^*}{1-\sigma}} \quad . \tag{48}$$

If a tract c is homogeneous in terms of observed variables, so that $\beta' x_{cm} = X_c^*$, then the last term in (48) vanishes, and the choice probability for the nested logit model (44) is exactly the Lerman model (47). 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 $X_c^* = \max_n \beta' x_{cn}$ or X_c^* given by the average of the "best" five percent of the disaggregate alternatives. Since Quigley uses N_c rather than $\log N_c$ as an explanatory variable, his model does not provide a direct estimate of σ .

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

(49) $$\frac{1}{N_c} \sum_{m=1}^{N_c} e^{\frac{\beta' x_{cm} - X_c^*}{1-\sigma}} \geqq 1 ,$$

and hence $I_c \geqq \frac{X_c^*}{1-\sigma} + \log N_c$, with the difference of the two sides of the inequality depending on the variance of $\beta' x_{cm}$. One limiting case of (48) that is of interest occurs when the number of dwellings within a tract is large, and the x_{cm} behave as if they are independently identically normally distributed with mean x_c^* . Let ω_c^2 denote the variance of $\beta' x_{cm}$. Then,

(50) $$E\, e^{\frac{\beta' x_{cm} - \beta' x_c^*}{1-\sigma}} = \exp\left\{\frac{1}{2}\left(\frac{\omega_c}{1-\sigma}\right)^2\right\} ,$$

and

$$(51)\qquad \frac{1}{N_c}\sum_{m=1}^{N_c} e^{\frac{\beta' x_{cm} - \beta' x_c^*}{1-\sigma}} \xrightarrow[\text{a.s.}]{} \exp\left\{\frac{1}{2}\left(\frac{\omega_c}{1-\sigma}\right)^2\right\} .$$

Hence,

$$(52)\qquad I_c \xrightarrow[\text{a.s.}]{} \frac{\beta' x_c^*}{1-\sigma} + \log N_c + \frac{1}{2}\left(\frac{\omega_c}{1-\sigma}\right)^2 .$$

The higher the variance ω_c^2 , the higher the inclusive value I_c for trace c . If $N_c = r_c N$ (i.e., the numbers of alternatives in tracts grow proportionately), then

$$(53)\qquad P_c \xrightarrow[\text{a.s.}]{} \frac{e^{\alpha' y_c + \beta' x_c^* + (1-\sigma)\log r_c + \frac{1}{2}\omega_c^2/(1-\sigma)}}{\sum_{b=1}^{C} e^{\alpha' y_b + \beta' x_b + (1-\sigma)\log r_b + \frac{1}{2}\omega_b^2/(1-\sigma)}} .$$

When the disaggregate data x_{cn} are not observed, but their distribution can be approximated or estimated, and ω_c is known, then model (53) can be used with standard multinomial logit estimation programs to provide estimates of σ and β . If r_c is unobserved, then it can be estimated when ω_c is known, although when y_c contains a tract-specific dummy variable, the tract-specific coefficient and r_c 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_c is not known, but is known to have the structure

$\omega_c^2 = \beta'\Omega_c\beta$, and the variables x_{cm} are multivariate normal with covariance matrix Ω_c , 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 $\beta'\Omega_c\beta$ as independent parameters and ignoring constraints.

9 CONCLUSION

This paper has considered the problem of modelling 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 Friedman (1975), Quigley (1976), and Lerman (1977), motivate the analysis and illustrate its applicability.

REFERENCES

Amemiya, T. (1976), "Specification and Estimation of a Multinomial Logit Model," Technical Report No. 211, Institute of Mathematical Studies in the Social Sciences, Stanford University, Stanford, CA

Ben-Akiva, M. and S. Lerman (1977), "Disaggregate Travel and Mobility Choice Models and Measures of Accessibility," Third International Conference on Behavioural Travel Modelling, Tanenda, Australia, 1977

Daly, A. and S. Zachary (1976), "Improved Multiple Choice Models," Mimeo.

Debreu, G. (1960), "Review of R. Luce, Individual Choice Behavior," American Economic Review, Vol. 50, pp. 186-188

Domencich, T. and D. McFadden (1975), Urban Travel Demand: A Behavioral Analysis, North-Holland, Amsterdam

Friedman, J. (1975), "Housing Location and the Supply of Local Public Services," Ph.D. dissertation, Department of Economics, University of California, Berkeley, CA

Gustafson, J.R., B. Harsman, and F. Snickars (1977), "Housing Models and Consumer Preferences," Mimeo.

Gustafson, J.R., B. Harsman, and F. Snickars (1977), "Medium Term Forecasting Models for the Housing Market," presented at the European Regional Science Association in Krakow, Poland, 1977

Harsman, B. and F. Snickars (1976), "Disaggregated Demand Models: Some Theoretical Approaches," Proceedings of the Fourteenth European Congress of the Regional Science Association

Lerman, S.R. (1977), "Location, Housing, Automobile Ownership, and Mode to Work: A Joint Choice Model," Transportation Research Board Record, No. 610

Luce, R.D. (1959), Individual Choice Behavior, Wiley, New York

Manski, C. and D. McFadden (1977), "Alternative Estimators and Sample Designs for Discrete Choice Analysis," Department of Economics, University of California, Berkeley, CA

McFadden, D. (1977a), "On the Use of Probabilistic Choice Models in Economics," presented at the MSSB/NBER Conference on Cognition, Choice, and Economic Behavior, National Bureau of Economic Research-West, Stanford, CA, October 1977

McFadden, D. (1977b), "Quantitative Methods for Analyzing Travel Behaviour of Individuals: Some Recent Developments," Cowles Foundation Discussion Paper No. 474, Yale University. Presented at the Third International Conference on Behavioural Travel Modelling, Tanenda, Australia, 1977

McFadden, D., W. Tye, and K. Train (1976), "Diagnostic Tests for the Independence from Irrelevant Alternatives Property of the Multinomial Logit Model," Working Paper No. 7616, Urban Travel Demand Forecasting Project, Institute of Transportation Studies, University of California, Berkeley, CA

McFadden, D. (1976), "Quantal Choice Analysis: A Survey," Annals of Economic and Social Measurement, 5/4

McFadden, D. (1975a), "Economic Applications of Psychological Choice Models," Working Paper No. 7519, Urban Travel Demand Forecasting Project, Institute of Transportation Studies, University of California, Berkeley, CA

McFadden, D. (1975b), "Tchebycheff Bounds for the Space of Agent Characteristics," Journal of Mathematical Economics, Vol. 2, pp. 225-242

McFadden, D. (1973), "Conditional Logit Analysis of Qualitative Choice Behavior," in P. Zarembka, ed., Frontiers in Econometrics, Academic Press, New York

McFadden, D. and M.K. Richter (1970), "Revealed Stochastic Preference," Department of Economics, Univeristy of California, Berkeley, CA

Nerlove, M. and J. Press (1973), "Univariate and Multivariate Log-Linear and Logistic Models," RAND Report No. R-1306-EDA/NIH

Oliveira, J.T. de (1961), "La Representation des Distributions Extremales Bivariees," Bulletin of the International Statistical Institute, Vol. 33, 477-480

Oliveira, J.T. de (1958), "Extremal Distributions," Revista de Faculdada du Ciencia, Lisboa, Serie A, Vol. 7, pp. 215-227

Pollakowski, H. (1974), "The Effects of Local Public Service on Residential Location Decision: An Empirical Study of the San Francisco Bay Area, Ph.D. dissertation, Department of Economics, University of California, Berkeley, CA

Quigley, J.M. (1976), "Housing Demand in the Short-Run: An Analysis of Polytomous Choice," Explorations in Economic Research, Vol. 3, Vol.1 (Winter), pp. 76-102

Williams, H.C.L. (1977), "On the Formation of Travel Demand Models and Economic Evaluation Measures of User Benefit," Environment and Planning, Vol. A.9, p. 285-344