PART I

THEORY AND ESTIMATION OF BEHAVIORAL TRAVEL DEMAND MODELS

CHAPTER 1

THE THEORY OF ECONOMETRIC CHOICE MODELS AND ESTIMATION OF PARAMETERS

Introduction

A major responsibility of transportation planners is to forecast those changes in travel demand induced by alternative transportation policies. In recent years, the range of analyzed policy alternatives and the range of considered policy questions have greatly expanded. Emphasis has shifted from long-run planning of highway networks to short-run planning and to management of integrated multi-modal transportation systems. These shifts have placed considerable strain on conventional forecasting tools, which were originally developed to address problems of highway network design.

Flexible demand forecasting methods have consequently been sought, particularly those capable of incorporating the behavioral forces linking individual transportation decisions and the relationships between individual travel choices and aggregate flows. The resulting behavioral disaggregate methods expand the policy sensitivity of forecasts. Tests and practical experience with these methods indicate that they are comparable or superior to conventional forecasting techniques in terms of data gathering and computational requirements and forecast accuracy. They provide, in short, a useful way of tackling the expanded list of contemporary planning questions.

We start with the observation that urban travel demand is the result of aggregation over the urban population, each member of which is making individual travel decisions based on his personal needs and environment. These individual decisions are complex, involving trip purpose, frequency, timing, destination, and mode of travel. Further, these choices should be analyzed in the context of simultaneous choice of automobile ownership, housing location, and end-of-trip activities

Define a homogeneous market segment to be a sub-set of the population with identical observed socioeconomic characteristics and observed transportation environments. The foundation of the behavioral approach to travel demand forecasting is to postulate that the distribution of travel behaviors of a homogeneous market segment will be independent of the date or location of observation, and will reflect underlying, stable patterns of human conduct. Behavioral travel demand forecasting will then model the behavior of homogeneous market segments, and aggregate the predicted demands of homogeneous market segments to obtain forecasts of aggregate transportation demand.

Consider, for example, a mode-split for work trips from an origin zone to a destination zone. The aggregate share of a mode is by definition the sum (over market segments) of the share of the mode in each market segment, weighted by the proportion of the total origin-zone population contained in this market segment. Suppose the segmentation is complete, with each observed combination of level-of-service variables and socioeconomic variables defining a market segment. Then,

(1)
$$\begin{aligned}\begin{bmatrix}\text{Aggregate}\\ \text{Share of}\\ \text{a Mode}\end{bmatrix} &= \begin{bmatrix}\text{Share of mode}\\ \text{in first market}\\ \text{segment}\end{bmatrix} \times \begin{bmatrix}\text{Proportion of first}\\ \text{market segment in}\\ \text{population}\end{bmatrix}\\ &+ \begin{bmatrix}\text{Share of mode}\\ \text{in second}\\ \text{market segment}\end{bmatrix} \times \begin{bmatrix}\text{Proportion of second}\\ \text{market segment in}\\ \text{population}\end{bmatrix}\\ &+ \ldots + \begin{bmatrix}\text{Share of mode}\\ \text{in last market}\\ \text{segment}\end{bmatrix} \times \begin{bmatrix}\text{Proportion of last}\\ \text{market segment in}\\ \text{population}\end{bmatrix}.\end{aligned}$$

This relation can be expressed, alternately, in mathematical symbols. Index the modes, $i = 1,2,...,J$. Index the market segments, $n = 1,2,...,N$. Let LOS_n denote the observed level-of-service variables in market segment n , and SE_n denote the observed socioeconomic variables in this segment. Let $p(LOS_n, SE_n)$ denote the proportion of the origin-zone population contained in market segment n . Let $P(i \mid LOS_n, SE_n)$ denote the share of mode i for market segment n . Then $P(i \mid LOS_n, SE_n)$ is the choice probability that individuals with socioeconomic characteristics SE_n will choose mode i when faced with level-of-service attributes LOS_n . Let $Q(i)$ denote the aggregate share of mode i in travel between the given origin and destination zones. Then, formula (1) can be written

$$(1a) \qquad Q(i) = \sum_{n=1}^{N} P(i \mid LOS_n, SE_n) p(LOS_n, SE_n) \quad .$$

In statistical terminology, $Q(i)$ is the expectation of the choice probability $P(i \mid LOS,SE)$ with respect to the distribution of the observed explanatory variables LOS and SE .[1]

In this chapter, we first consider the structure of individual choice behavior, and then the role of observed and unobserved variables in defining homogeneous market segments and their distributions of behavior. Next, concrete choice models derived from this theory are introduced. Finally, methods for estimation and evaluation of the models are discussed.

---

[1] In more general demand settings than mode-split, a generalization of equation (1) gives aggregate demand for a specified class of trips:

$$Q = \sum_{i=1}^{J} \sum_{n=1}^{N} \delta(i) P(i \mid LOS_n, SE_n) p(LOS_n, SE_n) \quad ,$$

where $\delta(i)$ equals the number of trips contributed to the specified class by choice i . For example, if i indexes joint destination and mode choice, and the specified class is trips to a specific destination by any mode, then $\delta(i)$ will be one for alternatives yielding a trip to the specific destination, and zero otherwise.

Individual Choice Behavior

An axiom of classical economic choice theory is that the individual is the basic decision-making unit, and that the individual can rank possible alternatives in order of preference and will always choose from available alternatives the option that he considers most desirable. Modified to take account of the psychological phenomena of learning and perception errors, economic choice theory has been used successfully in analyzing and forecasting in a wide variety of applications.

In the model of consumer behavior that follows, this theory is elaborated to focus on the relationship between consumer behavior and transportation. The consumer is assumed to have a utility function defined on both consumption and transportation attributes. The set of alternatives available to the individual is determined not only by the usual budget constraint, but also by the "household" technology for carrying out work and consumption activities in various locations, and the attributes of transport modes to these locations. Because transport often appears as a "fixed charge" concomitant of consumption activities and involves discrete choices, the set of available alternatives will not be a simple "budget set" of the type ordinarily encountered in consumer theory. Consequently, we will not obtain the usual consumer theory implications drawn from marginal analysis.

Our initial model will consider consumer choice in the abstract, without specific reference to transport. The reader may, however, find it useful to keep in mind the range of transport-related decisions made by the consumer:

(1) The locations of residence and job;

(2) Sales of labor and purchases of commodities, including vehicles;

(3) Frequency of work, shopping, recreation, and other trips;

(4) Destination of trips;

(5) Time of day of travel;

(6) Mode of travel.

To encompass these decisions, which involve short- and long-run choice and the dynamics of consumption activities, it is in general necessary to consider a fully intertemporal theory of behavior.

We formulate our description of the economic consumer within the framework of the Court-Griliches-Becker-Lancaster consumption-activity household-production model. This theory assumes that the individual has a series of basic wants, or drives, as for example "hunger," "thirst," and "rest," and the consumer is assumed to have a "utility" function defined for levels of satisfaction of these wants which "summarizes" his sense of well-being.

Over his lifetime, the individual has available a set A of mutually exclusive alternative choices, with each choice representing a lifetime program of activities, or acts. Each choice of a lifetime consumption activity determines the levels of satisfaction of wants the individual will experience.[1] On the other hand, each lifetime consumption activity determines a vector of attributes describing market commodities purchased, trips taken, work performed, etc. The individual chooses an activity from A that maximizes the derived utility; the corresponding vector of attributes defines his observed demands. In particular, this vector of attributes will specify transport demand behavior along the dimensions listed above.

The model above can be stated formally, although we will not make direct use of this formalism in the future. Assume the individual to have a lifetime extending over a finite sequence of short periods, indexed $v = 1,2,...,H$. Let w_v denote a finite vector of levels of satisfaction of wants in period v , and let $w = (w_1,...,w_H)$ denote the lifetime vector of want satisfaction levels. Individual utility is a function $r = W(w,s)$, where s is a vector of individual social and demographic characteristics influencing tastes. A consumption activity is assumed to be a finite vector a contained in a universe Ω . Associated with each $a \in \Omega$ is a vector of want satisfaction levels $w = Ma$, and a finite vector of attributes $x = Na$, where M and N are taken to be linear transformations. Given a set $A \subseteq \Omega$ of available actions, the consumer solves the problem

$$(2) \qquad \max_{a \in A} W(Ma,s) \quad ,$$

---

[1]In the most general case, each choice of a consumption activity will determine a lottery over levels of satisfaction whose outcome is determined by chance. It will be unnecessary to consider behavior under uncertainty explicitly in our analysis.

with observed demands satisfying $x^* = Na^*$ for the maximand a^* . Of particular interest is the case in which each attribute vector x is identified with a unique activity vector a ; i.e., N is a square non-singular matrix. Then, the set of available alternatives can be expressed directly in terms of observed attributes,

(3) $$B = \{Na \mid a \in A\} \quad ,$$

and utility can be written

(4) $$u = U(x,s) \equiv W(MN^{-1}x,s) \quad ,$$

so that (2) becomes

(5) $$\underset{x \in B}{\text{Max}}\ U(x,s) \quad .$$

To push the analysis beyond this very general statement of the mechanism for determining behavior, it is necessary to give more structure to the form of utility and of the budget constraints that determine the set of available alternatives. McFadden (1974) describes a dynamic model for individual choice that treats satisfaction levels as "states" that satisfy a first-order difference equation, with activity levels in each period influencing the evolution of the states. This framework allows the consideration of long-run choices such as residential and work location, and automobile ownership, as state variables against which short-run mode and destination decisions are made. In principle, evolution of budget possibilities over time, formulation of attitudes as state variables that evolve over time, and "switching" behavior under apparently static conditions can be treated within this framework.

It is usually possible, with little loss of generality, to express intertemporal utility as a discounted sum of "static" felicity functions of contemporary activity levels and a broadly defined list of state variables. Then, the consumer's decision problem is separable into "static" problems of maximizing felicity in each period, taking into account state variables, including the state of budget constraints that extend through time. For the analysis of "steady state" behavior, such as "usual mode to work," it is then sufficient to consider maximization of static felicity. The empirical work in this volume follows this approach, with the further assumption that auto ownership, residential and work location, and attitudes are the state variables that convey the effects of intertemporal decision-making on short-run mode and destination choices. This short list of state variables substantially restricts the generality of the behavioral model, although it is still more general than most models that have been or can be analyzed empirically. A specification of state variables imposes an implicit intertemporally separable structure on the consumer's decision problem. The assumption of intertemporal separability is widely used in empirical consumer demand analysis, but has not been extensively tested. In the application to travel behavior, introspection suggests that auto ownership, residential and work location, and attitudes do indeed capture the primary influences of the long-run environment on behavior, although additional variables such as wealth, habit, and physical condition may matter. In the empirical models in this study, additional state variables are absorbed into the long list of unobserved variables influencing behavior. To the extent that such variables are important, the explanatory power of the model will fall. However, forecasts will be unbiased unless the values of unobserved state variables shift between the calibration and forecast periods.

We now consider one-period utility, or felicity, with state variables included in the lists of observed and unobserved variables for the problem. The list of alternatives will be those available to the consumer in a particular period.

An alternative's attributes include the transportation level-of-service variables associated with its pattern of travel. The individual's utility of an alternative is a function of level-of-service variables for the alternative. Utility also depends on the individual's tastes and background--or socioeconomic characteristics. The individual chooses from the available alternatives the one which maximizes utility.

Some socioeconomic characteristics and level-of-service variables are observed by the transportation planner. Others are unobserved. For example, income and on-vehicle travel time are usually observed, while attitudes toward privacy and vehicle noise-level are usually not observed.

Behavior of a Homogeneous Market Segment

A disaggregate choice model is defined by specifying a probability distribution for the unobserved variables affecting utility, given the values of observed variables in a homogeneous market segment. This probability distribution then determines the choice probabilities--the proportions of the group with maximum utility for each alternative.

The mean utility of a homogeneous market segment is defined to be the average of the utilities of all individuals in the segment.[1] Mean utility depends on determinants of the distribution of unobserved variables. It does not depend on values of unobserved variables, which are averaged out.

The utility of an individual can be written as a sum of mean utility and a deviation due to unobserved idiosyncracies in the individual's tastes and alternatives, as in equation (6) below. The choice probabilities are then determined by the distribution in the market segment of the individual deviations from mean utility.

The determination of choice probabilities for homogeneous market segments can also be described symbolically. Suppose there are J alternatives, indexed $i = 1,...,J$. The utility of alternative i can be written as a function of both observed and unobserved level-of-service attributes and socioeconomic characteristics,

$$u_i = U(LOS^i, ULOS^i, SE, USE) \quad , \tag{6}$$

where LOS^i = vector of observed level-of-service variables;

$ULOS^i$ = vector of unobserved level-of-service variables;

SE = vector of observed socioeconomic characteristics;

USE = vector of unobserved socioeconomic characteristics.

---

[1]Utility scales are usually ordinal, with only the sign of utility differences, and not their magnitude, having significance. The operation of taking means requires a normalization of utilities which itself has no behavioral interpretation. Hence, mean utility in the sense used here should be treated as a computational device rather than a behavioral concept, with the normalization chosen to yield a convenient characterization of the distribution of unobserved components.

The choice probability for alternative i equals the probability of occurrence of unobserved variables such that the utility of i exceeds the utility of any other alternative:

$$(7) \qquad P(i \mid LOS,SE) = \text{Prob}\ \{(ULOS^1,...,ULOS^J,USE) \mid U(LOS^i, ULOS^i, SE, USE) > U(LOS^j, ULOS^j, SE, USE) \quad \text{for} \quad j = 1,..., J \ \text{and} \ j \neq i\} \ ,$$

where $LOS = (LOS^1, LOS^2,..., LOS^J)$ and "Prob" denotes the probability distribution of the unobserved variables.

The utility function in equation (6) can always be written in the form

$$(8) \qquad U(LOS^i, ULOS^i, SE, USE) = V(LOS^i, SE) + \varepsilon(ULOS^i, USE; LOS^i, SE) \ ,$$

where V is the mean value of utility in the homogeneous market segment with observations (LOS,SE) , and $\varepsilon_i = \varepsilon\ (ULOS^i,USE;LOS^i,SE)$ is a deviation from the average having a probability distribution induced by the distribution of the unobserved variables $ULOS^i$ and USE . Equation (7) can then be written

$$(9) \qquad P(i \mid LOS, SE) = \text{Prob}\ \{v_i + \varepsilon_i > v_j + \varepsilon_j \ \text{for} \quad j = 1, \ldots, J \ \text{and} \ j \neq 1\} \ ,$$

where $v_i = V(LOS^i,SE)$.

The Multinomial Logit Model of Choice Probabilities

Assuming a concrete probability distribution for the unobserved components of utility leads to a concrete formula for the choice probabilities. Unfortunately, most distributions of unobserved components yield computationally forbidding choice probability formulae, making them difficult to use in practical calibration and forecasting. One exception is the multinomial logit (MNL) model, which has choice probabilities of the form

(10) $$P(i \mid LOS,SE) = (\exp\ V(LOS^{i},SE)) / \sum_{j=1}^{J} \exp V(LOS^{j},SE) \quad ,$$

where $i = 1,...,J$ indexes alternatives;

LOS^{j} = observed level-of-service variables for alternative j ;

SE = observed socioeconomic variables;

$LOS = (LOS^{1},...,LOS^{J})$;

$V(LOS^{j},SE)$ = the mean utility of alternative j ;

$P(i \mid LOS,SE)$ = the choice probability for alternative i .

The MNL model can be derived from the theory of individual choice behavior by assuming that individual utility deviations from mean utility in a homogeneous market segment are statistically independent for different alternatives, and have a probability distribution called the extreme value distribution.[1]

The MNL model, viewed as a functional form for mode shares, but not necessarily a behavioral relationship, has been widely used in transportation planning. The two-alternative case yields the logistic curve graphed in Figure 1, which is widely used in aggregate mode-split modeling.

---

[1]Two random deviations are statistically independent if they are uncorrelated and the probability that the first has values exceeding some level does not depend on the value of the second. A random deviation Y is extreme value distributed if $\text{Prob}[\,Y \le y\,] = \exp(-\exp(-y))$. The extreme value distribution is bell-shaped like the familiar Normal distribution in statistics, but is skewed, with a right tail that is thicker than the left tail. The mean of the extreme value distribution is .57722 . Its mode, median, and standard deviation are zero, $-\log(\log 2)$, and $\pi\sqrt{6}$, respectively. Further discussion of this distribution and a demonstration that it leads to the MNL form is given in McFadden (1973) and Domencich and McFadden (1975), Ch. 4.

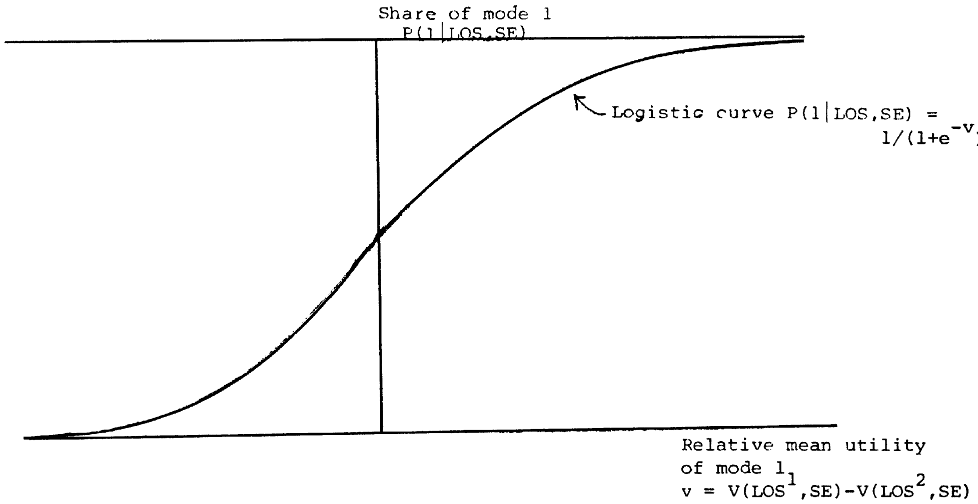

FIGURE 1

Consider a conventional singly-constrained aggregate gravity model for distribution,

(11) $$N_{kj} = O_k A_j / T_{kj}^h \quad ,$$

where N_{kj} = number of trips from zone k to zone j ;

A_j = attraction of zone j ;

T_{kj} = impedance between k and j ;

O_k = scale factor to equate trips distributed from zone k to trips originating in zone k .

Then, the share of trips from zone k to zone i satisfies

(12)
$$P(i \mid T_{k1},\ldots,T_{kj},A_j) = \frac{A_i / T_{ki}^h}{\sum_{j=1}^{J} A_j / T_{kj}^h} .$$

This is the multinomial logit function form in equation (10) with mean utility $V(LOS^i, SE) = \log A_i - h \log T_{ki}$. Thus, the multinomial logit model is a share model of the form familiar to transportation planners who use aggregate share models to describe trip generation and distribution. The primary differences between these traditional models and the formulation of the MNL model in a behavioral disaggregate context are:

1. The structure of the mean utility function $V(LOS^i,SE)$ in equation (10) is based on economic and psychological regularities in individual behavior, and will have a similar form in models of different aspects of transportation choice such as generation, scheduling, distribution, and mode-split.

2. The calibration and utilization of the model are carried through at the disaggregate level for homogeneous market segments, rather than applied to aggregate data.

Because of its simplicity and computational practicality, the multinomial logit (MNL) model has formed the basis for almost all the empirical disaggregate travel demand models studied to date, and has been the basic model used in the studies in this volume. Properties and limitations of the MNL model are discussed in Part IV, Chapter 1, and empirical tests for its validity are developed. Alternatives to the MNL model are discussed in Part IV, Chapter 2.

The Structure of Utility

The mean utility appearing in the multinomial logit formula for choice probabilities is usually assumed to have the form

$$\begin{bmatrix}\text{Mean utility of}\\ \text{an alternative}\end{bmatrix} = \begin{bmatrix}\text{Variable}\\ 1\end{bmatrix} \times \begin{bmatrix}\text{Coefficient}\\ 1\end{bmatrix} + \ldots + \begin{bmatrix}\text{Variable}\\ \text{K}\end{bmatrix} \times \begin{bmatrix}\text{Coefficient}\\ \text{K}\end{bmatrix} \quad , \tag{13}$$

where each variable is some transformation of observed socioeconomic and level-of-service data for the alternative. The coefficients are interpreted as weights giving the importance of each variable in determining an alternative's utility. Level-of-service attributes such as on-vehicle time and walk time are often entered directly as variables in the mean utility function. Travel cost divided by the wage rate of the traveler (or, in some applications, his income) is often introduced as a variable in utility.

Socioeconomic variables that influence the mean utility of every alternative in the same way have no influence on choice probabilities, as they change both the numerator and the denominator of the multinomial logit formula in equation (10) by a common factor that cancels out. Hence, there is interest only in those socioeconomic variables that interact with level-of-service variables to affect the mean utility of different alternatives differently. Travel cost divided by wage is one example of interaction. A second example is a variable that takes the value one for an alternative if this alternative requires driving a vehicle and the individual is able to drive, and is zero otherwise. (The variable in this example may be viewed as the product, or interaction, of a socioeconomic variable that is one if the individual can drive, and zero otherwise, and a level-of-service variable that is one if an alternative requires driving, and zero otherwise.)

An alternative-specific dummy variable is one for a particular alternative and zero for all other alternatives. Mean utility may include alternative-specific dummy variables, both alone and in interaction with other variables. The coefficient of an alternative-specific dummy variable in the mean utility function can be interpreted as reflecting those impacts of an alternative’s unmeasured level-of-service attributes that are not captured in the remaining variables.

Any variable that is the result of interaction between an alternative-specific dummy variable and a level-of-service or socioeconomic variable is termed an alternative-specific dummy. An example of an alternative-specific variable is one that gives the value of on-vehicle travel time for a transit alternative, and which has the value zero for all other alternatives. The coefficient of this variable, compared with the values of coefficients of other alternative-specific travel times, reflects the impact of specific attributes of transit on the onerousness of transit travel time.

A generic or homogeneous-effect variable is one that does not incorporate interactions with alternative-specific dummy variables. An example is a variable that gives on-vehicle travel time for each alternative without identifying the type of vehicle.

Individual utility, written as a function of observed and unobserved variables, should depend only on generic variables. The reason for this is behavioral--individual utility depends on the constellation of physical experiences associated with an alternative, and cannot depend on labels--such as "auto" or "transit"--attached to alternatives by the planner. Mean utility, on the other hand, may depend on alternative-specific variables that mimic the influence of unobserved generic variables. For example, suppose individual utility depends on generic on-vehicle travel time weighted by a generic index of comfort. Suppose the comfort index is unobserved, but varies between alternatives. Then, mean utility for an alternative will have a coefficient of on-vehicle time that reflects the average comfort index on this alternative. It will then appear to the planner that mean utility depends on alternative-specific travel times.

Alternative-specific variables in a multinomial logit model are evidence of failure to observe generic variables that are influencing behavior. A long-run objective of behavioral demand analysis is to improve model specification and data collection to the point where alternative-specific variables are not needed. Models based solely on generic variables are also desirable from the point of view of forecasting. Coefficients of alternative-specific variables do not isolate the behavioral sources of variation across alternatives, or establish that alternative-specific effects will be stable or extendable to new situations when forecasting. In the current state-of-the-art of disaggregate demand analysis, alternative-specific effects may capture the impacts of variables not observed in standard transportation data sets; their omission could bias the importance weights associated with other variables. Consequently, alternative-specific effects appear in most contemporary disaggregate models.

The usual disaggregate assumption that mean utility is linear-in-parameters is not very restrictive, because any smooth utility function can be approximated by a function that is linear-in-parameters (McFadden, 1975b). However, an assumption that level-of-service variables enter the mean utility function directly (or in some particular transformation) requires a behavioral justification.

One method of deriving a concrete form for the mean utility function will now be described. Suppose each individual has a utility function that depends on consumption of goods, leisure, and the amenities offered at various travel destinations, with hours spent at leisure evaluated differently from hours spent traveling. Suppose the alternatives describe possible trip patterns, including destination and mode choice. Assume that one of the alternatives is the no-travel option. Then, choice of an alternative will simultaneously determine generation, distribution, and mode choice for the individual.

Each individual is constrained by a budget: expenditure on goods plus travel cost cannot exceed income earned from labor plus other income. Time must be allocated between leisure, labor, and travel. Taking his budget into account, the individual will choose the travel alternative and the allocation of time between leisure and labor that maximize his utility. One can think of carrying out this maximization in two phases. First, choose the best labor-leisure mix for each possible alternative, and calculate the resulting utility levels. Second, choose the alternative that achieves a maximum among these utility levels.

Consider the first phase, in which the labor-leisure mix is determined for each alternative. The individual will weigh the loss in utility from an hour of leisure foregone against the gain in utility from the goods purchased with the wage from an added hour of work. The mix of labor and leisure will be adjusted until the net gain in utility from further adjustment is zero. Define the marginal utility of a variable to be the net increase in utility resulting from an added unit of this variable. For example, the marginal utility of leisure is the increase in utility from an added unit of leisure. At the utility-maximizing mix of labor and leisure, the marginal utility of goods, multiplied by the wage rate, will equal the marginal utility of leisure.

Utility comparisons are unchanged if the utilities of all alternatives are shifted up or down by equal amounts. Therefore, one can assume that all utility levels are shifted so that the utility of the no-travel option equals zero. The variable measuring the amenities offered at various destinations can also be scaled so that the amenity level associated with the no-travel option is zero.

Now consider the i-th travel alternative. In comparison with the no-travel option, this alternative will produce changes in the amounts of goods and leisure consumed, in travel time, in amenities, and in unobserved level-of-service variables. The utility of this alternative, compared with the no-travel option, can be written as a sum of the changes listed above, each weighted by the marginal utility of the corresponding variable.[1] From the individual's budget, one concludes that the net increase in goods consumption equals the net increase in labor hours worked, times the wage rate, less the cost of travel. The net increase in labor hours worked equals the net decrease in leisure less the time spent traveling. In short,

$$\begin{bmatrix}\text{Net increase}\\ \text{in goods}\\ \text{consumption}\end{bmatrix} = \begin{bmatrix}\text{wage}\\ \text{rate}\end{bmatrix} \times \left\{ \begin{bmatrix}\text{net decrease}\\ \text{in leisure}\end{bmatrix} - \begin{bmatrix}\text{travel}\\ \text{time}\end{bmatrix} \right\} - \begin{bmatrix}\text{travel}\\ \text{cost}\end{bmatrix} \quad . \tag{14}$$

The utility of the i-th alternative can then be written

$$\begin{aligned}\begin{bmatrix}\text{Utility}\\ \text{of i-th}\\ \text{alternative}\end{bmatrix} = {} & \begin{bmatrix}\text{Net increase}\\ \text{in goods}\\ \text{consumption}\end{bmatrix} \times \begin{bmatrix}\text{Marginal}\\ \text{utility}\\ \text{of goods}\end{bmatrix} \\ & + \begin{bmatrix}\text{Net increase}\\ \text{in leisure}\end{bmatrix} \times \begin{bmatrix}\text{Marginal utility}\\ \text{of leisure}\end{bmatrix} \\ & + \begin{bmatrix}\text{Travel}\\ \text{time}\end{bmatrix} \times \begin{bmatrix}\text{Marginal}\\ \text{utility of}\\ \text{travel time}\end{bmatrix} + \begin{bmatrix}\text{Amenities}\end{bmatrix} \times \begin{bmatrix}\text{Marginal}\\ \text{utility of}\\ \text{amenities}\end{bmatrix} \\ & + \begin{bmatrix}\text{Unobserved}\\ \text{attributes}\end{bmatrix} \times \begin{bmatrix}\text{Marginal utility}\\ \text{of unobserved}\\ \text{attributes}\end{bmatrix} \quad . \end{aligned} \tag{15}$$

---

[1] In the mathematical restatement of these arguments at the end of the section, this conclusion is stated more precisely and is shown to follow from the mean value theorem of calculus.

Substitute equation (14) and the relationship[1]

$$(16) \qquad \begin{bmatrix}\text{Marginal utility}\\ \text{of goods}\end{bmatrix} = \begin{bmatrix}\text{Marginal utility}\\ \text{of leisure}\end{bmatrix} \Big/ \begin{bmatrix}\text{Wage}\end{bmatrix}$$

into equation (15). After simplifications, one obtains

$$(17) \qquad \begin{bmatrix}\text{Utility}\\ \text{of i-th}\\ \text{alternative}\end{bmatrix} = -\left\{\begin{bmatrix}\text{Travel}\\ \text{time}\end{bmatrix} + \frac{[\text{Travel cost}]}{[\text{Wage}]}\right\} \times \begin{bmatrix}\text{Marginal}\\ \text{utility}\\ \text{of leisure}\end{bmatrix}$$

$$+ \begin{bmatrix}\text{Travel}\\ \text{time}\end{bmatrix} \times \begin{bmatrix}\text{Marginal}\\ \text{utility of}\\ \text{travel time}\end{bmatrix} + [\text{Amenities}] \times \begin{bmatrix}\text{Marginal}\\ \text{utility of}\\ \text{amenities}\end{bmatrix}$$

$$+ \begin{bmatrix}\text{Unobserved}\\ \text{attributes}\end{bmatrix} \times \begin{bmatrix}\text{Marginal utility of}\\ \text{unobserved attributes}\end{bmatrix} \quad .$$

If the terms in equation (17) involving unobserved attributes can be assumed to have appropriate probability distributions, then this utility function will lead to a multinomial logit model for the choice probabilities in a homogeneous market segment (see equation (10)). The mean utilities in this model satisfy

$$(18) \qquad \begin{bmatrix}\text{Mean utility}\\ \text{of the i-th}\\ \text{alternative}\end{bmatrix} = -b_T \times [\text{Travel time}] - b_C \times \frac{[\text{Travel cost}]}{[\text{Wage}]}$$

$$+ b_A \times [\text{Amenities}] \quad ,$$

where b_T , b_C , and b_A are parameters.

---

[1]Equation (16) is only an approximation when travel times and costs are large.

The preceding argument provides a justification--from the economic theory of utility maximizing behavior--for the entry of travel time and travel cost divided by wage as linear variables in the mean utility function. Generalizations of this model are possible in several directions.

Time, cost, and other attributes of alternatives may have sub-components. Time, for example, can be partitioned into on-vehicle time under congested or non-congested conditions, walk time, and wait time. Costs can be divided into overhead, indirectly charged per-trip costs such as fuel and maintenance, and daily out-of-pocket costs such as tolls. These components can be given separate coefficients in equation (18); the relative weights of components can then be determined as part of the calibration of the model, which is preferable to assigning traditional weights.

The coefficients b_T , b_C , and b_A may depend on observed socioeconomic variables. For example, the weight b_T associated with the walk time component of travel time may be a function of an individual's age and health status, or of those neighborhood characteristics correlated with safety. If this association is expressed in a linear-in-parameters form, then the mean utility function (18) is linear in these parameters, and the calibrated model will describe both the importance weight attached to walk time and the variation of this weight with socioeconomic factors.

The formulation of behavioral disaggregate models has been described for joint frequency, destination, and mode decisions. In practice, limited policy objectives, data, and budgets often make it desirable to analyze only some aspects of travel behavior, such as mode choice, while averaging-out or holding constant other aspects. This can be done sensibly, provided the contributions to utility from different aspects of demand can be disentangled and studied separately. The concept of a separable utility function is important in economic demand theory; it allows study of a particular aspect of demand in some markets in isolation from others. The concept has received extensive theoretical and empirical study. In terms of travel demand, separability requires that the relative weighting of attributes of one aspect of demand, such as mode choice, not depend on or vary with the attributes associated with other aspects of demand. Given such independence, it is possible to calibrate a model of mode choice separately from the analysis of other aspects of behavior. It is also possible to summarize the impact of the mode-choice decision in an "inclusive cost" measure, simplifying the analysis of other aspects of travel behavior. The concept of separability and its use in the construction of behavioral demand models is discussed in depth in

Domencich and McFadden (1975), Ben-Akiva (1973), McFadden (1975b), and McFadden (1976e) .

The argument from the theory of economic behavior that level-of-service variables may enter mean utility in a specific linear way can be restated in mathematical symbols. The individual's utility function is

$$(19) \qquad u_i = U(G, L, T_i, A_i, ULOS^i, SE, USE) \quad ,$$

where G = goods ;

L = leisure ;

T_i = travel time in alternative i ;

A_i = amenities offered in alternative i ;

$ULOS^i$ = unobserved level of service variables for alternative i ;

USE = unobserved socioeconomic characteristics ;

SE = observed socioeconomic characteristics ;

and i indexes trip patterns including destination and mode choice, with $i = 1$ corresponding to the no-travel option. Assume T_1 , A_1 , and $ULOS^1$ equal to zero. Then, choice of i will simultaneously determine generation, distribution, and mode choice for the individual.

The individual's budget, requiring that the total purchase of goods not exceed non-wage income plus wage income, less travel cost, is

$$(20) \qquad G = I + w(L^* - L - T_i) - C_i \quad ,$$

where I is non-wage income; w is the wage rate; and C_i is travel cost for i . (I, w, and C_i are measured in "real" terms; i.e., in units of goods.) L^* is the individual's total endowment of time, and $L^* - L - T_i$ is the number of hours devoted to work. For any i , the quantity of leisure is adjusted to maximize utility, the income constraint being considered during the adjustment,

$$\max_{L} \; U(I + w(L^{*} - L - T_i) - C_i, L, T_i, SE, USE, ULOS^{i}) \quad .$$

The calculus condition for this maximum is

$$(21) \qquad -w \frac{\partial U}{\partial G} + \frac{\partial U}{\partial L} = 0 \quad .$$

Applying the mean value theorem of calculus to the difference of utilities for alternatives i and 1 ,

$$(22) \qquad \begin{aligned} u_i &= \frac{\partial U}{\partial G}(G_i - G_1) + \frac{\partial U}{\partial L}(L_i - L_1) + \frac{\partial U}{\partial T}(T_i - 0) \\ &\quad + \frac{\partial U}{\partial A}(A_i - 0) + \frac{\partial U}{\partial ULOS}(ULOS^{i} - 0) \\ &= \frac{\partial U}{\partial L}\left\{\frac{1}{w}\left[-wL_i - wT_i - C_i + wL_1\right] + L_i - L_1\right\} + \frac{\partial U}{\partial T}T_i + \frac{\partial U}{\partial A}A_i \\ &\quad + \frac{\partial U}{\partial ULOS}ULOS^{i} + \text{higher order terms} \\ &= -b_C\left(T_i - \frac{C_i}{w}\right) - (b_T + b_C)T_i + b_A A_i + b_U ULOS^{i} \\ &\quad + \text{higher order terms} \quad , \end{aligned}$$

where $b_C = \partial U/\partial L$, $b_C + b_T = -\partial U/\partial T$, $b_A = \partial U/\partial A$, and $b_U = \partial U/\partial ULOS^i$, all evaluated at arguments between alternative i and alternative 1 and the higher order terms are approximately zero when travel times and costs represent relatively small fractions of the total time and income budgets of the consumer. Thus, utility can be expressed in the final form

(23) $$u_i = b_T T_i - b_C \frac{C_i}{w} + b_A A_i + \varepsilon_i \quad ,$$

where ε_i summarizes the effects of unobserved socioeconomic and level-of-service variables. If the terms ε_i are assumed to have independent, identical extreme value distributions, as in the previous argument leading to the multinomial logit model, then a concrete model of joint generation, distribution, and mode choice results,

(24) $$P_i = e^{-b_T T_i - b_C \frac{C_i}{w} + b_A A_i} \Big/ \sum_{j=1}^{J} e^{-b_T T_j - b_C \frac{C_i}{w} + b_A A_j} \quad .$$

Statistical Calibration of the MNL Model

The unknown parameters of a concrete behavioral disaggregate model can be estimated using transportation survey data on individual travel decisions or on the behavior of homogeneous market segments. Let $n = 1, \ldots, N$ index individuals or homogeneous market segments, and $j = 1, \ldots, J$ index alternatives.[1] Let S_{in} denote the number of choices of alternative i by individuals in a homogeneous market segment n; and let R_n denote the total number of individuals in segment n. For a single individual, S_{in} is one for the chosen alternative and zero for all other alternatives, and R_n is one. A concrete disaggregate model defines a choice probability for alternative i and market segment n, $P_{in} = P(i \mid LOS_n, SE_n, b)$, expressed as a function of level-of-service variables, LOS_n; socioeconomic variables, SE_n; and a vector of unknown parameters, b.

A statistical method known as maximum likelihood estimation is used in this study to calibrate the unknown parameter vector b. A computer program applying the maximum likelihood estimation method to the multinomial logit mode, QUAIL, was developed for this study, and is available in versions suitable for CDC and IBM machines.

A family of estimation methods that includes the maximum likelihood estimator can be defined by considering the system of equations

$$\text{(25)} \qquad \sum_{n=1}^{N} \sum_{i=1}^{J} W_{in}(S_{in} - R_n P_{in}) \frac{\partial P_{in}}{\partial b_k} = 0 \quad ,$$

where $k = 1, \ldots, K$ indexes the components of b, and W_{in} is a nonnegative weight that may depend on the choice probabilities. Consider solutions of this system of non-linear equations for the unknown parameters $b_1, \ldots, b_K$; an iterative computer algorithm is normally required to obtain a solution. One can show under very general conditions that solutions of this system are "consistent" estimates of the true values of the unknown parameters in the statistical sense that as the sample size grows, the probability that the estimates are more than a small distance away from the true parameters approaches zero.

[1] In general, the number of alternatives vary from case to case, depending upon availability. However, to avoid complex notation, assume the same number of alternatives for each case.

If the weights in (25) are $W_{in} = 1/P_{in}$, then the parameter vector b , which solves (25), also maximizes the function

$$(26) \qquad \mathcal{L} = \sum_{n=1}^{N} \sum_{i=1}^{J} S_{in} \log P_{in} \quad .$$

This is the <u>log likelihood function</u> for the sample, and solution of (25) with the weights $W_{in} = 1/P_{in}$ is the maximum likelihood estimation method. This method of disaggregate calibration yields the most precise estimates obtainable for the parameter vector b in very large samples.[1]

Statistical properties of maximum likelihood estimates of the MNL model are discussed in greater detail in McFadden (1973) and Manski and McFadden (1977) .

---

[1]The weights $W_{in} = 1/R_n$ yield a second useful method, <u>non-linear least squares</u>, for which the solution to (25) minimizes

$$\sum_{n=1}^{N} \sum_{i=1}^{J} (S_{in} - R_n P_{in})^2 / R_n \quad .$$

This method is not as efficient as maximum likelihood estimation in the sense of precision in large samples, but is more robust in the sense that errors in data or specification have less impact on the estimates.

An alternative approach to calibrating disaggregate multinomial logit models, available when splits among alternatives are observed for homogeneous market segments, or where independent variables can be grouped to yield relatively homogeneous market segments, is to estimate by ordinary least squares the equation

$$\log \frac{S_{in}}{S_{1n}} = \beta_1 (z_{in1} - z_{1nl}) + \dots + \beta_K (z_{ink} - z_{1nk}) + \varepsilon \quad .$$

This formula is derived from equation (10) by approximating the left-hand side of the formula

$$\log \frac{P_{in}}{P_{1n}} = \beta_1 (z_{in1} - z_{1n1}) + \dots + \beta_K (z_{inK} - z_{1nK}) \quad ,$$

where $\beta_1, \dots, \beta_K$ are unknown parameters and $z_{inl}, \dots, z_{inK}$ are mathematical functions of the observed variables LOS^i and SE . This procedure, called the Berkson-Theil method, has considerable computational advantages over maximum likelihood or non-linear least squares, and seems to yield equally satisfactory estimates. Thus, its use is recommended when share data on homogeneous market segments is available.

Model Evaluation and Validation

The transportation analyst usually has a number of alternative model specifications he considers *a priori* plausible, and wishes to determine empirically which alternative best fits the data. This calls for statistics that measure goodness-of-fit, and procedures that allow tests of hypothesized specifications.

General goodness-of-fit measures for discrete choice models that are now widely used are the log likelihood function, the likelihood ratio index, a multiple correlation coefficient, and a prediction success index.

An overall criterion for goodness-of-fit, one appropriate for models estimated by the maximum likelihood method, is the log likelihood function evaluated at the estimated parameters; in symbols,

(27) $$L = \sum_{n=1}^{N} \sum_{i=1}^{J} S_{in} \log P(i \mid LOS^{n}, SE^{n}, \hat{b}) \quad ,$$

where $\hat{b}$ denotes the estimated parameters. This function can be used in statistical "likelihood ratio" tests of the importance of particular sets of variables or parameter restrictions. Let L denote the log likelihood of a choice model estimated without restraints, and L_0 denote the log likelihood of this model estimated subject to M linear restrictions. Then, under the null hypothesis that the true parameters satisfy the linear restrictions, $2(L - L_0)$ is in asymptotically large samples distributed chi-square with M degrees of freedom. This statistic can then be used to carry out large sample tests of hypothesis.

The likelihood ratio index is defined by the formula

(28) $$\rho^2 = 1 - L / L_0 \ ,$$

where

(29) $$L = \sum_{n=1}^{N} \sum_{i=1}^{J} S_{in} \log P(i \mid z_n, \theta)$$

is the log likelihood function, with the S_{in} equal to one if i is chosen, zero otherwise;

$$(30) \qquad L_0 = \sum_{n=1}^{N} \sum_{i=1}^{J} S_{in} \log Q \quad ,$$

and Q_i equals the sample aggregate share of alternative i .

When the disaggregate model parameters are estimated by non-linear least squares, an appropriate goodness-of-fit measure is the sum of squared residuals,

$$(31) \qquad SS = \sum_{n=1}^{N} \sum_{i=1}^{J} (S - R_n P_{in}(\hat{\theta}))^2 / R_n \quad .$$

A transformation of this statistic yields a <u>multiple correlation coefficient</u> of the form familiar from regression analysis,

$$(32) \qquad R^2 = 1 - \frac{SS}{SS_0} \quad ,$$

where

$$(33) \qquad SS_0 = \sum_{n=1}^{N} \sum_{i=1}^{J} (S_{in} - R_n Q_i)^2 \quad ,$$

with Q_i the sample aggregate share of mode i as before.[1]

A third method of assessing the fit of a calibrated model is to examine the proportion of successful predictions, by alternative and overall. A <u>success table</u> can be defined as illustrated in Table 1, with the entry N_{ij} in row i and

---

[1]While the R^2 index is a more familiar concept to planners who are experienced in ordinary regression analysis, it is not as well-behaved a statistic as the ρ^2 measure for maximum likelihood estimates. Those unfamiliar with the ρ^2 index should be forewarned that its values tend to be considerably lower than those of the R^2 index and should not be judged by the standards for a "good fit" in ordinary regression analysis. For example, values of .2 to .4 for ρ^2 represent an excellent fit.

TABLE 1 A Prediction Success Table

		Predicted Choice 1	2	...	J	Observe d Count	Observed Share
Observed Choice	1	N_{11}	N_{12}		N_{1J}	$N_{1.}$	$N_{1.}/N_{..}$
	2	N_{21}	N_{22}		N_{2J}	$N_{2.}$	$N_{2.}/N_{..}$
	⋮						
	J	N_{J1}	N_{J2}		N_{JJ}	$N_{J.}$	$N_{J.}/N_{..}$
Predicted Count		$N_{.1}$	$N_{.2}$		$N_{.J}$	$N_{..}$	1
Predicted Share		$\frac{N_{.1}}{N_{..}}$	$\frac{N_{.2}}{N_{..}}$		$\frac{N_{.J}}{N_{..}}$	1	
Proportion Successfully Predicted		$\frac{N_{11}}{N_{.1}}$	$\frac{N_{22}}{N_{.2}}$		$\frac{N_{JJ}}{N_{.J}}$	$\frac{N_{11} + \ldots + N_{JJ}}{N_{..}}$	
Success Index		$\frac{N_{11}}{N_{.1}} - \frac{N_{.1}}{N_{..}}$	$\frac{N_{22}}{N_{.2}} - \frac{N_{.2}}{N_{..}}$		$\frac{N_{JJ}}{N_{.J}} - \frac{N_{.J}}{N_{..}}$	$\sum_{i=1}^{J}\left[\frac{N_{ii}}{N_{..}} - \left(\frac{N_{.i}}{N_{..}}\right)^2\right]$	
Proportional Error in Predicted Share		$\frac{N_{.1}-N_{1.}}{N_{..}}$	$\frac{N_{.2}-N_{2.}}{N_{..}}$		$\frac{N_{.J}-N_{J.}}{N_{..}}$		

column j giving the number of individuals who are predicted to choose i and observed to choose j .[1] Column sums give predicted shares for the sample; row sums give observed shares. The proportion of alternatives successfully predicted, $N_{ii} / N_{.i}$, indicates that fraction of individuals expected to choose an alternative who do in fact choose that alternative. An overall proportion successfully predicted, $(N_{11} + ... + N_{JJ}) / N_{..}$, can also be calculated.

Because the proportion successfully predicted for an alternative varies with the aggregate share of that alternative, a better measure of goodness of fit is the <u>prediction success index</u>,

$$\sigma_i = \frac{N_{ii}}{N_{.i}} - \frac{N_{.i}}{N_{..}} , \tag{34}$$

where $N_{.i} / N_{..}$ is the proportion that would be successfully predicted if the choice probabilities for each sampled individual were assumed to equal the predicted aggregate shares.[2] This index will usually be nonnegative, with a maximum value of $1 - N_{.i} / N_{..}$. If an index normally lying between zero and one is desired, (34) can be normalized by $1 - N_{.i} / N_{..}$.

An overall prediction success index is

$$\sigma = \sum_{i=1}^{J} [N_{ii} - N_{.i}^2 / N_{..}] / N_{..} = \sum_{i=1}^{J} \left[\frac{N_{ii}}{N_{..}} - \left(\frac{N_{.i}}{N_{..}} \right)^2 \right] . \tag{35}$$

---

[1]The formula for N_{ij} is

$$N_{ij} = \sum_{n=1}^{N} S_{in} P_{jn} .$$

This corresponds to a <u>randomized strategy</u> for classifying individuals, by drawing from the multinomial distribution with the estimated choice probabilities, and gives the <u>expected</u> success table for this strategy. An alternative prediction method is to forecast that the alternative with the highest probability will be chosen.

[2]In a model with alternative-specific dummies and the calibration data set, estimation of parameters imposes the condition $N_{i.} = N_{.i}$. If one predicted the choice probabilities for each individual to equal aggregate shares, then $N_{i.} = N_{..}$ would be the proportion successfully predicted to choose i . This represents a “chance” prediction rate for a model in which no variables other than alternative-specific dummies enter. Thus, σ_i measures the net contribution to prediction success of variables other than the alternative-specific dummies.

Again, this index will usually be nonnegative, with a maximum value of

$1 - \sum_{i=1}^{J} \left(\frac{N_{.i}}{N_{..}} \right)^2$, and can be normalized to have a maximum value of one if desired.

The use of measures of goodness-of-fit to assess model precision and validity will be discussed further in the context of specific applications.