Daniel McFadden
Massachusetts Institute of Technology

Econometric Models for Probabilistic Choice among Products*

This paper reviews several recent developments in econometric demand analysis which may be of interest in market research. Econometric models of probabilistic choice, suitable for forecasting choice among existing or new brands, or switching between brands, are surveyed. These models incorporate attribute descriptions of commodities, making them statistical counterparts of the Court-Griliches-Lancaster theory of consumer behavior. Particular attention is given to models which yield tree structures of similarities between alternatives. Also reviewed are methods for estimating econometric models of probabilistic choice from "point-of-sale" sample surveys.

An Econometrician's View of Marketing

I understand the discipline of marketing exists to answer questions such as: "Will housepersons buy more Brand A soap if its perfume content is increased?" Traditional econometric demand analysis provides no answer. Its attention has been concentrated on consumption levels of broad commodity classes (e.g., housing services), examined using aggregate market data, with demand models constructed on the twin pillars of economic rationality and consumer sovereignty. The market researcher has understandably looked elsewhere—to psychology and survey research—for answers to his questions.

Realities have forced econometric demand analysts to broaden their perspective. Public intervention in the supply of some commodities, notably in the areas of transportation, energy, and communications, have required economists to recognize the marketing considerations implicit in issues of policy. (The decision of whether to build and how to design a public

* Prepared for presentation at the Conference on Interfaces between Marketing and Economics, Graduate School of Management, University of Rochester, April 7, 1978. Research was supported in part by the National Science Foundation through grant SOC75-22657 to the University of California, Berkeley. Portions of this paper were written while the author was an Irving Fisher Visiting Professor of Economics at the Cowles Foundation for Research in Economics, Yale University.

(*Journal of Business*, 1980, vol. 53, no. 3, pt. 2)

0021-9398/80/5332-0003$01.50

transit system is not different in kind from the decision of whether to market and how to determine the perfume content of Brand A soap.) On the other hand, economists have begun to use, and even commission, sample surveys and to confront the problems of developing models appropriate for the analysis of survey data. There are four key elements in these economic analyses of marketing problems.

1. *The objective of market research is assumed to be the forecasting of behavior of consumers in economic markets.* Then, the analysis should be focused toward this objective, with questions of attitudes, intentions, perceptions, and psychological measurement entertained *only* to the extent that they can be linked to market behavior and can improve forecasting ability.

2. *The theory of the economically rational utility-maximizing consumer, interpreted broadly to admit the effects of perception, state of mind, and imperfect discrimination, provides a plausible, logically unified foundation for the development of models of various aspects of market behavior.*

3. *The core of a model of market behavior will be an equation, consistent with the theory of the economic consumer, which specifies the probabilities of choices* (e.g., brand, frequency, or volume of purchases) *as a function of the objective market environment of the consumer,* including measured attributes and prices of alternatives, socioeconomic characteristics, and past experience. If this function depends on nonmarket variables, such as perceptions and attitudes, then the system should be completed with equations relating the evolution of attitudes and perceptions to the market environment. The basic tool for forecasting will be a "reduced-form" equation relating behavior to market environment, with intervening nonmarket variables eliminated.

4. *The design of sample surveys, construction of questionnaires, specification of models of market behavior, and statistical methods for analysis should be integrated.* In particular, survey design and statistical procedure should be jointly optimized to yield consistent, maximally efficient forecasts.

Theories and Models of Probabilistic-Choice Behavior

I start from the axiom that consumers are "rational" in the sense that they make choices which maximize their perceived utility, subject to economic constraints on expenditures. To accommodate the demonstrated inability of individuals to discriminate perfectly, or of the analyst to measure exactly the environment of the choice, I follow Thurstone (1927) in assuming that utility is a random function. Sensible assumptions on the variables on which the utility of an alternative can

depend and on the probabilistic structure of the utility function will permit the construction of plausible, nonvacuous models of behavior. Although this structure is apparently static, it can in principle accommodate learning behavior and "behavior modification" via specification of the dependence of utility on experience, information, and perception. The following paragraphs summarize the development in McFadden (1973).

I shall concentrate on choice among a finite set of alternatives (e.g., brand choice), although the theory is more general. Suppose alternatives $j = 1, \ldots, J$ are offered, with alternative j having a vector of measured attributes z_j, and a (random) utility $u(z_j)$. Then, the *choice probability* that i will be chosen satisfies

$$P(i \mid z) = P[u \in E^J \mid u(z_i) \geq u(z_j) \quad \text{for} \quad j \neq i], \tag{1}$$

where we assume the distribution of u is such that the probability of ties is zero, and $z = (z_1, \ldots, z_J)$.

The random utility function $u(z_j)$ associated with a discrete alternative should be interpreted as the maximum utility attainable by the consumer, given his budget constraint and a fixed alternative j. Then $u(z_j)$ is a function of income and prices, including the price of alternative j. The sources of randomness in the utility function are unobserved variations in tastes and in the attributes of alternatives, and errors of perception and optimization by the consumer.

If underlying preferences are defined in terms of psychological needs, with commodities having attributes which meet these needs, then z_j will include these attributes. In particular, if a consumer preference model of the Court-Griliches-Lancaster type is applied to choice among discrete, mutually exclusive alternatives, then z_j will include the attributes of the portfolio of purchases made when alternative j is specified. Further, if "household production" of the type studied by Becker enters the consumer's optimization process, then $u(z_j)$ is interpreted as maximum utility obtainable with optimal household production, given the budget constraint and the fixed alternative j.

The models of econometric choice outlined above represent a statistical implementation of economic consumer theory for problems of discrete choice. They are consistent with a variety of special preference structures, including those implied by Lancasterian or Beckerian views of the consumer, for appropriate specifications of the random utilities of the discrete alternatives.

The choice probabilities will provide the key mapping from the environment of choice to market behavior. For econometric purposes, the measured variables influencing utility will be specified, as will a finite-dimensional parametric family of probability distributions for the random utility function.

The Luce Model

One historically and empirically important probabilistic choice model, due to Luce (1959), is

$$P(i \mid z,\theta) = v(z_i,\theta) / \sum_{j=1}^{J} v(z_j,\theta), \tag{2}$$

where $v(z_i,\theta)$ is a scale value for an alternative with measured attributes z_i and is assumed known up to a finite parameter vector θ. In its econometric implementation, with $\log v(z_i,\theta)$ linear-in-parameters, this model is also called the *multinomial logit* (MNL) model. This model can be derived from the random-utility model by assuming the utilities of the various alternatives to be independently distributed, with the extreme value distribution

$$P[u(z_i) \leqq u_i] = \exp\,[-v(z_i,\theta)e^{-u_i}]. \tag{3}$$

The Luce model has been employed for market analysis in economics with reasonable success, notably in transportation planning where it has been used to forecast market penetration of new travel modes (see McFadden [1974; McFadden and Associates 1977]; for a survey of other empirical applications, see McFadden [1976]). Some marketing applications are indicated in Silk and Urban (1978) and Punj and Staelin (1978). The model contains the structural restriction, *independence from irrelevant alternatives* (IIA), that the relative odds of two alternatives are independent of the attributes, or even the presence, of a third alternative. A classic example, with strong contrasts in similarities of alternatives, shows that this restriction is sometimes implausible. Suppose z_1, z_2, and z_3 are the attributes of a trip by red bus, blue bus, and auto, respectively. Suppose consumers treat the two buses as equivalent and are indifferent between auto and bus. Then, one expects $P(1 \mid z_1,z_2) = P(1 \mid z_1,z_3) = P(2 \mid z_2,z_3) = 1/2$ and $P(1 \mid z_1,z_2,z_3) = P(2 \mid z_1,z_2,z_3) = 1/4$. The relative odds of 1 and 3 depend on the presence of alternative 2, and these probabilities are inconsistent with the Luce model. The example suggests that the Luce model will be unsuitable for marketing applications where the patterns of perceived similarities of brands have a significant influence on market shares.

Thurstone (Multinomial-Probit) Model

The structural restrictions of the Luce model have led psychologists to develop alternatives without the IIA property. One conceptually appealing alternative is the *multinomial-probit* model, obtained from the random-utility model by assuming the utilities of alternatives to have a multivariate normal distribution. This model was first proposed by Thurstone (1927) and has been applied to psychological-choice data by

Bock and Jones (1968). Applications to market-choice problems have been inhibited by computation cost. However, in a series of economic applications, approximations have been introduced which reduce the computational barrier (see Hausman and Wise 1978; Daganzo, Bouthelier, and Shiffi 1977; and Lerman and Manski 1980).

Elimination-by-Aspects Model

Another model permitting a very general pattern of similarities is the elimination-by-aspects model of Tversky (1972). Conceptually, each alternative can be viewed as owning a set of aspects or features (e.g., for soap the relevant aspects may be color, fragrance, price category, size, etc.). The consumer is viewed as selecting an aspect at random, eliminating available alternatives which fail to own this aspect, and repeating this process until a single alternative remains. An example illustrates the structure of this model in a special preference-tree case. Suppose a choice set with three alternatives $\{z_1,z_2,z_3\}$ exists. Suppose the alternatives can be arranged in a treelike structure, as in figure 1, with common segments representing common aspects of alternatives and the length v_i of each segment a measure of the abundance (or desirability) of its associated aspects. The choice probabilities are formed by sampling randomly from aspects, discarding branches without the sampled aspect, and repeating the process until a single alternative remains. Thus, in figure 1, an aspect sampled from v_1, v_2, or v_3 determines a single alternative in one step. An aspect sampled from v_4 eliminates z_3, with additional step(s) required to select an aspect in either v_1 or v_2. The choice probabilities then satisfy

$$\begin{aligned} P(1 \mid z_1,z_2) &= v_1/(v_1 + v_2) \\ P(1 \mid z_1,z_3) &= (v_1 + v_4)/(v_1 + v_3 + v_4) \\ P(2 \mid z_2,z_3) &= (v_2 + v_4)/(v_2 + v_3 + v_4) \\ P(3 \mid z_1z_2,z_3) &= v_3/(v_1 + v_2 + v_3 + v_4) \end{aligned} \tag{4}$$

$$P(1 \mid z_1,z_2,z_3) = \frac{v_1}{v_1 + v_2 + v_3 + v_4} + \frac{v_4}{v_1 + v_2 + v_3 + v_4} \frac{v_1}{v_1 + v_2}.$$

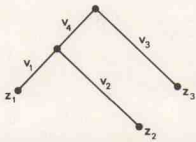

FIG. 1.—The elimination-by-aspects model for a preference tree

Note that when $v_4 = 0$ this model reduces to the Luce model. The model is readily extended to more complex trees and to general networks. Tversky shows that this model is consistent with random preference maximization. In a psychological laboratory setting, the v_i are treated as parameters. Application of the model to market choice would require that they be made parametric functions of the measured attributes of the alternatives. Since the general network case requires $2^J - 2$ terms v_i in a choice set with J alternatives, the model may become computationally infeasible for large choice sets. This problem is discussed further in McFadden (1980).

Generalized Extreme-Value Model

A third model due to me (McFadden and Associates 1978) is a direct generalization of the Luce (MNL) model to permit patterns of "non-independence" of alternatives. A random-utility model in which the utilities of the alternatives have *independent* extreme value distributions yields the Luce model. Considering nonindependent extreme value distributions leads to the *generalized extreme value* (GEV) models described below. The practical interest in these models lies in their form when specialized to preference trees such as the one in figure 2. For the GEV model, transition probabilities from one level of the tree to the next have multinomial-logit forms, with the scale value

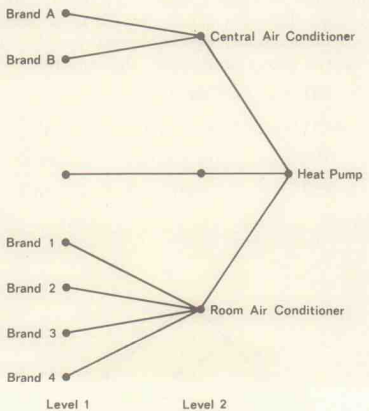

FIG. 2.—Example of a preference tree for air conditioners

associated with each branch obtained from the fitted probabilities for transitions further down the tree. Then, the GEV model can be estimated recursively by fitting a sequence of multinomial-logit models. The following result characterizes the GEV family.

Theorem

Suppose $G(y_1, \ldots, y_J)$ is a nonnegative, homogeneous-of-degree 1 function of $(y_1, \ldots, y_J) \geqq 0$ which is positive whenever any argument is positive. Suppose for any distinct $(i_1, \ldots, i_k)$ from $\{1, \ldots, J\}$, $\partial^k G/\partial y_{i_1}, \ldots, y_{i_k}$ is nonnegative if k is odd and nonpositive if k is even. Then

$$P_i = \partial \log G(v_1, \ldots, v_J)/\partial \log v_i, \tag{5}$$

where v_i is a scale value for alternative i, defines a choice model which is consistent with utility maximization.

The theorem is proved by showing that $\exp[-G(v_1 e^{-u_1}, \ldots, v_J e^{-u_J})]$ is a multivariate extreme value distribution and then obtaining (5) from (1) by construction. The special case $G(y_1, \ldots, y_J) = \sum_{j=1}^{J} y_j$ yields the MNL model. If G is symmetric in its arguments, then the choice probabilities satisfy the property of "simple scalability" (Tversky 1972), which has undesirable structural restrictions similar to the IIA property. However, in general G may depend on the "clustering" of alternatives in attribute space. An example of a more general G function satisfying the hypotheses of the theorem is

$$G(y) = \sum_{m=1}^{M} a_m H_m(y) \tag{6}$$

where

$$H_m(y) = \left(\sum_{k \epsilon B_m} y_k^{1/(1-\sigma_m)} \right)^{1-\sigma_m}$$

$$B_m \subseteq \{1, \ldots, J\}, \quad \bigcup_{m=1}^{M} B_m = \{1, \ldots, J\},$$

$$a_m > 0, \quad \text{and} \quad 0 \leqq \sigma_m < 1.$$

The parameter σ_m is an index of the similarity of the unobserved attributes of alternatives in B_m. The choice probabilities for this function satisfy

$$P_i = \sum_{m=1}^{M} P(i \mid B_m) P(B_m), \tag{7}$$

where $P(i \mid B_m)$ is the conditional probability that alternative i is chosen, given the event B_m. Then,

$$P(i \mid B_m) = v_i^{1/(1-\sigma_m)} \Big/ \sum_{j \in B_m} v_j^{1/(1-\sigma_m)} \qquad \text{if } i \in B_m$$
$$= 0 \qquad \text{if } i \notin B_m \tag{8}$$

and $P(B_m)$ is the probability of the event B_m, with

$$P(B_m) = a_m H_m(v) \Big/ \sum_{n=1}^{M} a_n H_n(v). \tag{9}$$

Functions of the form in (6) can also be nested to yield a wider class satisfying the theorem hypotheses. For example, the function

$$G(y) = \sum_{q=1}^{Q} a_q \left(\sum_{m \in D_q} H_m (y)^{1/(1-\delta_q)} \right)^{1-\delta_q} \tag{10}$$

where $\cup B_m = \{1, \ldots, J\}$, satisfies the hypotheses provided $1 > \sigma_m \geqq \delta_q \geqq 0$ for $m \in D_q$. The choice probabilities for (10) and analogous functions can be written as sums of products of conditional and marginal probabilities, in a manner generalizing (7), with each probability element having a MNL form and the denominator in each element equaling a representative term in the succeeding element.

Choice probabilities of the form (7) were apparently first derived for the case of three alternatives and $B_1 = \{1\}$, $B_2 = \{2, 3\}$ by Cardell (1977). For the case of disjoint B_m, the form (7) has been discovered, independently, by Daly and Zachary (1976), Ben-Akiva and Lerman (1977), and Williams (1977).

A particular advantage of the GEV model is that it can be specialized to forms with a "treelike" interpretation analogous to figure 2, but with the property that at each level of the tree choice can be interpreted as conforming to a MNL model. For example,

$$G(y_1, y_2, y_3) = (y_1^{1/\rho} + y_2^{1/\rho}) + y_3 \text{ with } \rho = 1 - \sigma$$

yields the choice probabilities

$$\begin{aligned} P(1 \mid z_1, z_2) &= v_1^{1/\rho} / (v_1^{1/\rho} + v_2^{1/\rho}) \\ P(1 \mid z_1, z_3) &= v_1 / (v_1 + v_3) \\ P(3 \mid z_1, z_2, z_3) &= v_3 / G(v_1, v_2, v_3) \\ P(1 \mid z_1, z_2, z_3) &= P(1 \mid z_1, z_2)\, (v_1^{1/\rho} + v_2^{1/\rho})^{\rho} / G(v_1, v_2, v_3). \end{aligned} \tag{11}$$

To compare this model with the elimination-by-aspects model (4), I have calculated the respective v_i parameters and σ so that binary-choice probabilities in the two models are the same and then compared their trinomial-choice probabilities. I found that these probabilities never differ by more than .015, or 3%. Thus, at least for three alternatives, these models appear to be indistinguishable for empirical purposes.

To see how (11) can be interpreted as a nest of MNL, or Luce, models, suppose that the scale values v_i can be specialized to a form log-linear-in-parameters, $\log v_i = \theta' z_i$. Then, the choice probability for alternative 1, conditioned on 1 or 2 being chosen, equals the binary-choice probability for 1 over 2,

$$P(1 \mid z_1, z_2) = e^{\theta' z_1/\rho}/(e^{\theta' z_1/\rho} + e^{\theta' z_2/\rho}). \tag{12}$$

Consideration of conditional-choice data then permits estimation of the parameter vector $\theta/(1 - \sigma)$ and calculation of the inclusive value, or composite value,

$$V_{12} = \ln (e^{\theta' z_1/\rho} + e^{\theta' z_2/\rho}). \tag{13}$$

The trinomial choice between 1, 2, and 3 can, at the upper level of the tree, be written as a binomial choice between z_3 and an inclusive alternative V_{12}, with choice probability

$$P(3 \mid z_3, V_{12}) = e^{\theta' z_3}/[e^{\theta' z_3} + e^{(1-\sigma)V_{12}}]. \tag{14}$$

This multinomial-logit form permits estimation of $(1 - \sigma)$. Then, all the parameters of this GEV model can be estimated by the empirically practical method of multinomial-logit estimation at each of the two nested levels of the decision tree. The parameter σ is a measure of the degree of similarity between 1 and 2, with $\sigma = 0$ corresponding to the "red bus/blue bus" example. Consistency with utility maximization requires $0 \leqq \sigma \leqq 1$.

The Williams-Daly-Zachary Theorem

A family of probabilistic-choice models consistent with the random-utility model, and containing the GEV family, are characterized by the following theorem, whose essential ingredients were proved by Williams (1977) and by Daly and Zachary (1979):

Theorem

Suppose $G(y_1, \ldots, y_J)$ is a nonnegative, homogeneous-of-degree 1 function of $(y_1, \ldots, y_J) \geqq 0$, and is positive whenever any argument is positive. Suppose for any distinct $(i_1, \ldots, i_k)$ from $\{1, \ldots, J\}$, $\partial^k \log G/\partial y_{i_1} \ldots \partial y_{i_k}$ is nonnegative if k is odd and nonpositive if k is even. Then,

$$P_i = \partial \log G(v_1, \ldots, v_J)/\partial \log v_i, \tag{15}$$

where v_i is a scale value for alternative i, defines a choice model consistent with utility maximization.

This result differs from the characterization of the GEV model in that the mixed partial derivatives of $\log G$, rather than G, are signed. It is

easy to show that the conditions of the GEV theorem are a special case of the conditions in this theorem.

The Williams-Daly-Zachary theorem characterizes random-utility models where the utility of an alternative can be written as the sum of a nonstochastic scale value and a random effect, with the distribution of the random effects independent of the scale values. Note that the distribution of the random effects can depend on aspects of alternatives other than their scale values, such as their direction or clustering in attribute space. (Without such dependence, the resulting choice models are simply scalable.) A proof and further discussion of the Williams-Daly-Zachary theorem is given in McFadden (1978). This theorem provides a useful foundation for constructing or checking probabilistic-choice models where consistency with the random-utility model is required.

Generic and Nominal Models

An alternative is described by a vector z of measured attributes. It may also have a vector y of unmeasured attributes (with the randomness of utility attributed to variations in y). The vector z contains information on generic (or, intrinsic) properties of the alternative, and in addition nominal (or extrinsic) information such as labels attached by the observer for purposes of identification. For example, a transportation-mode alternative may be described by generic variables such as time, cost, and number of stops as well as nominal labels such as "bus," "express," "Alternative No. 4." It is reasonable to postulate that behavior depends solely on generic properties of an alternative. However, an observed nominal label which is "correlated" with the unobserved generic variable may appear to be related to choice. For example, a label which identifies a transportation mode as bus may be correlated with an unobserved generic variable measuring the schedule flexibility of the mode, and hence may act as a "proxy" for the generic variable. The "similarity" between alternatives should also be perceived by the subject in generic terms. Again, nominal labels may serve as proxies for unobserved generic indicators. The plausibility of the red bus/blue bus example is based on the assumption that the label "bus" identifies alternatives with closely related unobserved attributes.

Generic models are desirable for forecasting market behavior, particularly the demand for new products. Knowledge of the historical effect of nominal variables, reflecting underlying unmeasured generic effects, is of little use in forecasting the demand for a product whose unmeasured generic attributes have changed.

Empirical experience in travel-demand forecasting suggests that it is difficult to construct purely generic choice models, using conventional market data alone, and obtain plausible fits to observed choice behav-

ior. Hence, the success of this approach to forecasting market behavior would appear to depend on comprehensive measurement of attributes of alternatives.

Measurement Issues

The measurement requirements imposed by generic probabilistic-choice models may be impractical in some marketing applications, where a full inventory of physical attributes of a product would be excessively long and expensive, and would raise fundamental questions of the appropriate dimensions for measurement. Since the proximate objective of these measurements is determination of their impact on the utility of alternatives, direct psychological scaling of consumer ratings of these attributes would appear to be a more direct and economical alternative. Two constraints should be imposed on the construction of these scales due to the context in which they will be used. First, the scales themselves will become explanatory variables in a probabilistic-choice model of market behavior. Hence, a primary criterion in scale construction should be their ability to provide predictive power in a choice model. Candidate scales can be compared by testing the forecasting accuracy of choice models employing them as variables. Note that this extrinsic criterion for scale construction is quite different from the intrinsic criteria commonly employed in multidimensional scaling.

Second, if the scales constructed in the manner described above are to be useful for forecasting purposes, they must be tied back to the physical attributes of products which are subject to design decisions by suppliers. This can be achieved by econometric models relating scale levels to physical attributes, or by psychological methods with trial products.

Attitudes and Behavior

Market behavior in response to an objective decision environment is presumably mediated by attitudes and perceptions. In fact, much of marketing seems to be concerned with modifying the link between the objective attributes of products and the consumer's perception of them. What, then, should be the role of attitude and judgment scales in a market-behavior forecasting system? As the preceding section indicated, psychological scales can be treated simply as attributes of alternatives in probabilistic-choice models, provided they are constructed so as to contribute to the explanation of behavior on one hand and to be tied to the experience of the consumer and the objective attributes of products on the other. There is no requirement that the explanatory variables in a choice model be exclusively objective attributes or ex-

clusively psychological scales; economy and plausibility should usually suggest a mix. The demand system will now be "structural." In addition to the choice model, there will be equations (or, more generally, mappings obtained by laboratory or trial-product experiments) relating perceptions and changes in attitudes to experience and to the objective attributes of products. For some forecasting purposes, a reduced form obtained from this system by solving out intervening attitude and perceptual scales and relating behavior to objective attributes will be relevant. For others, where the problem involves modifying the structural relationship between attitudes and objective attributes, each structural equation will be used separately. In both cases, psychological measurement has a well-defined interconnection with behavior and with the objective attributes of products.

Statistical Methods for Estimating Probabilistic-Choice Models

Most parametric probabilistic-choice models have a nonlinear structure requiring a general-purpose estimation technique such as nonlinear least-squares or maximum-likelihood estimation. For some models, notably the MNL model, efficient computer programs have been developed and extensive experience has been accumulated in their use.[1] More recently, an initial version of a "production" multinomial probit program has been released (see Lerman and Manski 1980).

Most of the statistical procedures and empirical applications of probabilistic-choice models to date have treated survey data obtained from exogenous random or stratified sample designs, where the actual choice behavior of subjects does not affect their probability of selection in the sample.

Several recent papers have considered the problem of statistical estimation of choice models using data collected by sampling procedures other than random sampling. Of particular interest are choice-based samples, utilizing data collected from "purchase" or "point-of-sale" surveys. Such data sources are often available to analysts from sales or warranty records or can be commissioned at low cost relative to random household surveys. Manski and Lerman (1977) have shown that treating choice-based samples as if they were random and calculating estimators appropriate to random samples will generally yield inconsistent estimates.[2] They introduce a weighted likelihood function whose maximization is shown to yield consistent estimates.

Manski and McFadden (1980) have considered more generally the problem of estimation of discrete-choice models under alternative

1. One example in the public domain is the QUAIL program for MNL analysis which I developed.

2. In a MNL model with alternative-specific dummy variables, the inconsistency is confined to the dummy variable coefficients.

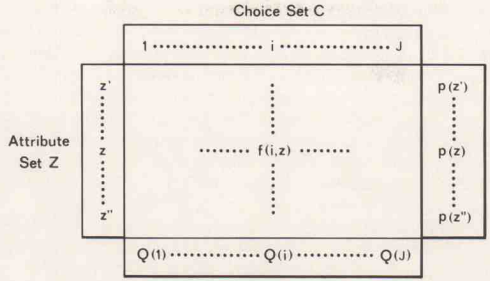

FIG. 3.—Contingency table layout of observations

sample designs. The discrete-choice problem can be defined by a finite set C of mutually exclusive alternative responses, a space of attributes Z, assumed to be a subset of a finite-dimensional vector space, a probability density $p(z)$ $[z \in Z]$, giving the distribution of attributes in the population, and *response probability* or *choice probability* $P(i \mid z,\theta^*)$, specifying the conditional probability of selection of alternative $i \in C$, given attributes $z \in Z$. Prior knowledge of causal structure is assumed to allow the analyst to specify the response model $P(i \mid z,\cdot)$ up to a parameter vector θ^*. The analyst's problem is to estimate θ^* from a suitable sample of subjects and their associated responses.

The probability density of (i,z) pairs in the population is given by

$$f(i,z) = P(i \mid z,\theta^*)p(z), \qquad [(i,z) \in C \times Z]. \tag{16}$$

The analyst can draw observations of (i,z) pairs from $C \times Z$ according to one of various sampling rules. The problem of interest is first, given any sampling rule, to determine how θ^* may be estimated, and second, to assess the relative advantages of alternative sampling rules and estimation methods.

The data layout can be visualized using a contingency table, as illustrated in figure 3. An observation (i,z) occurs in the population with frequency $f(i,z)$. The row sums give the marginal distributions of attributes $p(z)$, while the column sums give the population shares of responses $Q(i)$. The joint frequency $f(i,z)$ can be written either in terms of the conditional probability of i given z, or choice probability, or in terms of the conditional probability of z given i, as the formulae in the figure illustrate.

The feature of the probabilistic-choice problem which distinguishes it from the general analysis of discrete data is the postulate that the response probability $P(i \mid z,\theta^*)$ belongs to a known parametric family

and reflects an underlying link from z to i which will continue to hold even if the distribution $p(z)$ of the explanatory variables changes.[3] Alternatively, given a population $C \times Z$ with probability distribution specified by $f(i,z)$, one might, in the absence of any knowledge of the process relating i's to z's, obtain a random sample from $C \times Z$ and directly examine the joint distribution $f(i,z)$. This exploratory data analysis approach is exemplified by the literature on associations in contingency tables, where it is assumed only that Z is finite (see, e.g., Goodman and Kruskal 1954; Haberman 1974; and Bishop, Fienberg, and Holland 1975).

If one believes that the elements of C index conceptually distinct populations of z values, then the natural analytical approach is to decompose $f(i,z)$ into the product $f(i,z) = q(z \mid i)Q(i)$, where $q(z \mid i)$ gives the distribution of z within the population indexed by i and $Q(i)$ is the proportion of the population with this index. This is the approach taken in discriminant analysis. There, prior knowledge allows the analyst to specify $q(z \mid i)$ up to a parametric family, and a sample suitable for estimating the unknown parameters is obtained from the subpopulation i (see, e.g., Kendall and Stuart 1976 and Anderson 1958).

Models for data analysis of associations, or discriminant analysis, imply (by Bayes law) probabilistic-choice functions. These derived-choice probabilities may be inconsistent with economic-choice behavior. When they are consistent, a separate and interesting question is whether the parameter vector θ^* can be estimated conveniently from the associated model of association or discrimination.

Manski and McFadden (1980) attempt to provide a general theory of estimation for quantal response models. The scope of the investigation is as follows: Consider the problem of estimating θ^* from stratified samples of (i,z) observations. A stratified sampling process is one in which the analyst establishes an index set B, partitions $C \times Z$ into strata $S_b \subseteq C \times Z, b \in B$, and specifies a suitable probability distribution over B. To obtain an (i,z) observation, he draws stratum according to the specified distribution on B and then samples at random from within the drawn stratum.

Within the class of all stratification rules, two symmetric types of stratification are of particular statistical and empirical interest. In "exogenous" sampling, the analyst partitions Z into subsets $Z_b, b \in B$, and lets $S_b = C \times Z_b$. In "endogenous" or "choice-based" sampling, he partitions C into subsets $C_b, b \in B$, and lets $S_b = C_b \times Z$. Less formally, in exogenous sampling the analyst selects decision makers and ob-

3. This postulate is fundamental to the concept of "scientific explanation." If the response probability function is invariant over populations with different distributions of attributes, then it defines a "law" which transcends the character of specific sets of data. Otherwise, the model provides only a device for summarizing data and fails to provide a key ingredient of "explanation"—predictive power.

serves their choices, while in choice-based sampling the analyst selects alternatives and observes decision makers choosing them. In figure 3 exogenous sampling corresponds to stratifying on rows and then sampling randomly from each row, while choice-based sampling corresponds to stratifying on columns and then sampling randomly from each column.

Manski and McFadden make a detailed statistical examination of maximum-likehood estimation of θ^* in both exogenous and choice-based samples. They find that application of maximum likelihood is wholly classical in exogenous samples. In choice-based samples, however, the form of the maximum-likelihood estimate (MLE) depends crucially on whether the analyst has available certain prior information, namely, the marginal distributions $p(z)$, $z \in Z$, or $Q(i)$, $i \in C$, where $Q(i) = \Sigma_z P(i \mid z,\theta^*)p(z)$.

The maximum-likelihood estimator of θ in a choice-based sample when p is known and Q is unknown satisfies

$$\max_{\theta\in\Theta} \sum_{n=1}^{N} \log P(i_n \mid z_n,\theta) - \sum_{n=1}^{N} \log \sum_{Z} P(i_n \mid z,\theta)p(z), \tag{17}$$

where (i_n,z_n) is observed for a sample $n = 1, \ldots, N$. When Q and p are both known, (17) is maximized subject to the constraints

$$Q(i) = \sum_{Z} P(i_n \mid z,\theta)p(z). \tag{18}$$

When p is unknown, the classical conditions for maximum-likelihood estimation are not met. However, several alternative nonclassical maximum-likelihood and pseudo-maximum-likelihood methods are available which yield consistent estimators.

When Q is known and p is unknown, Cosslett (1980) has shown that the nonclassical full-information maximum-likehood estimator satisfies

$$\max_{\theta\in\Theta} \min_{\lambda\geq 0} \sum_{n=1}^{N} \log\left\{P(i_n \mid z_n,\theta)\left[\sum_{j\in C} \lambda_j Q(j)\right] \Big/ \left[\sum_{j\in C} \lambda_j P(j \mid z_n,\theta)\right]\right\}. \tag{19}$$

A second estimator, introduced by Manski and Lerman (1977) satisfies

$$\max_{\theta\in\Theta} \sum_{n=1}^{N} w(i_n) \log P(i_n \mid z_n,\theta), \tag{20}$$

where $w(i) = Q(i)/H(i)$ and $H(i)$ is the sampling frequency for alternative i.

If both p and Q are unknown in a choice-based sample, then, provided an identification condition is satisfied,[4] Manski and McFad-

4. An important case in which the identification condition *fails* is the MNL model, where in the absence of a knowledge of Q there is a confounding of the effects of Q and alternative-specific dummies.

den show that the nonclassical full-information maximum-likehood estimator satisfies

$$\max_{\theta\in\Theta}\max_{\lambda\geq 0}\sum_{n=1}^{N}\log\left[P(i_n\mid z_n,\theta)\lambda_{i_n}\Big/\sum_{j\in C}P(j\mid z_n,\theta)\lambda_j\right]. \qquad (21)$$

Note that one can, with some loss of efficiency, obtain consistent estimates for an information case by using a consistent estimator which ignores some available information. For example, the estimators (19) or (20) could be used in the case both p and Q are known, and the estimator (21) could be used in any of the information cases.

The choice-based sampling estimators above have been extended by Cosslett (1980) to a case which has particular potential value for application to marketing data. Suppose $p(z)$ is observed, say, from the U.S. Census Public Use Sample or other general data sources, and a choice-based sample is drawn for a *single* alternative, say, data from warranty registrations of purchasers. (This is termed an "enriched" sample.) Then, modified versions of the criterion (19) or (21) can be used to estimate the parameters of the probabilistic-choice function. More generally, this approach to the question of sample design and estimation method suggests that an integrated analysis of data availability, choice model structure, survey cost, and statistical method can improve the flexibility, precision, and cost effectiveness of market forecasts.

Conclusion

This paper has surveyed recent developments in the specification and estimation of econometric models of probabilistic choice. The motivation for this work has been the problem of forecasting the impact on demand of introducing new commodities or modifying aspects of existing commodities. Particular attention is given to correctly representing patterns of similarities between alternatives and to developing statistical methods for estimation of probabilistic-choice models from "purchase" surveys. The applications of these methods in transportation and labor economics appear similar to problems in market research, suggesting that some of these methods may be useful in the latter discipline.

References

Anderson, T. W. 1958. *An Introduction to Multivariate Statistical Analysis*. New York: Wiley.

Ben-Akiva, M., and Lerman, S. 1977. Disaggregate travel and mobility choice models and measures of accessibility. Mimeographed. Forthcoming in P. Stopher (ed.), *Third International Conference on Behavioural Travel Modelling*.

Bishop, Y.; Fienberg, S.; and Holland, P. 1975. *Discrete Multivariate Analysis*. Cambridge, Mass.: M.I.T. Press.

Bock, R. D., and Jones, L. V. 1968. *The Measurement and Prediction of Judgement and Choice*. San Francisco: Holden-Day.

Cardell, S. 1977. A choice model without independence from irrelevant alternatives. Mimeographed. Charles River Associates.

Cosslett, S. 1980. Efficient estimation of discrete choice models. In C. Manski and D. McFadden (eds.), *Structural Analysis of Discrete Data*. Cambridge, Mass.: M.I.T. Press.

Daganzo, C.; Bouthelier, F.; and Sheffi, Y. 1977. Multinomial probit and qualitative choice: a computationally efficient algorithm. *Transportation Science* 11, no. 4 (November): 338–58.

Daly, A., and Zachary, S. 1979. Improved multiple choice models. In D. Hensher and Q. Dalvi (eds.), *Identifying and Measuring the Determinants of Mode Choice*. London: Teakfield.

Goodman, L., and Kruskal, W. 1954. Measures of association for cross-classification. *Journal of the American Statistical Association* 49, no. 268 (December): 732–64.

Haberman, S. 1974. *The Analysis of Frequency Data*. Chicago: University of Chicago Press.

Hausman, J., and Wise, D. A. 1978. A conditional probit model for qualitative choice: discrete decisions recognizing interdependence, and heterogeneous preferences. *Econometrica* 48, no. 2 (March): 403–26.

Kendall, M., and Stuart, J. 1976. *Advanced Theory of Statistics*. Vol. 3. New York: Hafner.

Lerman, S., and Manski, C. 1980. On the use of simulated frequencies to approximate choice probabilities. In C. Manski and D. McFadden (eds.), *Structural Analysis of Discrete Data*. Cambridge, Mass.: M.I.T. Press.

Luce, R. D. 1959. *Individual Choice Behavior: A Theoretical Analysis*. New York: Wiley.

McFadden, D. 1973. Conditional logit analysis of qualitative choice behavior. In P. Zarembka (ed.), *Frontiers in Economics*. New York: Academic Press.

McFadden, D. 1974. Measurement of urban travel demand. *Journal of Public Economics* 3, no. 4 (November): 303–28.

McFadden, D. 1976. Quantal choice analysis: a survey. *Annals of Economic and Social Measurement* 5, no. 4 (January): 363–90.

McFadden, D. 1978. Modelling the choice of residential location. In A. Karlquist et al. (eds.), *Spatial Interaction Theory and Planning Models*. Amsterdam: North-Holland.

McFadden, D. 1980. Econometric models of probabilistic choice. In C. Manski and D. McFadden (eds.), *Structural Analysis of Discrete Data*. Cambridge, Mass.: M.I.T. Press.

McFadden, D.; Talvitie, A.; Cosslett, S.; Hasan, I.; Johnson, M.; Reid, F; and Train, K. 1978. *Demand Model Estimation and Validation*. Institute of Transportation Studies, Urban Travel Demand Forecasting Project, Final Report Series, vol. 5. Berkeley: University of California Press.

Manski, C., and McFadden, D. 1980. *Alternative Estimators and Sample Designs for Discrete Choice Analysis*. In C. Manski and D. McFadden (eds.), *Structural Analysis of Discrete Data*. Cambridge, Mass.: M.I.T. Press.

Manski, C., and Lerman, S. 1977. The estimation of choice probabilities from choice-based sampling. *Econometrica* 45, no. 8 (November): 1977–88.

Punj, G., and Staelin, R. 1978. The choice process for graduate business schools. *Journal of Marketing Research* 15 (November): 588–98.

Silk, A., and Urban, G. 1978. Pre-test-market evaluation of new packaged goods: a model and measurement methodology. *Journal of Marketing Research* 15 (May): 171–91.

Thurstone, L. 1927. A law of comparative judgement. *Psychological Review* 34 no. 4 (July): 272–86.

Tversky, A. 1972. Elimination by aspects: a theory of choice. *Psychological Review* 79, no. 4 (July): 281–99.

Williams, H. C. L. 1977. On the formation of travel demand models and economic evaluation measures of user benefit. *Environment and Planning* A.9, no. 3 (March): 285–344.