Behavioral Travel-Demand Models

Edited by
Peter R. Stopher
Northwestern University

Arnim H. Meyburg
Cornell University

Lexington Books
D.C. Heath and Company
Lexington, Massachusetts
Toronto

1976

17 The Mathematical Theory of Demand Models

Daniel L. McFadden[a]

Charge to Workshop

The concern of this workshop is with alternative theories of travel-demand models and the operational implications of these theories. For example, the independence-of-irrelevant alternatives axiom leaves open the operational question of the definition of an alternative. Disaggregate demand theories have also tended to ignore questions of aggregation and definition of choice sets. These and other problems should be addressed, and definition of needed work in theoretical developments should be attempted.

Summary of Recommendations

A point that will be obvious to conference members but must be made in presenting behavioral approaches to more general audiences, is that behavioral-demand modeling is *not* synonymous with the multinomial-logit model (MNL) and its independence-of-irrelevant alternatives (IIA) property, and that the attention concentrated on IIA is directed to the questions of when it is valid and what procedures to adopt when it is not. It must be emphasized that use of the MNL model entails many computational advantages and convenient properties, and that as a structural restriction the IIA is on a par with many of the properties, often unstated, that restrict other demand models, behavioral or not. It should also be emphasized that, as a practical matter, it is likely to be feasible to develop satisfactory forecasting models for most applications *within* the MNL framework, with minor respecifications to correct for failures of the IIA.

We conclude that a substantial research effort should be devoted to behavioral alternatives to the MNL model, but that further development and application of models based on the MNL form should also be encouraged. In particular, we conclude that emphasis should be placed on refining the definition and measurement of variables, and detailed specification, estimation, and validation of existing behavioral models.

[a]The generous support of the National Science Foundation, through grant SOC 72-05551 A02, Division of Social Sciences, and grant GI-43740, Office of Research Applied to National Needs, to the University of California, Berkeley, is gratefully acknowledged.

The development of sound statistical procedures for estimation, forecasting, diagnostic tests of assumptions, and tests of validity, has lagged behind model development. There is a general need to raise the consciousness of planners as to the efficient utilization of data and the reliability of empirical results and to develop statistical procedures tailored to the use of transportation planners (i.e., to develop a field of "transmetrics").

Among the remaining research topics discussed in the workshop, we recommend further research on the following, roughly in order of priority:

1. The problem of defining alternatives and identifying the choice set, including the question of appropriate models for choice among aggregates of alternatives
2. The problem of aggregating from individual to population behavior, and the development of correction procedures for aggregation
3. The problem of integrating the use of strictly behavioral data with observations on intentions, attitudes, perceptions, and responses to hypothetical situations, and with more generalized observations on aggregate transportation flows and revenues, trip patterns, and "life styles," and political and social attitudes related to transportation
4. The workshop recommends the formation of a "Committee on Analytic and Statistical Methods for Transportation Planning," which would serve both the Travel Demand Forecasting and Travel Behavior Committees.

Discussion of Workshop Recommendations

This report reviews topics recently analyzed or suggested in behavioral demand modeling, and suggests topics for further research. It should be noted at the outset that this topic is concerned with the description, understanding, and prediction of the full complex of transportation-related decisions of consumers.

Briefly, it is assumed that travel-demand behavior and consequently the travel-demand models incorporate not only the short-run travel choices of trip frequency, choice of destination, mode, and route, but also the longer-run mobility choices of location of residence and job, car ownership, and perhaps travel mode to work. These choices are presumed to be intertwined with the mobility choices made with a longer time frame in mind and the travel choices made conditional to the mobility choices (see the preceding chapter by Talvitie).

The specific concern of this workshop is with the underlying axioms assumed to govern choice, and their validity, with the development of practical demand models consistent with plausible axiomatic foundations, and with the development of statistical and computational methods for the empirical application of these models.

Basic Principles

Several basic principles should govern the development of behavioral demand models:

1. The axiomatic theory of travel demand should be formulated at the level of the individual decision maker, and should consider the full set of alternatives facing the individual. Models of choice among aggregates of alternatives, or for aggregates of individuals, should be derived explicitly (although not necessarily constructively) from the theory of individual choice.
2. A general axiom for individual choice behavior is that each individual has a preference ordering of conceivable alternatives and chooses among available alternatives to maximize these preferences. This axiom should be a common starting point for all behavioral theories of demand. (Note that until "alternative" is defined, this axiom is useful only as an organizing concept.)
3. The goal of theory is to construct models that yield valid insight and prediction with minimum complexity and measurement cost. Occam's razor should be applied to each modeling effort. Suppose, for example, the primary objective of the theory is to make accurate forecasts for aggregates of individuals. Ideally, a valid model can be deduced in which aggregate behavior can be described as a simple function of readily measured aggregate engineering-system variables. More likely, to obtain a valid model it will be necessary to take into account socioeconomic and psychological variables and the distribution of these variables in the population aggregate. As indicated above, the emphasis and extensiveness of variables used to define alternatives should be governed by the objectives of obtaining simple, cost-effective, valid models. Taking into consideration collection cost, reliability, and reproducibility, demand modeling should not hesitate to use any of the engineering, socioeconomic, attitude, perceptual, or psychological variables that can be collected on consumers.

Outstanding Issues

We have identified the following outstanding issues to be addressed by current and future research:

1. Functional forms for choice probabilities and their implications; particularly, the validity of the "Independence-of-Irrelevant-Alternatives" property of the widely used multinomial-logit (MNL) model and alternatives to the MNL model
2. Specification of variables, including: (a) the use of generic versus al-

ternative-specific variables; (b) joint dependence of an observed aspect of choice with an "apparent" explanatory variable; and (c) the use of simple scales to represent complex constellations of engineering or psychological attributes

3. Identification of the "set of feasible alternatives" (the choice set), including the issues of constructing practical models with large numbers of alternatives, construction of models in which "nonfeasible" alternatives are deleted as part of the decision process, and development of theories in which perceptual aggregation of alternatives in tree-decision structures is captured adequately
4. Aggregation of large numbers of "feasible" alternatives into groups suitable for estimation and forecasting
5. Aggregation of individual decision models into aggregate demand models for use in forecasting
6. Development of statistical estimation procedures and diagnostic tests for the analysis of choice models, the validity of their assumptions, and the accuracy of their predictions

Functional Forms and Independence

The Independence-of-Irrelevant-Alternatives (IIA) property of the MNL model is discussed in detail by Talvitie in the preceding chapter and by Daniel McFadden.[b] The following paragraphs summarize current findings.

The IIA property has extremely desirable implications for estimations and forecasting. Hence, it should be assumed and exploited in any application where diagnostic tests support its validity.

All choice models currently feasible for flexible production demand analysis either have the IIA property, or deviate from it in ways that fail to capture the effects introduced by patterns of similarity and dependence. Talvitie has conjectured that upper and lower bounds on choice probabilities can be obtained from the MNL model and the "maximum" model for subalternative choices.

Several models based on expansions about the MNL with interdependency terms introduced have been investigated. McLynn's "fully competitive" model is a one-parameter expansion that has proven unsuccessful in dealing with the independence problem. Monte-Carlo analysis of data generated from the "red bus/blue bus" example shows that the fully

[b] Daniel McFadden, "On Independence, Structure, and Simultaneity in Travel Demand Analysis," Working Paper No. 7511, Travel Demand Forecasting Project, Institute of Transportation and Traffic Engineering, University of California, Berkeley, 1975.

competitive model yields choice probabilities agreeing with the MNL model to the second decimal place. Neither model gives a reasonable reproduction of the true "red bus/blue bus" probabilities. A more general expansion in powers of jointly normal interdependencies, first suggested by Talvitie, has proven to have unsatisfactory convergence properties, making it computationally unsatisfactory. General alternatives such as multinomial probit with general interdependencies are not currently practical for more than three alternatives because of computational requirements, and appear to involve massive problems of numerical analysis.

At the current stage of development, in which the MNL model is the most practical, there are immediate needs for diagnositic tests of independence, and for ad hoc methods of adjusting the MNL model when independence tests fail. For example, a procedure suggested by McFadden is to consider a choice model of the MNL functional form shown in equation (17.1).

$$P_i = \exp V_i \Big/ \sum_{j=1}^{J} \exp(V_j) \tag{17.1}$$

V_i is given by equation (17.2)

$$V_i = \beta' \ln z^i + \sum_{j=1}^{J} \gamma_{ij} \ln z^j \tag{17.2}$$

where P_i is the choice probability for alternative i from alternatives $1, \ldots, J$, β, γ_{ij} are parameters normalized so that $\gamma_{ij} = 0$ and $\Sigma_{i=1}^{J}\Sigma_{j=1}^{J}\gamma_{ij} = 0$; and $\ln z^i$ is the vector of logs of independent variables z^i. For this model, the cross-elasticity of demand for P_i with respect to z_ℓ^k is given by equation (17.3) (for $i \neq k$).

$$E_{k,\ell}^i \equiv \frac{z_\ell^k}{P_i} \frac{\partial P_i}{\partial z_\ell^k} = -P_k \beta_\ell + \gamma_{ik\ell} - \sum_{j=1}^{J} P_j \gamma_{jk\ell} \tag{17.3}$$

Differences in elasticities for different alternatives are determined by the $\gamma_{ik\ell}$, as shown in equation (17.4).

$$E_{k,\ell}^i - E_{k,\ell}^j = \gamma_{ik\ell} - \gamma_{jk\ell} \tag{17.4}$$

Thus, it is possible to obtain a pattern of cross-elasticities matching a pattern of similarities among alternatives. For P_k bounded away from zero and appropriate parameter values, these elasticities will have the expected signs. Since this model can be estimated with a standard MNL program, it provides a practical current alternative if the data set fails to satisfy the diagnostic tests for IIA. However, because this model is not derived from behavioral foundations and does not permit an easy interpretation of coefficients from the standpoint of individual choice, it should be replaced as soon as practical behavioral models with dependence become available.

Diagnostic tests for IIA can be based on models with interaction terms as described above, by testing the stability of coefficients in MNL for restricted choice sets, and by analysis of residuals.

Specification of Variables

Three areas are currently being investigated which deal with the specification of variables. The first deals with the question of generic versus alternative-specific models. Because of ease of interpretation and flexibility in forecasting, generic variables should be used whenever possible. Scattered empirical results indicate that generic models do well. Nevertheless, the presence of unobserved attributes that are correlated with nongeneric variables may make it necessary to introduce these variables.

The second area is a concern with jointly dependent variables, and with the use of explanatory variables, which are themselves influenced by the decision process. Examples are the use of auto ownership variables in explaining mode choice when ownership and mode are joint decisions, or the use of variables correlated with location (time, cost, neighborhood density) to explain mode choice when location and mode choice are joint decisions. Untreated, such effects can bias seriously model-parameter estimates. A currently available method for dealing with this problem is explicit consideration of the joint choice, as has been done by Ben-Akiva for mode and destination. An important problem for investigation is the development of statistical methods for treating joint dependence without data collection and estimation for the complete system.

The third problem in specification of variables is concerned with the use of scales summarizing the information contained in a large number of attitude variables or engineering-system attributes. The use of these variables in their raw form makes model estimation computationally difficult and unstable due to high multicollinearity, and this complicates interpretation. An alternative is to amalgamate closely related variables into one or more scales that have plausible interpretations. Examples would be a scale of mode convenience constructed from an inventory of attitude questions, or a mode-comfort scale constructed from engineering features. Psychological scaling techniques are currently used to carry out these constructions. A problem deserving investigation is the development of consistent statistical procedures for scale construction and choice-model estimation, particularly the use of choice-residual data, reduced by removing the effects of major variables such as time and cost, in the scaling criterion function.

Definition of the Choice Set

Upon close examination, simple specifications of alternatives often dissolve into a bewildering complex of subalternatives (e.g., auto drive, auto ride with family member, auto ride with non-family members, bus with walk access, etc.). Further, it is often unclear from the application what alternatives are viewed as feasible by the individual in making a choice, and the extent to which the individual forms perceptual aggregates of alternatives during the choice process. Finally, it is critical for data collection and estimation that the models be able to describe choice among aggregates of alternatives.

Within the framework of the Independence-of-Irrelevant-Alternatives (IIA) assumption, Lerman has investigated the effects of aggregating alternatives. We summarize his results in a way that also covers nonfinite choice sets. Let X denote the space of elemental alternatives, assumed to be a metric space, $\mathfrak{X}$ the Borel field of subsets of X, and μ a sigma-finite measure on X. Interpret $\mu(A)$ as a measure of the number of elemental alternatives in A, expressed as a proportion of the total. If, for example, X is the plane on which the individual must choose a residential location, μ could be ordinary area (Lebesgue) measure. Let $P(A, B)$ denote the probability that an alternative from A is chosen when B is the feasible set. The IIA condition is that for any measurable sets $A \subseteq B \subseteq C$, $P(A, B)P(B, C) = P(A, C)$. Suppose the choice probabilities satisfy this condition, and suppose that $P(\cdot, X)$ is absolutely continuous with respect to μ; i.e., $\mu(B) = 0$ implies $P(B, X) = 0$. Then by the Radon-Nikodym theorem there exists a unique (i.e., μ) function g:$X(0, \infty)$ such that $P(A, X) = \int_A g(x)\mu(dx)$. Combining this result with the IIA property implies equation (17.5).

$$P(A, B) = \int_A g(x)\mu(dx) \Big/ \int_B g(x)\mu(dx) \tag{17.5}$$

wherever $P(B, X) > 0$.

Suppose now that X is partitioned into "alternative groups" $X_1, \ldots, X_N$, with $A = X_1$ and $B = X_1 \cup \ldots \cup X_J$. Then equation (17.6) results.

$$\begin{aligned} P(X_1, B) &= \int_{X_1} g(x)\mu(dx) \Big/ \sum_{j=1}^{J} \int_{X_j} g(x)\mu(dx) \\ &= g(\bar{x}^1)\mu(x_1) \Big/ \sum_{j=1}^{J} g(\bar{x}^j)\mu(X_j) \end{aligned} \tag{17.6}$$

where the mean value theorem has been used to re-express the integrals.

Eliminate any groups with

$$\int_{X_j} g(x)\mu(dx) = 0$$

and for the remainder define $v_i = \log g(\bar{x}^1)$. Then, we have equation (17.7).

$$P(X_1\ B) = \exp[v_1 + \log \mu(X_1)] \Big/ \sum_{j=1}^{J} \exp[v_j + \log \mu(X_j)] \qquad (17.7)$$

Interpreting this formula, the choice probabilities for "alternative groups" are determined by "average" utility levels v_i for the groups and by the sizes of these groups $\mu(X_i)$. In the case of counting measure for finite X, this formula coincides with that developed by Lerman. If the IIA assumption is correct, then the effects of further aggregation or disaggregation can be accounted for through the terms $\log \mu(X_i)$. A second method for utilizing the choice model (equations (17.5) or (17.7)), suggested by Ben-Akiva, is to parameterize or idealize the choice set directly. Suppose, for example, individuals are choosing shopping destinations. Suppose the "impedance per mile" of a trip is a constant z, and $v(x) = -\beta zx$, where x is distance. Suppose there are $2\pi x\, dx$ alternatives at distance x. Then, the probability of choosing to make a trip of length exceeding y is proportional to equation (17.8).

$$\int_y^\infty e^{v(x)} 2\pi x\, dx = 2\pi \int_y^\infty e^{-\beta z x} x\, dx = \frac{2\pi}{\beta^2}(1 + \beta zy)e^{-\beta zy} \qquad (17.8)$$

Since this probability must equal one at $y = 0$, we obtain the following distribution of trip lengths by individuals facing the impedance level z per mile, equation (17.9).

$$\text{Prob [distance from CBD} \geqq y] = (1 + \beta zy)e^{-\beta zy} \qquad (17.9)$$

Alternative assumptions on the form of $v(x)$ or the distribution of opportunities could give other trip-length distributions consistent with behavioral assumptions.

While the preceding approach is reasonably tractable analytically, the IIA assumption is of dubious validity. Sociological and psychological results suggest that individuals are limited in their capacity to perceive large numbers of alternatives as distinct, and that seven is a good rule of thumb for the maximum number of perceived alternative groups considered at a time. This would suggest a tree-decision structure, with the dimensions of the tree determined subjectively and unavailable to the transportation planner. It may be possible to identify perceptual groups by psychological scaling methods, or, at minimum, to use scaling methods to determine likely grouping for definition of various choice models. Statistical criteria could then be used to select among likely decision trees. An open research question (to which the discussion of Talvitie in the preceding chapter is relevant) is the definition of "average" attributes of alternative groups.

Aggregation and Aggregate Demand

For applications, it is necessary to aggregate individual demands into a population demand. If all population members are identical with respect to their characteristics and the environments they face, then aggregation is trivial. If these explanatory factors vary in the population, then their distribution must be taken into account in the aggregation process. This can be accomplished by explicit aggregation of individuals, or by numerical or analytic integration over the distribution of explanatory variables. Some findings supporting this notion are the following.

Talvitie has obtained some results showing that when aggregate network information is used to provide time and cost data, a naive procedure of using the individual model with means of independent variables outperformed several more sophisticated methods.

Koppelman found that simple classification methods are effective, and that the naive method performed better when means are adjusted for dispersion. Studies of the Taylor's approximation method developed by Talvitie have found that it often has unsatisfactory convergence properties, making it impractical. An open problem of considerable interest is that of seeking alternative expansion methods or distributional assumptions that would lead to flexible and practical aggregation formulae.

Aggregation biases appear to develop very quickly, even for small zones, and then do not worsen significantly as zone size increases. Thus, once aggregation starts, there appears to be little harm in going to large zones. This conclusion is consistent with empirical findings of McFadden and Reid that most regional variation in socioeconomic characteristics occurs within small zones.

Overall, aggregation biases appear to be modest, suggesting that relatively crude and inexpensive methods for a correction should be adequate. There is a need for further empirical experience with simple bias correction procedures, particularly in the case of multiple choices.

There is a need for a systematic analysis of error propagation in aggregate prediction, taking into account errors in the measurement of explanatory variables, errors in functional specification, statistical errors in estimation, and errors of aggregation. An assessment of the relative seriousness of these errors would form a basis for determining where future research could be directed most profitably.

From the standpoint of model applications, emphasis should be placed on the development of disaggregated models that can be calibrated directly from aggregate data (e.g., zonal averages and measures of intrazonal dispersion) and aggregated conveniently to forecast aggregate demands.

Statistical Methods

An underdeveloped part of transportation-demand analysis is the statistical methodology for calibration and tests of hypotheses. Historically, this is clearly a consequence of a carry-over of engineering methods from laboratory settings where experimental control and abundant data made statistical issues inconsequential. However, the relatively uncontrolled and complex phenomena determining transportation demand and nonrepetitiveness of field data have made questions of model specification and statistical method critical in the interpretation of results. The following specific areas where investigation is needed can be identified:

1. Statistical quality control of data, including the questions of how to proceed with data containing measurement errors and missing observations, with missing variables or imperfect proxies for missing variables, and with imperfectly defined or aggregated variables
2. Development of good estimation procedures for behavioral models, taking into account efficiency, bias, consistency, and robustness (Two methods currently used for the MNL model are maximum likelihood and the Berkson-Theil method. Where data can be formatted so that the second method can be used, it is typically more efficient than maximum likelihood, but more prone to bias. There may be modifications of the maximum-likelihood procedure, such as data adjustments or censoring of observations, which will improve finite-sample efficiency. Manski has introduced a maximum-score estimation procedure, which is less efficient than the preceding methods, but is more robust when the distributional assumptions underlying these methods fail.)
3. Development of diagnostic tests for assumptions, including tests for the validity of the IIA property, for the exclusion of sets of explanatory variables, for joint dependence of choice variables, and for transferability of models from one location to another
4. Development of validation tests of models against baseline data for prediction
5. Development of practical computation procedures for estimation of complex models such as the general purpose multinomial-probit model with dependencies