TRAVEL DEMAND FORECASTING: SOME FOUNDATIONS AND A REVIEW

DANIEL BRAND
Department of City and Regional Planning
Harvard University

This paper has 2 overall objectives with regard to travel demand forecasting:

1. To bring together and discuss the rationale for various stands of previous work, and
2. To provide a common point of departure for discussion of improved use of existing methods and of development of research needs.

Hundreds of millions of dollars have been spent on travel forecasting for design and planning of urban and intercity ground transportation systems in the United States alone during the past 20 years. Only a small fraction of that money has been spent specifically for new travel demand model development. Even so, many transportation studies tried in a professional way during that period to make incremental improvements in the methods they inherited.

In the 1940s and 1950s, trip-generation models were developed to predict "generated" traffic on facilities, namely, "traffic created by one or more land uses" (62). Similarly, trip-distribution models were developed to predict shifted traffic, namely, "trips whose desire lines have shifted due to a change in origin and destination" (62). And in the 1960s as substantial new federal money became available for planning transit, modal-split models were developed to predict "diversion" of trips from highways to transit facilities. All these models, applied sequentially, provide input to shortest and multipath route-finding techniques that assign total travel by mode to links at particular locations. The models use as input data aggregate values of zonal population, employment, and link capacity and average values of zonal incomes, car ownership, and interzonal travel times and costs. They are based on aggregate travel definitions that describe what happens to facilities when changes are made to them.

More recently, a different perspective on modeling travel has emerged. This is the perspective that asks, What happens to individuals when changes are made in the transportation system? In 1962 in a university setting, Warner applied this individual-choice perspective to the just-emerging popular subject: transit-usage forecasting. He used disaggregate data to develop the first probabalistic model of individual travel behavior—(binary) modal-choice behavior. Since then, research in, but not application of, the disaggregate approach has been extensive. Generally, its purpose has been to explore the kinds of models and descriptions of travel behavior (e.g., value of time) that result if travel choices are viewed from the new perspective. Travel choices at the individual traveler level can include trip frequency (including the no-trip option), choice of destination, choice of mode, choice of time of day, and choice of route within mode.

More recently, in the 1970s, information is being sought by planning agencies on relative trip peaking at the aggregate level. This corresponds to individual choice of time of day of travel. Transportation agencies seek "to measure the magnitude of peak loads, how long they last, and the extent of accompanying congestion" (84). Descriptive models are being developed that relate travel-peaking percentages to aggregate measures of city size and socioeconomic characteristics (53); they are similar to their precursor, aggregate trip-generation models. This relative trip-peaking modeling corresponds to modeling an individual's choice of time of day of travel, which only recently has been attempted (10).

In the last few years, representing travel demand directly as a function rather than as a fixed quantity has been introduced to travel forecasting from economic demand theory. "Induced travel" as a term describes the change in travel resulting from shifts along a demand curve. The term incorporates the older aggregate descriptive terms of trip generation, trip distribution, and modal split. The first attempt to combine (short-run) travel-choice definitions and behavioral assumptions at an aggregate level was in 1963 (33) when the trip-generation and modal-choice decisions were combined and modeled by using interzonal system data in a direct demand model. The traveler was considered to evaluate simultaneously all the alternative modes available in the Northeast Corridor. Choices were not modeled separately (i.e., sequentially or indirectly). The data were limited to the relatively few intercity zonal pairs in the corridor.

Such a direct demand model was first used for an urban area in 1967 (9). Alternative-route and time-of-day choices were consciously excluded from these early direct-demand models, and the destination choice was modeled without cross relations (i.e., without cross elasticities between destinations). Because the number of choice combinations to be considered and modeled simultaneously is the product of the number of alternatives within each of the previously described sequential choices, the choice environment quickly becomes very complex and difficult to describe in a direct-demand model. Nevertheless, in 1969 a direct demand model was used (54) that explicitly considered alternate destinations for the Northeast Corridor divided into 8 "metrodistricts."

The issue of aggregate versus disaggregate "probability" models permeates the above discussion. Most urban travel forecasting is still carried out "in the field" with the earlier aggregate "choice" models by state highway departments and regional planning agencies with the help of the U.S. Department of Transportation. Research is under way with disaggregate models in several universities and in consulting firms under contract to various agencies of the U.S. Department of Transportation and a few state departments of transportation. The often-used term "disaggregate behavioral" models gives the impression that individual-choice models have a monopoly on incorporating travel behavior. That is clearly unfair, for travel demand models can be derived from behavioral assumptions independently of whether they will use aggregate or disaggregate data.

Choice behavior in disaggregate models must be interpreted as probabalistic. Deterministic choice (i.e., 0, 1 binary) behavior produces uninteresting results when aggregated over all individuals to describe aggregate behavior in a planning application. However, the probability process is assumed to be in static equilibrium (see Appendix) and incorporates no time parameter in a behavioral sense; e.g., learning or experience does not change the probabilities (43). Disaggregate travel models should, therefore, be referred to as probabilistic and not stochastic if they are used with cross-sectional data.

The generally strong arguments for using disaggregate models usually include data efficiency arguments. That is, more information on travel choice situations and behavior is usually available with disaggregate data than with aggregate data. For example, Fleet and Robertson (86) showed that aggregation of trip data to zones reduced the variation in trip-making (trip generation) between observations to only 20 percent of the value at the dwelling unit level. In the process of aggregation, nonlinear relations may also be lost by using averages of explanatory variables. However, disaggregate travel models have not yet demonstrated practical superiority in providing travel information to decision-makers. In fact, we have as yet a way to go in getting models based on individual-choice behavior into the field. [Disaggregate models of some of the conventional UTP steps (i.e., trip generation) will be easy to introduce "in the field" (31).]

However, there is little doubt that the emerging techniques (72) for using travel models based on the behavior of individuals and not the behavior of aggregate numbers of trips will accelerate our understanding of travel-choice behavior. The empirical results of the next few years should greatly improve our understanding of and our ability to base models on behavioral assumptions appropriate to the circumstances under which the modeling is undertaken. In most cases, travel models, whether aggregate or disaggregate, should be based on a well-specified structural or behavioral representation of the decision process. Such models can be disaggregate or aggregate. Models should be avoided that are merely "best fit" curves, for they are impossible to interpret. Also, whether aggregate or disaggregate, the models should be evaluated on the basis of their applicability in a given situation, e.g., ease of use or efficiency in the use of data.

Unfortunately, current travel forecasting procedures fall short of satisfying current demands on their use. The needs and requirements of today's transportation decision-makers for travel information are rapidly changing. The U.S. Department of Transportation noted in its preliminary statement for this conference (76):

> Present passenger travel demand forecasting procedures . . . are most responsive to the issues of the 1950s and early 1960s concerning long-range regional transportation plans and the development of information that was required to design the facilities.
>
> The planning issues of the late 60s and 70s are broader and more numerous. First, they involve a much wider range of alternatives that need to be evaluated. These include highway-transit trade-offs, low and noncapital alternatives such as pricing schemes, new technological systems, and "do-nothing" alternatives. Second, it is now insufficient to evaluate facilities on the issues of capacity and cost alone. Additional measures have become important in the planning process and include levels of service and price. Third, the environmental and social effects of transportation-facility construction and operation must become integrated into the planning process. Fourth, the incidence of travel service, environmental, and social consequences on various groups within the study area must be considered in the evaluation of transportation facilities. Fifth, as a consequence of greater involvement by elected officials and citizens in the planning process, travel forecasts for transportation facilities must be made expeditiously and information must be summarized in a manner that facilitates communication.
>
> Travel forecasts are essential elements in reaching decisions on transportation. To be more responsive to the issues, travel forecasting methodology will have to be modified and improved. Travel forecasting procedures must be quicker and less costly to operate, be sensitive to the wide range of policy issues and alternatives to be considered, and produce information useful to decision-makers in a form that nontechnical people can understand.

In some places, current travel-forecasting models are successfully providing useful information on very short notice. However, such instances normally occur only at large agencies that have several highly trained professionals and large continuing computer budgets. Costs are high not only to continue the operation of current procedures in a given location but importantly also to initially develop and install the methods in a given region. Calibration of existing travel models and procedures takes considerable skill and effort. Until travel demand models are transferable from area to area, very high start-up costs in the form of new data collection, program development, and model calibration will continue to seriously impede the ability of the profession to produce relevant and responsive travel information for decision-makers.

TRAVEL BEHAVIOR

Travel forecasting procedures must have a basis in behavior if planners and decision-makers are to be able to understand and interpret the results of the forecasts. This is true for many reasons. The forecasts that result depend on the behavioral assumptions. Behavioral models are needed for transferability (in space and time) to situations other than those for which the models were developed. Behavioral models are needed also for evaluation, if the (usual) assumption is to be made that the trade-offs between time and money in a travel choice situation are valid for user benefit calculations.

In travel demand forecasting, therefore, we must confront squarely the validity of our theories describing relations among people and their locations on the one hand and travel on the other. This involves consideration in particular of how and in what sequence, if any, people view the origins and destinations of their journeys and the transportation system that connects or potentially connects their origins and destinations.

A travel demand model implements in a purposeful way the understanding that the modeler has of the behavior of the system of interest. A system can be defined as a set of objects and a set of relations among those objects and among their attributes (23). Every time we make or contemplate a decision, the complexity of urban and transportation systems confronts us with a need to make a simplified and intelligible imitation of reality (i.e., a model). This involves abstracting the important parts, to us, of the decision situation that confronts us. Clearly, the set of objects that describe the travel choices confronting travelers is important in travel demand forecasting. Transportation planning concerns itself with making, or contemplating making, changes to the transportation system or changes that will affect that system. Our interest is in describing the behavior of travelers as they respond to travel choices and to changes in travel choices that confront them. The ability to predict the amount and distribution of travel in any situation is, therefore, only as good as our understanding of the underlying perceptions that travelers have of the choices that confront them.

Modeling Choices

There are developing some basic modeling choices based both on explicit statements of alternate understandings of travel-choice perceptions and decisions and on the realization that a travel demand model, like any model, is ultimately a subjective imitation of reality. The basic modeling choices are founded on differing behavioral premises, for ultimately the modeler's view of behavior in the system of interest must be the starting point.

Strategy of Paper

In this paper, certain basic modeling choices will be described at the outset. Where possible, the analytically derivable implications of each modeling choice on appropriate mathematical-structural forms of travel demand models are also described. Finally, the travel demand models that have implemented or might implement the modeling choices are described.

Issues exist when there are unsolved problems or unresolved conflicts over appropriate solutions. This paper was written specifically for a conference dealing with such problems and conflicts. We made the initial presumption in the conference, as in this paper, that issues relating to theory and practice in travel demand forecasting are researchable and in many cases can be made subject to empirical testing. [Causality, unfortunately, cannot be empirically demonstrated, although empirical results can be demonstrated to be inconsistent with certain causal chains (68).]

It may be clear from this review paper that our theory and prior understanding of how travelers perceive their travel-choice environment are weak. This is certainly not a criticism so much as a description of the state of the art of understanding choice behavior in the social sciences in general. Our weakness in understanding is evidenced by the variety of different assumed perceptions of the travel environment on which existing travel demand models can be shown to be based. This paper attempts to organize several of these perceptions into alternate modeling choices, without making strong statements about which choices seem preferable, or more plausible, to the author. All the basic modeling choices are indeed worthy of further research and application and will be shown to be combinable for still additional modeling choices.

The supplier's perspective and concern with describing and evaluating what happens to facilities when changes are made to them may be fairly credited with leading to the earlier aggregate travel forecasting models. Those models respond directly to the question of what happens to flows on transportation facilities when changes are made in the facilities.

The social science (academic) disciplines are more concerned with what happens to individuals and groups of individuals. Thus, it is no surprise that Warner's early work on individual travel-choice models took place in a university setting. [Wilson et al. (83) make the useful distinction between primarily academic disciplines concerned with analysis (i.e., the social sciences, including economics) and the professional disciplines concerned with design and policy-making (i.e., engineering, city planning, and architecture). The latter can plausibly be said to be traditionally concerned with the objects of their design and their use in the aggregate, while the former are concerned with analysis of cities and regions at all levels of (dis)aggregation.] However, the issue of aggregation has been argued to be separable from the issue of travel behavior.

The more fundamental behavioral choice is whether the attributes of travel choices are considered or perceived independently from or together with the objects or facilities that carry or support or propel the traveler. That is, the most basic behavioral modeling choice is whether travel attributes are perceived by themselves or whether they are mapped on particular supply-side choices (e.g., mode and route, or choice of technology). The argument can similarly be extended to attributes of alternative destination choices. These alternate perceptions of the travel environment imply that attributes of the transportation system can be included in travel demand models in 1 of 2 ways: as choice abstract or attribute specific, or as choice-specific attributes.

Particular names for these 2 modeling choices are not yet settled on. Manheim (44) calls the first choice the "hypothesis of commodity-independent utilities." The authors (57) of the best known example of the first type of model, the abstract mode model, have more recently referred to their model as an "attribute-specific" model. This gets away from the needlessly restrictive modal-choice emphasis indicated by their original "abstract-mode" name. In this paper, the terms choice abstract and choice specific are used to describe these 2 basic travel modeling choices.

FOUNDATIONS: BASIC MODELING CHOICES

In general, demand models relate quantities demanded to resources that must be expended to obtain those quantities. In travel demand modeling, the first behavioral question is, Whose resources? Are they the resources of the individual traveler, i.e., his money, and the use of his most basic resource, his time? (In theory, of course, the "behavioral" resources expended are always those of the "demanders.") Or are they the resources of society that provides facilities that "produce" travel, i.e., the aggregate of individual trips on the transportation system? This divergence in viewpoints or "values" has led fundamentally to the development of different kinds of travel-forecasting models. The alternate perceptions of the travel-choice environments resulting from each view provide the most basic (behavioral) modeling choice for travel demand forecasting.

That is, by whom shall the important parts of the transportation system be defined? By the supplier who considers the objects that he is able to provide, and who finds it useful to differentiate among modes, routes (path) within modes, and the locations, sizes, and technical characteristics of the means of producing transportation? Or by the individual traveler who may or may not consider the same description of the hardware of transportation as the supplier? Is there any overlap whatsoever between systems defined from each point of view? Or are the important parts of the system so defined completely disjointed? That is, does the traveler consider only the services provided by the transportation system to the complete exclusion of any identification of the objects (facilities) provided?

CHOICE-ABSTRACT TRAVEL DEMAND MODELS

Attributes

In classical utility analysis, consumers maximize some function of quantities of various commodities that can be consumed (see Appendix). Travel is, of course, a commodity. Depending on how travel is defined, the number of alternate commodities possessing utility that can be consumed is very large (i.e., ultimately all combinations of alternative trip origins, destinations, times of day, modes, and paths).

Utility theory may be modified to base utility on attributes or characteristics of the quantities to be consumed. According to Lancaster (36), "Utility or preference orderings are assumed to rank collections of characteristics and only to rank collections of goods indirectly through the characteristics that they possess.... Furthermore, the same characteristic may be included among the joint outputs of many consumption activities so that goods which are apparently unrelated in certain of their characteristics may be related in others." The traveler is assumed to derive utility, U, from the attributes, Z, consumed and obtained as a result of the transportation activity.

Simultaneous Choice: Abstract-Mode Model

The abstract-mode model (57) is derived consistent with this modification of utility theory. The model provides a striking example of the modelers' perspective on the problem determining the forecasting model that is developed.

The Northeast Corridor project, for which the model was developed, was charged with analyzing and predicting the demand for new transportation services in the corridor. This required that travel forecasts be made for travel modes that might not currently exist (the new-mode problem). Therefore, the introduction of a new mode should not change the demand function (model) derived from a utility function, U = U(Z), estimated on the basis of the attributes, Z, of existing modes by using existing data (see Appendix). Technology or production function equations, $Z_m = g_m(X)$, could indeed be mode specific and describe choice environments having different attribute levels as a function of amount of travel, X. However, travel (demand) choices were to be determined only by the attributes of the choice environment so produced, independent of mode.

In the derivation of this choice-abstract, or attribute-specific, demand model, the concept of attributes is used "to define a mode in terms of the type of service it provides to the traveler and not in terms of the administrative entity that controls its operations or the sort of physical equipment it employs" (57). However, the derivation of the model did not proceed analytically from consideration of personal utility. The modification of utility theory was (only) relied on to justify characterizing modes "by the values of the several variables that affect the desirability of the mode's service to the public: speed, frequency of service, comfort and cost" (57).

The estimated travel-forecasting equations are, therefore, not mode specific but mode-attribute specific. They take the following form:

$$V_{klm} = \phi_0 (P_k P_l)^{\phi_1} (Y_k Y_l)^{\phi_2} t_{klb}^{\theta_1} \left(\frac{t_{klm}}{t_{klb}}\right)^{\theta_2} c_{klb}^{\theta_3} \left(\frac{c_{klm}}{c_{klb}}\right)^{\theta_4} \tag{1}$$

where

V_{klm} = volume between k and l by mode m,
P_k = population in zone k,
Y_k = median income in zone k,
t_{klm}, c_{klm} = travel time and (money) cost between k and l by mode m,
t_{klb} = travel time by fastest mode,
c_{klb} = cost by cheapest mode (not necessarily same as fastest mode!), and
ϕ, θ = parameters of the model.

This is a simplified statement of the model. Separate parameters for each variable can be added, and the variable list can be extended to include others such as frequency of service and employment. Note, however, that the model has only one set of parameters regardless of the subject mode, m, for which travel is being predicted. Thus, the equations are mode-attribute specific and not mode specific. The introduction of a new mode, if not the best mode in any attribute (and not the subject mode), does not change the travel prediction for the subject mode.

Particular assumptions are made about the perceived interaction of modal attributes in determining travel demand. For example, there are cross elasticities (cross relation) only with respect to the best competing mode in any attribute. These are equal in magnitude to the direct elasticities for the subject mode.

Young (85) changed the representation of the competing modal attributes in Eq. 1 from only the best values among all the modal choices to weighted averages of the attribute values of the competing modes. That is (58),

$$T_{ijk} = a_0 X_{ij}^{a_1} \frac{F_{ijk}}{\sum_k F_{ijk}} \tag{2}$$

where

i = origin,
j = destination,
k = mode,
a = constants,
T = travel volume,
X_{ij} = exogenous economic and demographic variables,
$F_{ijk} = D_{ijk} C_{ijk}^{a_2} H_{ijk}^{a_3}$,
D_{ijk} = number of trips by mode k,
C_{ijk} = cost (money) on mode k, and
H_{ijk} = journey time on mode k.

Consistency with the independence axiom (see next section) is obtained if the D's are removed from the product term for F and made a separate relative frequency term in Eq. 2. That is,

$$T_{ijk} = a_0 X_{ij}^{a_1} \left(\frac{D_{ijk}}{D_{ij}}\right) \frac{F_{ijk}}{\sum_k F_{ijk}} \tag{3}$$

Practical difficulties must be noted in completely reducing travel-related (dis)utility to mode-independent attributes. These difficulties can include quantifying the time and space restrictions from car-pooling or transit travel, as contrasted with automobile-driver travel (not to mention quantifying the comfort and privacy differences) and between transit mode combinations as represented by its several access modes (walk, park-ride, kiss-ride, feeder bus). To the extent that such differences, as they affect travel-choice behavior, can be subsumed in door-to-door travel times, departure frequencies, and fares, the abstract-mode model can be considered applicable. However, if the list of attributes that must be quantified to adequately describe travel alternatives in terms only of the perceived levels of attributes becomes extensive, the alternative modeling choice of identifying the attributes together with the modes may be more practical as a strategy. However, the new-mode problem (if relevant) must then be faced.

These 2 models, Eqs. 1 and 2, are examples of choice-abstract direct demand models, which assume that the traveler considers all the attributes of alternative travel choices simultaneously when making a travel decision. The result is a simultaneous-choice or direct demand model.

However, there is a choice-abstract modeling alternative. That is an assumption of nonsimultaneous, or sequential consideration of, system-independent or choice-abstract attributes.

Sequential Consideration of Attributes

An important alternative modeling choice is to formulate travel behavior models that are not based on the simultaneous consideration of values of attributes across all alternatives. Probability mechanisms can be proposed based on the individual's attending to different aspects of the choice situation at different times. One proposal, (75) based on earlier work by Marschak, is the notion of eliminating alternatives in a multiple-choice situation by successively considering single aspects (attributes) of the choice situation. Each successive choice is governed by one aspect selected from those included in the available alternatives "with probability proportional to its weight" (75). All alternatives are eliminated that do not include the selected aspect, and the process continues until only one choice remains. Aspects that are common to all the alternatives do not affect the choices made. Obviously, the way aspects are defined is critical. The theory might be extended to include groups of aspects (factors) not easily described by a single measure.

A scenario of the elimination-by-aspects method of modeling travel-choice behavior might be as follows:

The most important aspect results from the trip purpose. For example, for shopping trips, only destinations containing the aspect, retail stores, are considered as alternative destinations. A more precise definition of the shopping purpose (e.g., shopping goods as opposed to convenience goods) serves to delimit further the allowable alternative destinations. The next most important aspect (following the findings of Hille and Martin, 27) is "reliability of destination achievement." Unsafe and unreliable modes and routes are eliminated. This will generally not eliminate many alternatives in U.S. urban areas because, through nonuse, most unsafe travel alternatives have been eliminated as economically nonviable. However, because random elements might be allowed, some alternatives for some individuals may be eliminated because they did not meet some stated safety threshold. The next most important aspect, comfort, with emphasis on flexibility and ease of departure, is used to eliminate the transit mode for all travelers from all origins to all destinations not near a transit line. The automobile mode is eliminated for travelers with no car (or car pool) available. The possibility of a trip is eliminated if no car is available, no transit is available to the "available" destination alternatives, and walking distance is too far to all of the available destinations not yet eliminated through the purpose and reliability aspects. Again, random elements allow this to be a probability model of choice. Other aspects of travel time are considered next, then cost, and so on, according to the sequence of importance in, for example, the Hille and Martin (27) findings.

Summary

A diagram may be useful in summarizing the travel demand modeling choices described thus far (Fig. 1). The lowest level of the hierarchy is not the result of choice forks but rather contains examples of models that have implemented or might implement the modeling choices. The elimination-by-aspects method of Tversky is not likely to be the only possible model structure that implements travel behavior that considers choice-abstract attributes sequentially.

CHOICE-SPECIFIC TRAVEL DEMAND MODELS

The alternate assumption about how travelers perceive their choice environment is that the attributes of the travel-choice environment are not perceived or at least modeled independently of the objects provided, i.e., the facilities that constitute the transportation system. This modeling choice, as before, breaks down into the behavioral modeling subchoices of (a) simultaneous consideration of all the attributes and (b) sequential consideration of the attributes.

The distinction between direct and indirect demand models has already been made. In the former, all attributes of an entire trip are assumed to be known and considered simultaneously by the traveler. As shown in Figure 2, this behavior can be described as involving the simultaneous consideration of all the attributes normally associated with each of the 5 conventional descriptors of travel: frequency, time of day, destination, mode, and path. If each path through the travel decision tree is considered an alternative travel choice whose attributes are considered simultaneously "in competition" with the attributes of all the other travel choices, the models can become very complex. The number of choice combinations to be considered and modeled simultaneously is the product of the number of alternatives within each of the travel choices. For example, a simultaneous model of travel that considers 3 modes, 2 times of day, 20 destinations, and 1 path requires the modeling of ($3 \times 2 \times 20 \times 1$) or 120 travel choices for each origin. [This number may be reduced by eliminating zero-probability choices in calibrating models that satisfy the independence axiom (see next section).] The number of explanatory variables and the allowable interactions among variables that may be assumed to explain (model) simultaneous travel behavior can multiply very rapidly for realistic travel-choice situations in urban areas.

The need for "simple robust models" has been well articulated (2). Calibrating models for large numbers of alternatives (choices) with very low probabilities of choice is difficult in the extreme. Attributing properly the separate effects of large numbers of (possibly highly correlated) attributes describing complex choice environments (where calibration techniques often require certain assumptions, e.g., normality or homoscedasticity) boggles the mind. (One may speculate that the "number of variables required to predict probability of choice is finite and rapidly approaches the limit of human discrimination.") For these reasons, travel demand models must be reduced in complexity in some plausible way.

Restricting the choices available restricts the products or attributes the traveler is assumed to evaluate in making his travel decision. Restricting the choices that are presumed available to the traveler appears to be the way in which choice-specific travel demand models can be reduced in complexity. However, this involves making some important assumptions on the separability and the sequence of travel choices.

The assumption that travelers behave as though they sequentially consider (travel) choice-specific attributes (Fig. 2) means that there is a hierarchy of travel decisions in which certain travel decisions are made independently (separately) of others. In turn, other travel choices (e.g., higher level choices like destination, Fig. 2) are made given that lower level choices (e.g., mode) are predetermined.

There are 2 ways to model such sequential travel behavior. The first assumes that the relative valuation of choice attributes is constant throughout the set of travel choices. This requires that models of the independently made lower level travel decisions be calibrated based only on a subset of attributes describing those choices. The estimated (and preserved) utilities from the lower level choices are then added to a set of attributes on the basis of which the higher level choices are made. The traveler, it is assumed, makes some sequence of choices, and the earlier choices are based on independent and separate evaluations of personal utility (separate) from the "later" conditional or "constrained" choices. For example, the time of day (shopping purpose) choice was modeled (10) on the assumption that "there is a utility associated with the trip itself which is additive to the utility or disutility associated with the choice of time of day, which is additive with the utility associated with the place to which the trip is made. . . ."

Thus, the choice of mode is modeled separately and prior to the destination choice and is assumed to be independent of the overall number of trips between the origin and destination. Similarly, the choice of time of day is assumed to be made independently of the choice of destination.

The attributes that are assumed additive must map on the (sequential) choices. Otherwise, a choice-abstract model results. If difficulty is encountered, either the travel choices can be redefined or the supply side description of choices (e.g., mode) can be abandoned and sequential choice-abstract models can be developed.

The assumption of sequential travel choices, given that travelers perceive their choices as described by attributes inseparable from choices, is a difficult assumption to make. Yet it is an attractive strategy for reducing the complexity of travel demand models because it greatly reduces the number of interaction terms in the model. The other strategy is to reduce the number of independent variables that are assumed to influence travel behavior. That is, reduce the number of attributes the traveler is assumed to evaluate in his travel decision-making process without excluding interaction. Because the attributes that the traveler evaluates are identified with particular travel choices, this second strategy for reducing model complexity is more appropriate to choice-abstract models than to choice-specific travel demand models.

A second way to model sequential travel behavior requires the still stronger (more difficult) assumption that some travel choices are made completely independently of other travel choices and that the relative valuation of choice attributes common to 2 or more travel choices is not necessarily the same in successive travel choices. This represents a third-level assumption regarding the consideration and valuation of the attributes (i.e., the relative marginal utilities) of the choice situation confronting the traveler. These 3 levels of assumptions are summarized in order from the weakest to the strongest (or most heroic) assumption.

1. All the attributes of the choice situation confronting the traveler are considered simultaneously. The complete trip is one decision. The relative valuation of the attributes is constant in any travel choice in the hierarchy shown in Figure 2.
2. There is a hierarchy of travel decisions in which certain travel decisions are made independently of other decisions. However, the relative valuation of choice attributes is constant in any complete travel decision (i.e., any single path through the travel decision tree shown in Fig. 2).
3. As in assumption 2, there is a hierarchy of travel decisions in which certain travel decisions are made independently of other decisions. However, the relative valuation of choice attributes common to 2 or more travel choices is not necessarily the same in successive travel choices.

The first assumption is the easiest to make. It requires the concomitant assumption of constant relative valuation of attributes in component travel choices of a complete travel decision.

The second (strict utility) assumption is made for ease of estimation (reducing the number of variables in the models to be estimated relative to the first and third assumptions). It requires some sequence of travel choices to be assumed for purposes of estimation as discussed above. Inclusive prices must be used to preserve the previously estimated utilities in strict utility models. The separately calibrated models using inclusive prices may be combined and applied simultaneously, or sequentially in any order.

The third assumption is the present assumption of UTP models that completely and independently estimate the different travel choices with different valuations of the independent variables in each model. The traveler, nevertheless, must face the same values of the independent variables in more than one component travel choice. For example, "the costs of the various modes influence not only the choice of mode but also the selection of destination and the determination of whether the trip should be made at all" (14). The most damaging indictment of the third assumption is that the sequence of application of the models determines the results. That is, no unique equilibrium can be reached with these models so long as flow and congestion conditions and the resulting travel costs change in any way from those used to calibrate the models. That is, even if the conventional series of models (including trip generation) were system sensitive, the sequence of their application determines the network equilibrium reached after more than one iteration. In addition, of course, the third assumption poses the problem of what appropriate value to place on user benefits (e.g., time savings) in evaluation of transportation system alternatives when different valuations of the independent variables are assumed in each component travel choice.

From the above discussion, the conclusion may be drawn that the assumption is easier to make that travel choices are separable than that travel choices are made in some sequence. This assumption implies only that the marginal rates of substitution (trade-offs) among attribute variables that govern one travel choice do not vary among travel choices. Stated another way, this means that the trade-offs or ratio of "weighted" attributes that explain one travel choice are independent of the other choices.

It is with the last statement that 2 important results from separate disciplines can be joined. In mathematical psychology, this is a statement of separability property of the independence-of-irrelevance-alternatives axiom (41, 42). In economics (utility theory), at the conditions assumed at equilibrium (see Appendix), the ratio of the marginal utilities of 2 choices is equal to the ratio of their "weighted" attributes (i.e., their revealed "prices"). The relative marginal utilities of the attributes of a choice situation can be solved for (inferred from) observed data on the choices made (61).

Thus, the assumption of separable travel choices potentially allows complex travel choices to be broken down into simple travel choices whose relative marginal utilities can be inferred from observed data. However, a sequence assumption is necessary to determine which (separable) travel choice will be "simply" modeled, the inferred relative marginal utilities from which will be preserved in the remaining travel choices. Before the possible plausibility of any sequence and separability assumptions is discussed, the important properties and implications for travel demand modeling of the independence axiom will be described.

Independence-of-Irrelevant-Alternatives Axiom

The independence-of-irrelevant-alternatives condition (41) implies that, for any 2 alternatives i and j having a positive (nonzero) selection probability, the relative odds of choosing j over i in a set containing only the 2 alternatives are equal to the ratio of their probabilities of being selected from any larger set of alternatives containing both i and j. This can be expressed as (48)

$$\frac{P_{ji}}{P_{ij}} = \frac{P(j:A_1)}{P(i:A_1)} \tag{4}$$

where

P_{ji} = probability of selecting j in a 2-element set $A_1 = i, j$;
P_{ij} = probability of selecting i in a 2-element set $A_1 = i, j$;
$P(j:A_1)$ = (nonzero) selection probability of choosing j contained in any set A_1; and
$P(i:A_1)$ = (nonzero) selection probability of choosing i contained in any set A_1.

This condition states that the odds that alternative j will be chosen over i in a set containing both are independent of the presence of irrelevant "third" alternatives in A_1. This is the separability property of the independence-of-irrelevant-alternatives axiom (41, 42).

"Strict utility" is defined by Luce (41) as being the function $h(Z_{ki})$ that satisfies Eq. 4 for the binary case i = 1, 2. That is, the relative odds of choice or share of, say, travel, P_i/P_j, between any 2 alternatives i and j are simply some function of the variables describing the 2-choice alternatives (and no others!).

$$\frac{P_i}{P_j} = \frac{h(Z_{ki})}{h(Z_{kj})} \tag{5}$$

where

P_i = probability of choosing i;
P_j = probability of choosing j;
$h(Z_{ki})$ = strict utility of i; and
Z_{ki} = (scale) variables, k, describing i.

The actual odds or probability P_i of choosing alternative i from a larger set of alternatives can vary, of course.

The binary-choice strict-utility model, Eq. 5, generalizes into a multiple-choice model only if the independence axiom holds, that is, only if the probability of a choice from a subset of alternatives is independent of what other choice alternative may also have been available. The resulting multiple-choice strict-utility model is (41)

$$P(i{:}A) = \frac{h\,(Z_{ki})}{\sum_{j \epsilon A} h\,(Z_{kj})} \tag{6}$$

for j = 1, ..., i, j, ..., where

$P(i{:}A)$ = probability of choosing i from a set of alternatives A;
$h(Z_{kj})$ = strict utility of alternative j in the set A, a monotonic function of the scale variables Z_k describing j; and
$j \epsilon A$ = complete set of alternatives between which a choice is made.

An exponential transformation of the strict utilities (and an abandonment of set notation) yields the multinomial logit formula:

$$P_i = \frac{e^{v(z_{ki})}}{\sum_{j=1}^{J} e^{v(z_{kj})}} \tag{7}$$

for j = 1, ..., i, j, ..., J.

Equation 7 says that the probability that a traveler will choose alternative i out of a set of J alternatives is directly proportional to its strict utility $V(Z_{ki})$ (a monotonic function of attributes k of the alternative i) and that the probabilities of choosing one alternative in the set of available alternatives, each with a nonzero probability of being chosen, must sum to one. ["Perhaps the most general formulation of the independence axiom is the assumption that the alternatives can be scaled so that the choice probability is expressible as a monotone function of the scale variables, k, of the respective alternatives" (75). This assumption is called simple scalability by Krantz (35).]

The function $V(Z_{ki})$ in Eq. 7 can, of course, be interpreted and estimated. In the language of the psychologist, it represents some function of the environment that stimulates a decision (70). In utility terms, it represents some function of the attributes of value to travelers of the alternative travel choices. A correct model specification is needed to capture appropriate effects on behavior of variables (attributes) describing the choice situation. A constant term, θ, in an equation for $V(Z_{ki})$, e.g., $\theta_i \prod_k Z_{ki}^{a_{ki}}$, will include the effects of all attributes not explicitly included in the model.

Separability Property

The independence axiom is a general statement that has consequences that can be tested. For example, it says that, if alternative i is preferred to j in one context (choice situation), it is preferred to j in any context for which both are available. Furthermore, if the odds of choosing i over j are 0.7 in one context, those odds will be preserved in any choice situation. The traveler is assumed to exhibit transitivity in his behavior with respect to his "strict utility" $h(Z_{ki})$ versus $h(Z_{kj})$. That is, he values the attributes, Z_k, of any choice, i, the same (ratio scale) relative to choice j regardless of the context. Thus, the probability that an alternative (choice) will be chosen is exactly proportional to its strict utility (therefore, Eq. 6). And from Eq. 5, the relative odds that an alternative will be chosen from 2 alternatives is constant and a function only of the strict utilities of the 2 alternatives. This allows the introduction of new alternatives in a model application without calibration of the model, provided the previously estimated strict utilities are preserved.

In 1962, the author used the separability property of Eq. 6 to calibrate a share model of (multiple) choice among 4 access mode (walk, park-ride, kiss-ride, and feeder bus to line-haul rapid transit) alternatives being tested in Washington, D.C. The model was calibrated with paired aggregate modal-split data from a number of surveys because of the lack of data describing the relative usage of all 4 feeder modes together. This was allowable because of the "startling" behavior of the model (Eq. 6) that "the relative substitutability of any two sub-modes without the third being available is assumed equal to the relative attractiveness of the two in the presence of the third" (6).

McLynn and Woronka (50) used this property extensively to calibrate their "single pair" market share model developed for the Northeast Corridor project (see Appendix). In their model, automobile was used as the "base mode" (16). When difficulties were encountered with certain nonsensical parameter estimates and the single-pair estimates, all single-pair equations were estimated simultaneously. From Eq. 5, it follows that such simultaneous estimation is irrelevant from the point of view of the behavioral grounding of the model, however much it may be desirable to constrain certain parameter estimates. [The derivation of the model from strict-utility considerations highlights certain of its behavioral groundings that may not be clear from the McLynn derivation (see earlier sections).]

The property of "separability" of alternatives is not restricted to alternatives among modes. Alternatives can characterize the entire range of choices of trip frequency, destination, time of day, mode, and path, as already discussed. Thus, separate choice models can be calibrated separately and later combined into a travel-demand model. However, behavioral assumptions as to the sequence of travel decisions are required, as already discussed. The separability property of the independence axiom was first explicitly recognized and used to calibrate a travel-demand model by Charles River Associates (CRA) (10).

Share models have been used in travel forecasting without recognition of their separability properties for many years. For example, the gravity model of trip distribution (77) is a share model whose standard derivation is simple and general (18).

$$\left.\begin{aligned}
V_{ij} &\sim G_i A_j Z_{ij}^k \\
V_{ij} &= C_i G_i A_j Z_{ij}^k \\
G_i &= \sum_j V_{ij} = \sum_j C_i G_i A_j Z_{ij}^k \\
G_i &= C_i G_{ij} \sum_j A_j Z_{ij}^k \\
C_i &= \frac{1}{\sum_j A_j Z_{ij}^k}
\end{aligned}\right\} \qquad (8)$$

$$V_{ij} = G_i \frac{A_j Z_{ij}^k}{\sum_j A_j Z_{ij}^k} \tag{9}$$

$$\frac{V_{ij}}{G_i} = \frac{A_j Z_{ij}^k}{\sum_j A_j Z_{ij}^k} \tag{10}$$

Equation 8 states that the volumes between zones i and j are proportional to the previously estimated trips generated, G_i, and attracted, A_j, and to the attributes, k, of travel between i and j. C_i is the constant of proportionality, which is solved for in the remaining equations. The result, Eq. 9, is the usual form of the gravity model, which is equivalent to a share model, Eq. 10, for the split fraction of total trips from a zone i destined to zone j. However, the previously estimated "strict utilities" that (may have) resulted in the estimation of the G_i and A_j are not normally preserved.

In fact, of course, no transportation attributes are normally used in the estimation of the productions, G_i, and the attractions, A_j. Empirical evidence to support the use of strict utilities is the juggling necessary to bring the V_{ij}'s into line with the G_i and A_j in any gravity model application. That is, the results of the separately calibrated trip-generation and -distribution models are not (internally) consistent with each other.

The separability property implies that the conventional gravity model should be calibrated only with subregional structures (partitionings) that define distinctly different destination alternatives with nonzero probabilities of being chosen from a particular origin by a particular traveler (type) for a particular trip purpose. This would considerably simplify calibration but would appear to complicate gravity model application, i.e., predicting trip distribution (see discussion in section on applying forecasting models). An understanding of the separability property may thus lead to substantially more effective gravity models. Empirical research is clearly needed.

The derivation of the gravity model (Eqs. 8, 9, and 10) from a simple proportionality statement can easily be generalized to derive any split fraction (e.g., fraction of total regional trips emanating from an origin zone, or fraction of total interzonal trips on each mode). Each split fraction is in turn dependent on the previously derived trip universe being split. The models can then be "solved," one in terms of the next, in one multiple-choice share model. The result is similar to Manheim's "general share model" (45):

$$V_{klmp} = \alpha\beta_k\gamma_{kl}\delta_{klm}\omega_{klmp} \tag{11}$$

where

V_{klmp} = travel between origin k and destination l by mode m and path p,
α = total (regional) travel,
β_k = split fraction of α from origin k,
γ_{kl} = split fraction of $\alpha\beta_k$ to destination l,
δ_{klm} = split fraction of $\alpha\beta_k\gamma_{kl}$ to mode m, and
ω_{klmp} = split fraction of $\alpha\beta_k\gamma_{kl}\delta_{klm}$ to path p.

Each of the terms on the right side of Eq. 11 is intended to be a function of activity system and transportation system variables in Manheim's model.

In summary, in the calibration of a travel demand model, the separability property of the independence axiom implies that the (marginal) probability distribution of choice of mode can be separately estimated and multiplied by the conditional probability distribution of another travel choice, e.g., P (destination, mode), to give the joint probability distribution of both:

$$P(M,D) = P\ (M)\ P(D|M) \tag{12}$$

provided the previously estimated strict utilities from the modal-choice model are preserved. This operation requires 2 assumptions: (a) that destination choices are made conditional on mode choices and not the reverse, and (b) that the (dis)utility from the mode choice is additive to the utility from the destination choice. Thus, the mode choice is assumed to be independently made from the destination choice (in this case) but not the reverse. Given the separability and sequence assumptions, the choices can be separately modeled, assuming negligible income effects, and later recombined into one joint probability model by simple multiplication of the separately calibrated probability models, as in Eq. 12. Conversely, the joint distribution, P(M, D), must be estimated directly if the sequence and separability assumptions appear too strong. The possible behavioral bases for sequential and separable choice assumptions are discussed in the next section.

Travel-Choice Behavior

Existing travel demand models are classified as short-run or long-run demand models, according to whether (short-run) travel decisions (choices) are modeled separately from (long-run) activity-location decisions. The additional classification of direct and indirect demand models is used to describe whether the short-run travel decision is modeled as one simultaneous "joint" choice or as a series of separate choices (e.g., mode, destination, frequency, and so on). In this section, certain behavioral assumptions in these choice classifications are discussed.

Activity (Land Use) Location

In travel demand forecasting, activity-location choices are assumed to take place in a much larger market than travel choices. Also, the time periods over which activity-location choices are made is assumed to be much longer. If activities are considered substitutes for each other in one market, this requires long-run demand models where activity locations and intensities are allowed to vary. The recent mixed success in land use modeling (38) testifies to the difficulty of describing the attributes of all the related choices in this larger market (which also includes travel choices). Thus, the present state of the art of travel demand forecasting with a few exceptions allows only amount of travel to vary, i.e., to be the dependent variable. [Some demand models have been formulated and calibrated that forecast (long-run) residential location, car ownership, and modal split in one equation set (1, 30). However, these models do not forecast quantity of travel. Nevertheless, the models provide a direction for further work.]

In modeling travel separately from activity location, the attribute variables describing the choice situation must be limited to those "highly" involved in the decision (i.e., close substitutes and complements). Indeed, a necessary condition for utilities derived from separately modeled travel decisions to be considered additive is that their components must be neither competitive (substitutes) nor complementary (43).

Trip purpose is the first way of describing the restricted set of choices that are said to be available to the traveler as an individual decision-maker. No substitution is assumed among trip purposes because the purpose of the trip corresponds to the activities at the trip destinations. The activities in place are taken as given in the partial equilibrium framework. If activities are taken as substitutes, a long-run demand (land use) model results.

The choice ordering implied by assuming that travel choices are made, conditional on activity locations, is represented in Eq. 13.

$$P(T, A) = P(T \mid A)\, P(A) \tag{13}$$

where

P(T, A) = joint probability distribution of travel and activity location;

$P(T \mid A)$ = conditional probability distribution of travel, given activity location; and
$P(A)$ = marginal probability distribution of activity location.

Equation 13 implies the sequence assumption that activity-location choices are made first and precede travel choices. The sequence requires that the strict utilities inferred from activity-location behavior be used in the calibration of the travel demand model. This is, of course, not the way travel models are currently calibrated.

It is, of course, possible to assume that travel and activity location are independent. That is,

$$P(T \mid A) = P(T) \tag{14}$$

This is exactly the assumption that is made when one assumes that there is a sequence of travel-choice decisions in which mode and route choice precede destination choice. That is, these choices are assumed to be made solely on the basis of the (dis)utility of the trip itself. Making this particular assumption of travel choice ordering (discussed in the next section) is at least consistent with Eq. 14.

In summary, although the logical conclusion of the theory of travel as a derived demand is to allow both short- and long-run travel activity to vary as complements in a general equilibrium framework (7), the assumption is made that we can eliminate the imposing structure this would require and model travel choices separately as an activity with a set of complements (activities) in place and fixed.

The resulting set of attributes needed to describe the choice environment for input to a travel-choice model is correspondingly (greatly) reduced. Further, the choice ordering implied by this assumption is that travel choices are adjusted much more quickly to a change in travel conditions than in residence and work-place location. Modeling the latter requires a dynamic model where changes are measured over relatively long periods. Thus, if a static travel model is assumed, the effects of changes in travel conditions on travel can be modeled (inferred), it is assumed, separately from their effects on activity location. This assumption and its implications are worthy of considerable research.

Travel Choices

The open question is, What does the traveler perceive in his evaluation of his travel alternatives? Modeling travel directly as a simultaneous decision means including the attributes of every conceivable alternative to a specific choice in any model of that choice. By modeling long-run demand separately from short-run travel, we exclude moving the traveler's residence and work-place location as alternatives to his travel choice. However, such alternative choices remain as traveling to activities at varying locations as an alternative to staying put (destination choice versus no-trip choice); an automobile trip at a different, say, off-peak, time of day as an alternative to a transit trip at the peak hour; and so on.

As noted in the introduction, the conventional breakdown of individual travel choices is to separately model trip frequency, trip destination, time of day, mode choice, and route choice. Such a breakdown involves a stronger set of assumptions than the assumption of simultaneous travel decisions. The trade-off is generally between a stronger set of assumptions but less complex models and weaker assumptions but more complex and difficult-to-calibrate models. The unanswered questions are, How difficult to calibrate are models that combine travel decisions, and how difficult are they to forecast with?

At least 2 of the conventional travel choices might plausibly and relatively easily be combined, at least for purposes of empirical testing. That is, combining trip frequency and trip destination into 1 set of alternative choices appears theoretically plausible and convenient. Zero-trip frequency is the equivalent of no change in traveler location. Other combinations may also be speculated on. However, some appear more difficult than others, not because of the difficulty in assuming that travel-choice behavior is a simultaneous decision, but because of the separability property of most existing travel models. For example, combining mode and route choice into one decision may be difficult because of the similar characteristics of alternative routes within modes and the overly strong separability property in this situation. [The evidence is that "the addition of an alternative to an offered set 'hurts' alternatives that are similar to the added alternative more than those that are dissimilar" (75).]

Because the basis of calibrating travel demand models using the separability property is to constrain some decisions on the basis of attribute (utility) evaluations made in decisions modeled earlier in the chain, a discussion of travel-choice-separation assumptions cannot proceed far without including consideration of the ordering of the separate choice assumptions.

Choice Ordering

The assumed order of the travel decisions, given a separation, determines which choice situation is used to estimate the initial strict utilities. Empirical testing with alternate orderings and breakdowns can provide some evidence as to "natural" orderings, given the underlying assumption of "conditional" choice behavior. Is there a logical or natural ordering of travel choices? If there is any separation at all, hypotheses can be attempted for specific orderings of the choices. The following hypotheses are some that support the assumption that travel choices are separable and proceed in some sequence or order.

1. Sequential choice ordering based on timing. Traveler decision-making proceeds from the latest to the earliest decisions in time. For example, for a particular trip purpose (choice-of-destination activity), the traveler may be hypothesized to have some notion of the conditions on the available modes and routes when choosing his destination. That is, he has already considered the modes and routes that are available to him. He anticipates and makes choices on routes and modes that may then limit or constrain his available destinations and departure times. (Within a mode, he is apt to have anticipated the conditions on the alternative routes within the mode when he makes his mode choice. This suggests that mode-choice decisions are made after path decisions as opposed to both decisions being made simultaneously.) This implies a logical order of travel-choice decisions running counter to their sequence in time.

The possibility of a logical order of decisions running counter to their sequence in time in the case of travel decisions was discussed already by Beckmann et al. in 1955 (3). This reverse order also gets us around the practical difficulties (probably impossibility) of having to compute supply-sensitive system characteristics (travel attributes) on an area-wide basis for input to (disaggregated!) trip-frequency decisions made at a point (or zone), or for input to a modal-split model that precedes trip distribution. Production functions g(x) for, say, travel times, are well known on a link and route within modal basis (28).

2. Sequential choice ordering based on adjustment time. Models that assume some choice ordering in a sequence could rest their plausibility on the time it takes to adjust behavior to a change in policy. Some decisions (e.g., route choice) can be adjusted more quickly by an individual than others (e.g., an origin change involving a house purchase or a mode change involving a car purchase) because they involve less commitment to their former situation. Thus, sequential choice models that involve adapting to changes in supply considerations can be considered in this sense dynamic or stochastic (5). Conversely, simultaneous-choice assumptions result in models that are in this sense static. Unfortunately, only cross-sectional data exist at present to empirically test most travel demand models.

3. Sequential choice ordering based on experience. Traveler decision-making proceeds from those choices on which there is the most experience to those choices on which there is the least experience. Most, if not all, current travel demand models are based on or can be shown to be equivalent to rational "economic man" assumptions. These yield plausible (if normative) descriptions (models) of travel behavior, but they

demand more of man's capabilities than he can generally "deliver." In addition, they assume that the traveler's values, and the choices he confronts, are constant over time. Conversely, there are other descriptions of behavior that assume less (or a bounded set of) knowledge on the part of the individual decision-maker. These provide alternate but as yet largely unexplored bases for modeling travel behavior, and the dynamics of commitment to old and selection of new travel choices as families move spatially and socially over time.

Important theoretical support for separate and sequential choice modeling comes from the theory of decision-making called "satisficing" (46). This theory rejects the notion that there exists a rational economic man who is perfectly knowledgeable and perceptive about all the possible alternatives that confront him and who can compare all possible alternatives with one another to find his optimal choice by manipulating stored criteria describing the alternatives. Satisficing substitutes for this true or complete rationality a hypothesis of bounded rationality. This implies sequential search and limited sets of criteria used for evaluation. That is, in place of simultaneous (or separable and transitive) comparison of all alternatives, alternatives are examined sequentially according to satisficing. And rather than being compared to one another on the basis of a set of (interval scale) operational criteria, the alternatives are compared to a simpler set of minimal criteria until an alternative is found that satisfies the decision-maker. Alternatives are discovered or searched sequentially until a satisfactory alternative is encountered. No attempt is made to exhaust all possible alternatives. Moreover, search for new alternatives will only occur if the traveler perceives a discrepancy between his level of aspiration and his level of reward from the existing behavior.

This "model" in its general formulation can be interpreted as supporting models of sequential travel behavior. Travelers can be considered to evaluate sequentially well-defined travel alternatives in terms of the objects that provide the travel service (modes) and in terms of the benefits from the travel service (destinations). Conversely, the traveler may sequentially apply a limited set of criteria that are used to reject alternatives that do not meet threshold levels of those criteria. (This latter interpretation provides support for choice-abstract sequential models.) In both cases there is support for the hypothesis of choice behavior that involves sequential examination of choices.

We may describe the present trip of a traveler as one path through the tree shown in Figure 2 (assuming he presently makes a trip). If he is dissatisfied with any aspect of his present trip or, if confronted by a new alternative with a promised or expected improved level of service, does he sequentially examine "near" alternatives at only one level of choice? Or does he reconsider many paths involving changes throughout the hierarchy? Or does he simply consider only the new alternative if available and accept it or reject it?

According to the theory of satisficing, there is generally a conservative bias in the system of choice. That is, over time, levels of aspiration tend to adjust to levels of achievement. (It is the difference in the levels that is said to motivate search for new alternatives.) A new alternative may or may not change the traveler's perception of difference between present and possible (future) alternative states if he changes his travel behavior. We clearly need to better understand what those perceptions of difference are, at what level in the hierarchy they occur, in what sequence they occur, and how their relative requirements of adjustment time may operate to eliminate certain choices from the sequence.

The above hypotheses that support sequential travel decision-making are not made as a matter of idle speculation. The current conventional procedure of travel forecasting assumes sequential travel choice and a very particular choice ordering. The choice ordering is allowed to vary only slightly in practice. For example, the place of modal split in the order of trip-choice decisions has been called "the most actively debated issue in modal split" (80). The context of this statement referred to whether modal split should precede or follow trip distribution. The alternatives can be represented by the following 2 model structures (probability statements in this case):

$$P(M, D) = P(D|M)\, P(M) \tag{15}$$

$$P(M, D) = P(M|D)\ P(D) \tag{16}$$

where M = mode, and D = destination. If Eq. 16 were true and Eq. 15 false, destination choice would be independent of the availability of a mode (say, automobile) to reach the destination. This does not seem plausible except possibly in the case of work trips. (In such a case, the car is assumed to be purchased if not available and if necessary for reaching the destination.) In the reverse case (Eq. 15 is true, and Eq. 16 is false), the choice of mode is assumed to be made independently of the choice of destination. For example, the automobile, if available, might be selected for the trip, and the destinations that can be reached by automobile are then considered by the traveler. This appears somewhat plausible (say, for convenience shopping trips), at least more plausible than the reverse sequence. (If this is true, at least for some important trip purposes, it augers badly for transit usage. That is, choice of mode, e.g., transit usage, would be independent of origin-destination transportation system characteristics, including origin-destination pairs in larger cities where transit service may be excellent.)

There is an alternative model structure that poses a way out of the above dilemma if the order of travel behavior is not stable or must be subjected to further empirical testing. Equations 15 and 16 may be rewritten in the following form (17):

$$P(M, D|MEX_D) = P(D|M)\ P(M|MEX_D) \tag{17}$$

$$P(M, D|DEX_M) = P(M|D)\ P(D|DEX_M) \tag{18}$$

where X_D is the set of all decisions made prior to the choice of destination, and $P(M, D|MEX_D)$ is, therefore, the conditional probability that M and D will be chosen if mode choice precedes destination choice. Analogous statements apply to Eq. 18. Because MEX_D and DEX_M are mutually exclusive, Eqs. 17 and 18 can be added together to yield

$$P(M, D) = P(D|M)\ P(M|MEX_D) + P(M|D)\ P(D|DEX_M) \tag{19}$$

This is an exact expression for P(M, D). Equation 19 is equivalent to Eq. 15 or 16 only if mode choice always precedes destination choice or vice versa. It is also possible to expand Eq. 19 to include all aspects of travel decision-making.

The logical place of the time of departure decision in an assumed sequence of decisions is diffucult to establish even in theory. It may, for example, plausibly come before or after the trip-destination decision. The separation of time-of-day utility from destination-place utility and trip (dis)utility, as noted before, may make this the weakest assumed separation, leading to confusion as to its place in a logical order of travel decisions. The choice of time of departure might best be combined with frequency or destination or both, even though this would make travel models more complex.

Unfortunately, a solid case cannot be made for many trip-choice sequence assumptions. Our theory is weak, and we must look at whatever empirical evidence is available. Ben-Akiva (5) showed empirically that mode choice, assumed before or after destination choice, or the 2 travel choices modeled jointly all lead to different valuations (relative marginal utilities) of the trip attributes, (e.g., time and money costs of travel). (But this is insufficient evidence to lead to the conclusion that both sequences are wrong or that the separation assumption is incorrect.) His work on estimating the joint probability of mode and destination choice directly is the first demonstration that disaggregate data can be used for simultaneous travel-choice models, though not all travel choices were included. [The first simultaneous choice model using aggregate (zonal) data was by Kraft in 1963. The trip-generation and mode-choice decisions were combined and modeled simultaneously. Again, not all travel choices were included.] By combining choices and modeling them simultaneously, the need for sequence assumptions, but not separability assumptions (except when applying the model directly), is avoided. That is, the separability property of any formula satisfying Eq. 6 (e.g., multinomial logit) allows travel choices to be separated while still preserving the strict utilities. The separability property allows the conditional and marginal probabilities of the travel choices to be computed from the joint probability distribution estimated from the simultaneous model. Thus, for forecasting purposes, models satisfying Eq. 6 may be separated and applied sequentially (indirectly) or combined for application in a direct model (see later discussion of alternative methods).

When travel-choice models are calibrated separately, the alternatives allowed are determined by the conditional probabilities. That is, in Eq. 15, the only alternatives allowed are the destinations that are available or can be reached by mode m. The estimated strict utilities from this set of choices are then assumed to be independent of the choices as soon as the separability property of Eq. 6 is used in travel forecasting (see later discussion of definition of alternative choices).

The hypothesis of simultaneous (i.e., not conditional) travel choices can be easily tested by using standard chi-square tests for differences between marginal and conditional distributions of the same random variable. If there are no differences, the hypothesis of no relation between, say, mode and destination could not be rejected. Because it is relatively easy to show a relation by the chi-square test with large sample sizes, an inability to reject no sequence might be considered evidence that the decisions are being made simultaneously. (However, the power of the test is low.)

Theories of choice that consider different choice-abstract aspects of travel attended to at difference times and in some specific order were discussed earlier. Aspects of travel can overlap with the definitions of travel choices because attributes in the definitions of each are often common to both. Some arguments against transitive value (strict-utility) models can be used in part to advance the case for assuming sequential travel choices and thus advantageous use of the separability property to calibrate demand models.

Similarly, arguments against a logical ordering of travel-choice decisions argue also for strict-utility travel-choice models because such arguments are consistent with assuming a single monotonic function of the scale variables of the alternatives and the single estimation of joint probability distributions of simultaneous travel choices (i.e., "direct" demand models). Therefore, uncertainties as to whether travel choices can be assumed to be separable and occur in some logical order do not point to abandoning strict-utility models. They may point to combining choices and making less use of the separability property in model calibration.

In summary, there may be some clear-cut travel-choice ordering that can be assumed from the standpoint of travel behavior and, thus, lead to the conclusion that probability models for combined choices should be calibrated directly wherever possible. Fewer sequence assumptions can lead to improved use of the separability property for combining separately modeled choices into a demand model. Because the independence axiom excludes, in any event, alternatives with zero probability of being chosen, the data requirements for estimating strict-utility models of combined travel choices can be greatly reduced. Simultaneous (direct) demand models rather than sequential choice models seem indicated from a behavioral point of view, although the discussion cannot be closed in view of the above hypotheses.

Combining Strict-Utility Sequential Travel-Choice Models

CRA (10) used the separability property of the independence axiom to calibrate a series of shopping-trip travel models in the following assumed sequence: mode choice, destination choice, time-of-day choice, and trip frequency (including whether to make the trip). Data at the individual traveler level were used. The relative marginal utilities of modal attributes revealed (estimated) in the mode-choice decision were preserved in the next choice modeled, namely, trip destination, by weighting the attributes of travel by mode to each destination by the probability that the mode would be chosen, given the selection of the destination. The weighting and aggregation are done with the estimated parameters from the previous (mode-choice) decision. The previously estimated strict utilities or "inclusive prices" are preserved. A proof is given that this method of combining separately calibrated travel-choice models is consistent with the assumption of additive utilities. There is no summation over the estimated number of trips because the choice of mode is assumed to be independent of the number of trips between an interzonal pair. "Tastes about modes are (assumed) independent of tastes about trip frequency" (10).

The method can be schematically portrayed for the 4 sequential shopping-trip decisions as follows:

$$\left.\begin{aligned} P(\text{mode}) &= f_m(p, s) \\ P(\text{time of day}) &= f_t(\hat{p}, s) \\ P(\text{destination}) &= f_d(\hat{\hat{p}}, s) \\ P(\text{frequency}) &= f_f(\hat{\hat{\hat{p}}}, s) \end{aligned}\right\} \tag{20}$$

where

p = vector of travel attributes,
$\hat{p}$ = previously estimated strict utility = "inclusive price,"
$\hat{p}, \hat{\hat{p}}$ = inclusive prices previously estimated, and
s = vector of socioeconomic variables.

This is the logical conclusion of the assumption of transitive tastes. (Strict utility suggests that "behavioral time values" have a legitimate place in transportation benefit measurement, assuming transitive tastes!

Summary

Figure 3 shows all the travel demand modeling choices considered thus far. The assumption of individuals' evaluating choices such that their probability of choice is expressible as a monotonic function of the choice-specific attributes of all the alternatives (simple scalability or strict utility) has been shown to be the expression of the independence of irrelevant alternatives axiom. This means that the relative probability of choice between 2 alternatives is independent of the attributes of other alternatives in the offered set of alternatives. The transitive nature (strict utility) of the resulting choice behavior results in multinomial, multivariate probability or share models. The separability property of the independence axiom and its resulting multiple-choice share models allow big, complicated travel decisions (e.g., those modeled in direct demand models) to be broken up into smaller, more easily modeled choices. However, these models may be separately calibrated only if separation and sequence assumptions are made. The separately calibrated models can then be linked through their previously estimated parameters into a demand model (i.e., a direct or one stage-pass demand equation). To do so requires use of probabilities (or relative frequencies), not summation of numbers of trips from the prior travel choice in the assumed sequence.

There is, in addition, a set of travel-choice models based on the strong assumption that the choice probabilities are expressible as a function of attributes of subsets of travel choices making up one complete travel decision. This requires the assumption of sequential and completely independent travel choices where the relative valuation of attributes common to 2 or more travel choices, making up one trip decision, is not constant throughout the hierarchy of travel choices (Fig. 2). These models (e.g., the present UTP models) cannot be combined into one direct demand model, but must be applied sequentially in the order in which they have been calibrated, as discussed in the next section.

Figure 1. Incomplete diagram of travel-modeling choices based on alternate travel-behavior assumptions.

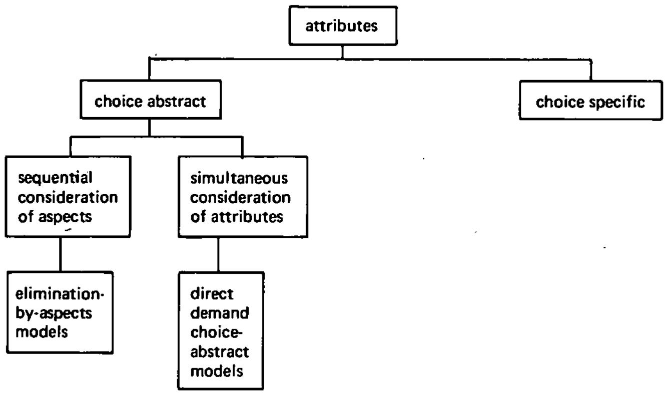


Figure 2. Presumed hierarchy of travel choices.

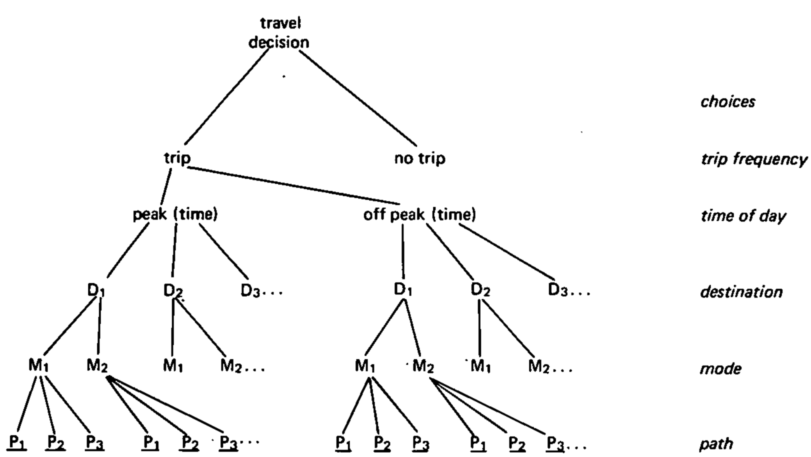
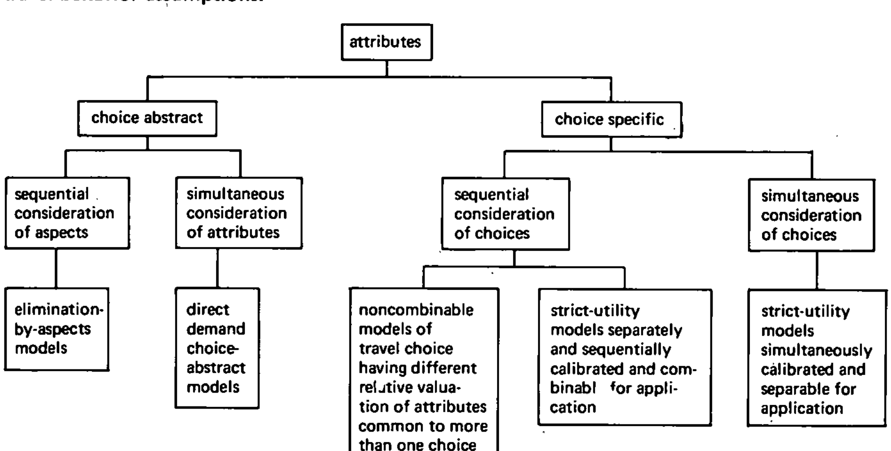


Figure 3. Less incomplete diagram of travel-modeling choices based on alternate travel-behavior assumptions.



APPLYING TRAVEL FORECASTING MODELS

Alternative Methods

The question remains of how to apply travel forecasting models. Five alternative methods are apparent.

1. Apply the models in chains in their usual UTP order (i.e., trip generation, trip distribution, modal split, traffic assignment);
2. Apply the models in chains as travelers are assumed to order their choices;
3. Link sequentially calibrated travel-choice models parametrically and apply them in one stage (i.e., as a direct demand model);
4. Apply simultaneously calibrated travel models in one stage (i.e., as direct-demand models); or
5. Apply sequentially the conditional and marginal probabilities of separate travel choices derived from the joint probability of a simultaneously calibrated model.

In the first (conventional) strategy of chaining independently calibrated travel-choice models with different relative valuations of independent variables common to 2 or more choices, the sequence of application determines the results. In such cases, the separability property of the independence axiom does not apply among choices. For example, in the application of binary-choice modal-split models in a chain, shown in Figure 4 (65), the results (i.e., splits) calculated higher in the chain are preserved lower in the chain. And in conventional UTP, the trips calculated higher in the chain are normally preserved lower in the chain on any pass through the chain.

The critical problem in method 1 is how to input the system characteristics (attributes) of the choices lower in the chain at points higher in the chain. For example, how in trip generation-trip frequency can the system characteristics for the entire region be aggregated to a single point or zone for input to this first step? The choice attributes can either be summed over (weighted by) trips calculated lower in the chain (e.g., potential functions or gravity-model weighted sums) and brought "up" to be input to higher models in the chain. Or the estimated parameters common to all the ordered-choice models can be used to probabilistically aggregate the choice-specific attributes from the lower level choices. The latter method, as noted before, is the only method consistent with the assumption of additive utilities from sequentially calibrated separable multiple-choice travel models.

If sequential models are derived and calibrated consistently with the (implicit or explicit) behavioral assumptions of preservation of strict utilities in separable multiple-choice models, there is no difference among methods 1, 2, and 3 in the resulting computed network-equilibrium travel patterns. That is, the same separable model may be applied sequentially in a series of separate travel-choice forecasts, or the joint probability distributions of choices may be calculated directly by parametrically combining the separately calibrated choice models as per the independence axiom. However, the sequential application of the models in this case can actually be in any order including methods 1 and 2. The estimated strict utilities are independent of the choices, as per the original behavioral assumption implemented by using the separability property of Eq. 6.

Conversely, from a simultaneously calibrated model satisfying the independence axiom, the conditional and marginal probabilities of travel choice may be derived, and the separate submodels of travel choice may be applied sequentially. Submodels so derived may be applied in any order, including methods 1 and 2. Joint estimation of the choice probabilities eliminates the need for the sequence assumption, but not the separation assumption, for models based on or consistent with the independence axiom.

Models based on or consistent with the independence axiom are separable multiple-choice models. Preference for any method of application is a matter of convenience, control, and purpose of the transportation systems analysis. For example, it is often desirable to be able to compute travel in sequential steps (generation, distribution, and so on) in order to be able to check the intermediate results and exert control over the forecasting process in some way. A direct application of the parametrically combined or simultaneous model may be appropriate if the user is confident of his results and wants to save time and money. If the model has been derived in a fashion consistent with its behavioral assumptions, both methods will produce the desired output for calculating the flow volumes on links in a transportation network. The choice of method should be based on the requirements of different planning environments.

Because the aggregate of trips, not the probabilities, are assigned to a network, a complete run through the sequence will be required to produce the joint probability distributions of travel (including trip-frequency probabilities) needed for aggregating over the total number of individual trip-makers to calculate the aggregate demand. Assignment of trips must also be made to update link and path supply functions for computation of an appropriate network equilibrium. Network equilibration can proceed either through incremental (fractional) loading or by iterating.

Defining Alternative Travel Choices

In the application of separable, multiple, choice-specific travel models (models having the separability property of the independence axiom), great care must be taken in choosing alternatives in order that the separability property not be too strong for the application. The strict utilities in these models are estimated in choice-specific situations even though the separability property of Eq. 6 allows travel choices to be separated for forecasting purposes while still preserving the strict utilities. Truly independent and distinct alternatives as perceived by travelers should be chosen in the application of separable multiple-choice share models. A black bus following the same route as a yellow bus, when chosen as an "independent" alternative, has the effect of reducing the use of automobile (the third choice) in order to preserve the relative odds of choosing automobile over either of the bus alternatives taken singly. This is a misapplication of the separability property because the property would appear to be too strong in this application. In model calibration, the color of the bus does not usually specify or identify a choice, so this seems perfectly clear. The black bus running on a different route from that of the yellow bus between the same origin and destination would have the same effect; and again this effect appears too strong, unless the strict utilities are clearly identified as route (choice) specific. If the yellow bus were now changed to yellow rail transit, and if the multiple choice-specific model were calibrated specifically with rail and bus transit parameters, as well as with automobile parameters, the separability property would appear not to be troublesome. Caution, however, is certainly advised.

Alternative destinations are rarely if ever defined in such a way that choice-specific strict (destination place) utilities are estimated for each destination. That is, the use of socioeconomic variables to describe the (static) trip-end activities amounts to the behavioral assumption of choice-abstract destination-place attributes embedded in an otherwise choice-specific travel demand model. Even more troublesome for the use of separable travel models are the implications of changing the destination alternative set from a small set of alternatives used for model calibration, each having nonzero probabilities of choice, to the usual large number of alternatives, among which trips are forecast in order that a high degree of resolution may be obtained for traffic-assignment purposes. In such cases, forecasting should probably be a 2-step process. That is, forecasts of trips should be made to large aggregations of zones, grouped on the basis that they are distinctly different and real (known) alternative destinations to travelers at the origin. Such grouped destinations might be based on a hierarchy of increasingly regionally oriented work or shopping places for the type of worker or shopper in each zone. Destinations not likely to be known to travelers at each origin would be eliminated from consideration. Forecasts to these zonal aggregations would then be allocated in some way to the small component zones for traffic-assignment purposes (e.g., based on employment share). Another possible way of forecasting is simply to truncate to zero trips to low (calculated) probability destinations, just as low or zero probability destinations were excluded from the data used in model calibration, as per the separability property of the independence axiom.

In summary, in an application of a separable multiple-choice share model (Eq. 6) within a hierarchical level (e.g., mode choice), the implication of the independence axiom is that the introduction of an additional transit alternative (mode or submode other than one for which the choice-specific strict utilities were estimated) will change the probability of choice (modal split) for all the existing modes. The relative share of all the existing modes included up to then in the analysis will be preserved because of the independence axiom. This also means that the cross elasticity of the modal fraction for each old mode with respect to an attribute of the new mode is the same for each of the old modes. For example, the cross elasticity of modal fraction on the old modes with respect to fare on a new transit submode will be equal for all automobile and transit alternatives considered thus far. This precludes a pattern of differential substitutability among modes and, in effect, implies a (mode) choice-abstract model with respect to the modal fraction, but not with respect to aggregate demand, however (10, 50).

A number of specific examples, such as the above black and blue bus versus the yellow and red bus, can be and have been used as criticisms of the overly strong separability properties of the independence axiom in many instances. Much practice will be required in defining alternatives before multiple-choice share models are usable in any but the most straightforward mode-choice situations in which they have thus far been applied with apparent success (e.g., by Rassam, Ellis, and Bennett, 60). One set of arguments in certain situations consists of citing examples where the relative odds of choice in a binary-choice situation are unlikely in fact to be preserved when new choices are offered [i.e., the black and yellow bus argument, or a second Beethoven record added to an original Debussy and Beethoven binary choice (12)]. Luce and Suppes (43) state:

> We cannot expect the choice axiom to hold over all decisions that are divided in some manner into two or more intermediate decisions. It appears that such criticisms, although usually directed towards specific models, are really much more sweeping objections to all our current preference theories. They suggest that we cannot hope to be completely successful in dealing with preferences until we include some mathematical structure over the set of outcomes that, for example, permits us to characterize those outcomes that are simply substitutable for one another, and those that are special cases of others. Such functional and logical relations among the outcomes (alternatives) seem to have a sharp control over the preference probabilities, and they cannot long be ignored.

COMBINING MODELING CHOICES: RESEARCH DIRECTIONS

Previous sections have described the major choice-behavior assumptions (stated or unstated) of existing travel forecasting models and discussed some of their implications. This section discusses briefly how those modeling choices might be combined and suggests some further research directions in this area.

Combining Modeling Choices

The choice-specific sequential and the choice-abstract sequential (elimination-by-aspects) models of choice behavior can be combined in their use. That is, when all available (noneliminated) alternatives contain all the remaining aspects (as, for example, if travel time and cost were the entire set of remaining aspects in the scenario in an earlier section), the independence axiom is shown by Tversky (75) to again hold. Thus, the elimination-by-aspects model can be used to select the "independent" alternatives having non-zero-choice probabilities among which choice is allowed. These allowable choices may then be modeled by using forecasting models based on monotonic functions of the remaining important attributes. The remaining attributes may or may not be perceived by the traveler as identified (or modeled) with specific supply-side choices (i.e., as choice-specific attributes). Figure 5 shows these modeling choices (as arrows) added to the previously described set of modeling choices. This "completes" the diagram of modeling choices based on alternate travel-behavior assumptions.

Figure 4. Modal-split chain for commuter travel.

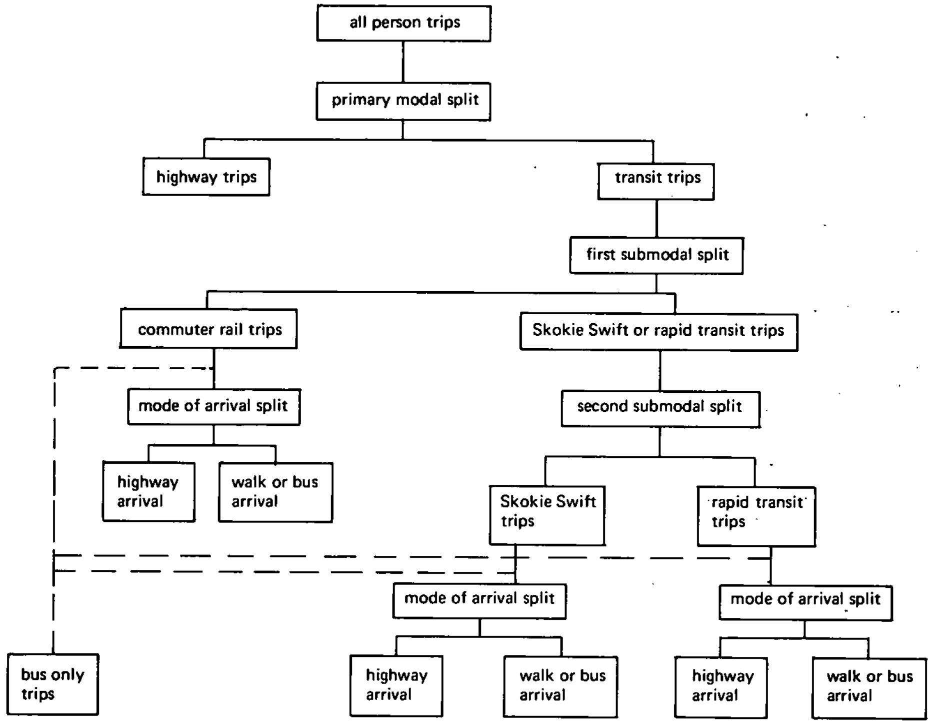
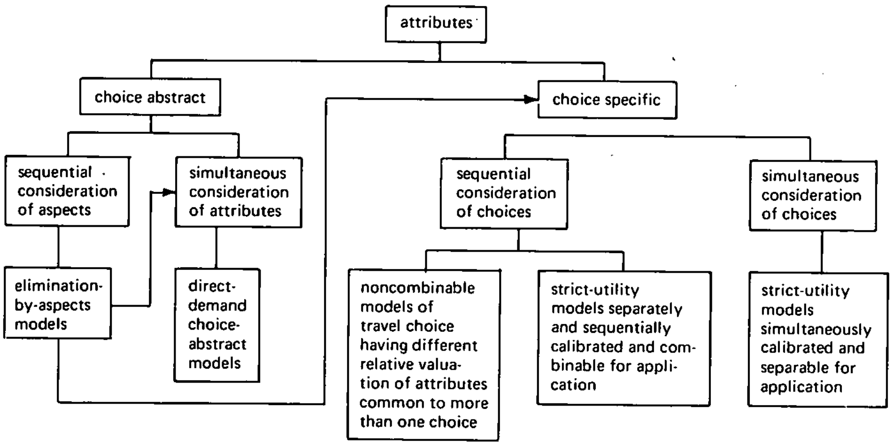


Figure 5. Complete diagram of travel-modeling choices based on alternate travel-behavior assumptions.



It is perhaps also possible that the arrows can be drawn symmetrically from right to left, that is, from choice-specific models to choice-abstract models. For example, this might more accurately describe a travel demand model having choice-specific mode and route attributes (assumed first in the choice ordering) and choice-abstract destination and origin-place attributes. This highlights the difficulty that existing travel demand models have in discriminating among competing activity locations. That is, there are no specific cross relations among place (choice) specific trip destinations in practically any existing travel (forecasting) models. However, the diagram need not be additionally embellished at this writing.

Additional Research Directions

Other decision rules can also be imagined in the sequential choice-abstract case. For example, more than one aspect at a time can be applied to eliminate alternatives. However, this produces the same results as applying aspects one at a time because all alternatives not containing the aspects are eliminated either way. A search of the mathematical psychology literature will no doubt turn up additional possible sequential choice rules.

Is there a remaining possibility that certain travel choices are decided on the basis of different weightings of the attributes than other choices? This would require that trip choices be perceived as fundamentally different, independent, nonhierarchical choices and that alternatives considered for each choice be disjoint (no aspects or attributes contained in common) with the alternatives for another choice. This appears to be the strongest (most heroic) assumption, as discussed earlier. If the assumption can be verified, it would certainly strengthen the basis in behavior of present UTP models. Clearly, some important research questions remain.

ACKNOWLEDGMENTS

I have benefited greatly by discussions during the writing of this paper. I would like particularly to thank Moshe Ben-Akiva, Tom Domencich, Greg Ingram, Jim McLynn, Mel Cheslow, Tom Lisco, Pat Burnett, Marvin Manheim, and Fred Dunbar, as well as several of my students, particularly Nicos Polydorides, David Alshuler, and Elaine Li. Responsibility for errors and improper attribution of previous works is mine.

REFERENCES

1. Aldana, E. Towards Microanalytic Models of Urban Transportation Demand. Dept. of Civil Eng., M.I.T., Cambridge, PhD thesis, 1971.
2. Alonso, W. The Quality of Data and the Choice and Design of Predictive Models. HRB Spec. Rept. 97, 1968, pp. 178-192.
3. Beckmann, M., McGuire, C. B., and Winsten, C. B. Studies in the Economics of Transportation. Yale Univ. Press, New Haven, 1955, Ch. 2.
4. Beckmann, M., Godfrey, M., Golob, T., Huber, M., Jones, A., and Whorf, R. Prediction of Demand for New Urban Transportation Systems. Urban Transp. Workshop Proc., Oper. Res. Soc. of Amer., Baltimore, 1973.
5. Ben-Akiva, M. E. Structure of Travel Demand Models. M.I.T., Cambridge, PhD dissertation, 1973.
6. Brand, D. The Derivation and Application of Program Centipede. Natl. Cap. Transp. Agency, Washington, D.C., 1962.
7. Brand, D. Theory and Method in Land Use and Travel Forecasting. Highway Research Record 422, 1973, pp. 10-20.
8. Brunner, G. A., Hille, S. J., Nash, A. N., Paine, F. T., Schellenberger, R. E., Smerk, G. M., and Taff, C. A. User Determined Attributes of Ideal Transportation Systems. Dept. of Bus. Admin., Univ. of Maryland, College Park, 1966.

9. A Model of Urban Passenger Travel Demand in the San Francisco Metropolitan Area. Charles River Assoc., Inc., 1967.
10. A Disaggregate Behavioral Model of Urban Travel Demand. Charles River Assoc., Inc., 1972.
11. Crow, R. T., Young, K. H., and Cooley, T. Models of Intercity Demand—Part 1: Theoretical Aspects. Northeast Corridor Transp. Proj., U.S. Dept. of Transp., Washington, D.C., 1971.
12. Debreu, G. Topological Methods in Cardinal Utility Theory. In Mathematical Methods in the Social Sciences (Arrow, K. J., Karlin, S., and Suppes, P., eds.), Stanford Univ. Press, 1960.
13. Dial, R. A Probabilistic Multipath Traffic Assignment Model Which Obviates Path Enumeration. Transp. Res., Vol. 5, No. 2, 1971.
14. Domencich, T. A., Kraft, G., and Valette, J. P. Estimation of Urban Passenger Travel Behavior: An Economic Demand Model. Highway Research Record 238, 1968, pp. 64-78.
15. Dyckman, J. W. Planning and Decision Theory. Jour. of Amer. Inst. of Planners, 1961.
16. Endriss, J. R., McLynn, J. M., and Smith, D. E. Calibration of the Multimode Assignment Model: Choice of the Base Mode. Davidson, Talbird and McLynn, Bethesda, Md., 1967.
17. Ewing, R. Sequential Decision Making and Travel Demand Forecasting. Harvard Univ., Cambridge, Mass., unpublished term paper, 1973.
18. Urban Transportation Planning, General Information and Introduction to System 360. Fed. Highw. Admin., U.S. Dept. of Transp., 1970.
19. Fertal, M. J., Weiner, E., Balek, A., and Sevin, A. F. Modal Split. Bur. of Public Roads, U.S. Dept. of Commerce, 1966.
20. Forbord, R. J. Twin Cities Modal Split Model. Minnesota Dept. of Highw., St. Paul, 1966.
21. Golob, T. F., and Beckmann, M. A Utility Model for Travel Forecasting. Transp. Sci., Vol. 5, No. 1, 1971.
22. Gur, Y. Development and Application of Direct Traffic Estimation Method. Highway Research Record 392, 1972, pp. 76-88.
23. Hall, A. D., and Fagen, R. E. Systems Engineering. Bell Teleph. Lab., New York, 1956, Ch. 1.
24. Hartgen, D. T., and Tanner, G. H. Investigations of the Effect of Traveler Attitudes in a Model of Mode-Choice Behavior. Highway Research Record 369, 1971, pp. 1-14.
25. Heanue, K. E., and Pyers, C. C. A Comparative Evaluation of Trip Distribution Procedures. Public Roads, Vol. 34, No. 2, 1966.
26. Hill, D. M., and von Cube, H. G. Development of a Model for Forecasting Travel Mode Choice in Urban Areas. Highway Research Record 38, 1963, pp. 78-96.
27. Hille, S. T., and Martin, T. K. Consumer Preference in Transportation. Highway Research Record 197, 1967, pp. 36-43.
28. Irwin, N. A., Dodd, N., and von Cube, H. G. Capacity Restraint in Assignment Programs. HRB Bull. 297, 1961, pp. 109-127.
29. Irwin, N. A., and von Cube, H. G. Capacity Restraint in Multi-Travel Mode Assignment Programs. HRB Bull. 347, 1962, pp. 258-289.
30. Kain, J. F. A Contribution to the Urban Transportation Debate: An Econometric Model of Urban Residential and Travel Behavior. Rev. of Econ. and Stat., Vol. 46, 1964.
31. Kannel, E. J., and Heathington, K. W. The Temporal Stability of Trip Generation Relationships. Highway Research Record 472, 1973.
32. Kassoff, H., and Gendell, D. S. An Approach to Multiregional Urban Transportation Policy Planning. Highway Research Record 348, 1971, pp. 76-93.
33. Kraft, G. Demand for Intercity Passenger Travel in the Washington-Boston Corridor. Bur. of Public Roads, U.S. Dept. of Commerce, 1963, Pt. 5.
34. Kraft, G., and Wohl, M. Special Survey Paper: New Directions for Passenger Demand Analysis and Forecasting. Transp. Res., Vol. 1, No. 3, 1967.

35. Krantz, D. H. Conjoint Measurement: The Luce-Tukey Axiomatization and Some Extensions. Jour. of Math. Psychol., 1964.
36. Lancaster, K. J. A New Approach to Consumer Theory. Jour. of Polit. Econ., Vol. 74, 1966.
37. Ledyard, J. O. Cost Benefit Analysis as a Statistical Hypothesis Test: An Example From Urban Transportation. Environ. and Planning, Vol. 2, 1970.
38. Lee, D. B. Models and Techniques for Urban Planning. Cornell Aeronaut. Lab., Inc., Buffalo, 1968.
39. Lisco, T. E. The Value of Commuter's Travel Time—A Study in Urban Transportation. Univ. of Chicago, PhD thesis, 1967.
40. Loubal, P., and Potts, R. A Mathematical Model for Trip Distribution. Univ. of California, Berkeley, 1970.
41. Luce, R. D. Individual Choice Behavior. John Wiley, New York, 1959.
42. Luce, R. D., and Raiffa, H. Games and Decisions. John Wiley, New York, 1957.
43. Luce, R. D., and Suppes, P. Preference, Utility and Subjective Probability. In Handbook of Mathematical Psychology (Luce, R. D., Bush, R. R., and Galanter, E., eds.), Vol. 3, John Wiley, New York, 1965.
44. Manheim, M. Fundamentals of Transportation Systems Analysis, prel. ed. Cambridge, Mass., 1971, Ch. 3.
45. Manheim, M. L. Practical Implications of Some Fundamental Properties of Travel-Demand Models. Highway Research Record 422, 1973, pp. 21-38.
46. March, J. G., and Simon, H. A. Organizations. John Wiley, New York, 1958.
47. Mayberry, J. P. Structural Requirements for Abstract Mode Models of Passenger Transportation. In The Demand for Travel: Theory and Measurement (Quandt, R. E., ed.), Heath, Lexington, Mass., 1970.
48. McFadden, D. Revealed Preferences of a Public Bureaucracy. Technical Inst. of Internat. Studies, Univ. of California, Berkeley, Tech. Rept. 17, 1968.
49. McLynn, J. M., Goldman, A. J., Meyers, P. R., and Watkins, R. H. Analysis of a Market Split Model. Jour. of Res. of Nat. Bur. of Stand., Vol. 72B, No. 1, 1968.
50. McLynn, J. M., and Woronka, T. Passenger Demand and Modal Split Models. Northeast Corridor Transp. Proj., U.S. Dept. of Transp., 1969.
51. McLynn, J. M. The Metric Model. Off. of High-Speed Ground Transp., U.S. Dept. of Transp., 1971.
52. Mansod, S. Relative Shares Model. In Studies in Travel Demand. Mathematica, Princeton, Vol. 5, 1969.
53. Ockert, W., Easler, R., and Spielberg, F. Analysis of Travel Peaking. Highway Research Record 369, 1971, pp. 159-180.
54. Mathematical Model for Estimating Total Intercity Travel and Trip Interchanges Between Cities in the Northeast Corridor. Peat, Marwick, Livingston and Co., New York, 1969.
55. Pratt, R. H., and Deen, T. B. Estimation of Sub-Modal Split Within the Transit Mode. Highway Research Record 205, 1967, pp. 20-30.
56. Pratt, R. J. A Utilitarian Theory of Travel Mode Choice. Highway Research Record 322, 1970, pp. 40-53.
57. Quandt, R. E., and Baumol, W. J. The Demand for Abstract Transport Modes: Theory and Measurement. Jour. of Reg. Sci., Vol. 6, No. 2, 1966.
58. Quandt, R. E., and Baumol, W. J. The Demand for Abstract Transport Modes: Some Hopes. Jour. of Reg. Sci., Vol. 9, No. 1, 1969.
59. Quarmby, D. A. Choice of Travel Mode for the Journey to Work: Some Findings. Jour. of Transp. Econ. and Policy, Vol. 1, No. 3, 1967.
60. Rassam, P. R., Ellis, R. H., and Bennett, J. C. The n-Dimensional Logit Model: Development and Application. Highway Research Record 369, 1971, pp. 135-147.
61. Samuelson, P. A. Foundations of Economic Analysis. Harvard Univ. Press, Cambridge, Mass., 1947.
62. Schmidt, R. E., and Campbell, M. E. Highway Traffic Estimation. Eno Found. for Highw. Traffic Control, Saugatuck, Conn., 1956.

63. Creighton, R. L., Hoch, I., Schneider, M., and Joseph, H. Estimating Efficient Spacing for Arterials and Expressways. HRB Bull. 253, App. B, 1960, pp. 1-43.
64. Schneider, M. A Direct Approach to Traffic Assignment. Highway Research Record 6, 1963, pp. 71-75.
65. Schultz, G. W., and Pratt, R. Estimating Multimode Transit Use in a Corridor Analysis. Highway Research Record 369, 1971, pp. 39-46.
66. Shaffer, M. T. Attitude Techniques in Action. Highway Research Record 305, 1970, pp. 112-122.
67. Shunk, G. A., and Bouchard, R. J. An Application of Marginal Utility to Travel Mode Choice. Highway Research Record 322, 1970, pp. 30-39.
68. Simon, H. A. Models of Man. John Wiley, New York, 1957.
69. Stopher, P. R. A Probability Model of Travel Mode Choice for the Work Journey. Highway Research Record 283, 1969, pp. 57-65.
70. Stopher, P. R., and Lisco, T. E. Modelling Travel Demand: A Disaggregate Behavioral Approach, Issues and Applications. Transp. Res. Forum Proc., 1970.
71. Stopher, P. R., and Lavender, J. O. Disaggregate, Behavioral Travel Demand Models: Empirical Tests of Three Hypotheses. Transp. Res. Forum Proc., 1972.
72. Theil, H. Principles of Econometrics. John Wiley, New York, 1971.
73. Thomas, T. C., and Thompson, G. Value of Time Saved by Trip Purpose. Highway Research Record 369, 1971, pp. 104-117.
74. Townsend, H. Two Models of Freight Modal Choice. In Studies on the Demand for Freight Transportation, Mathematica, Princeton, Vol. 3, 1969.
75. Tversky, A. Elimination by Aspects: A Theory of Choice. Psychol. Rev., Vol. 79, No. 4, 1972.
76. Study of Passenger Travel Demand Forecasting Techniques. U.S. Dept. of Transp., 1972.
77. Voorhees, A. M. A General Theory of Traffic Movement. Inst. of Traffic Eng., Washington, D.C., Spec. Rept., 1955.
78. Travel Demand in 1990. Alan M. Voorhees and Assoc., 1966.
79. Warner, S. L. Stochastic Choice of Mode in Urban Travel: A Study in Binary Choice. Northwestern Univ. Press, Evanston, Ill., 1962.
80. Weiner, E. Modal Split Revisited. Traffic Q., Jan. 1969.
81. Wilson, A. G. A Statistical Theory of Spatial Distribution Models. Transp. Res., Vol. 1, No. 3, 1967.
82. Wilson, A. G. Entropy Maximizing Models in the Theory of Trip Distribution. Jour. of Transp. Econ. and Policy, Vol. 3, 1969.
83. Wilson, A. G., Bayliss, D., Blackburn, A. J., and Hutchinson, B. G. New Directions in Strategic Transportation Planning. In The Urban Transportation Planning Process, Organization for Economic Cooperation and Development, Paris, 1971.
84. Wohl, M. A Methodology for Forecasting Peak and Off-Peak Travel Volumes. Highway Research Record 322, 1970, pp. 183-219.
85. Young, K. H. An Abstract Model Approach to the Demand for Travel. In Studies in Travel Demand, Mathematica, Princeton, Vol. 5, 1969.
86. Fleet, C. R., and Robertson, S. R. Trip Generation in the Transportation Planning Process. Highway Research Record 240, 1968, pp. 11-31.

APPENDIX

Utility Analysis

Perhaps the most plausible descriptions and interpretations of travel-choice behavior derive from utility theory. That theory describes the traveler as an individual welfare maximizer, one who maximizes his own personal welfare from travel, subject to constraints, such as not exceeding his total time or resources available. Considerable scholarship in the field of economics has been devoted to developing a "science of rational choice," resulting from the utilitarian foundations of modern economics that people do (or tend to) act rationally (15). That is, that people act to maximize their own utility. Whether or not the reader agrees with utility analysis is inconsequential to the theoretical development in the body of this paper. Certain important travel models that can be analytically derived from consideration of an individual traveler's maximizing personal utility from travel can more simply be derived on the basis of logic from assumptions on consistent choice behavior. However, utility theory derivations highlight certain additional assumptions of these models, which are usefully documented in a state-of-the-art paper.

Principles

Travel forecasting can be based on consideration of the rational individual's maximizing his own welfare or benefit from travel. Travelers are presumed to be rational decision-makers, acting in their own behalf. This constitutes the basic normative statement of behavior of the system that has as its objective adequately describing that behavior. For this property to be used to solve rigorously (analytically) for the state of the system at any time requires that the assumption be made that the system is in static equilibrium. Comparison of alternatives via the comparison of (travel) outcomes of alternatives is made by the method of comparative statics.

The equilibrium solution proceeds from the property that "the conditions of equilibrium are equivalent to the maximization of some magnitude" (61). In utility analysis, personal utility is maximized subject to certain time and resource constraints. "The individual confronted with given prices and confined to a given total expenditure selects that combination of goods which is highest on his preference scale" (61). At equilibrium, the ratio of the marginal utility of 2 choices is equal to the ratio of their "weighted" attributes (i.e., their revealed "prices"). The relative marginal utilities of the attributes of a choice situation can be solved for (inferred from) observed data on the choices made.

In general, therefore, the utility, U, of a trip is related to the attributes of characteristics, Z, of a trip through some constants of proportionality, u_k. For example, in linear form,

$$U = U(Z) = u_o + \sum_k u_k Z_k \tag{21}$$

In equilibrium analysis, this function is maximized, subject to certain constraints (e.g., budget). The attributes are related to the amount of travel, X, and the characteristics of the choices by means of "supply" or production functions,

$$Z_k = g_k(X) \tag{22}$$

where g(X) is specified by the choices (e.g., the transportation "technology" and link characteristics in the case of travel time over a single link in the usually depicted speed and volume supply function). In the general case, the attributes Z_k are outputs of the consumption activity **X**, travel.

The problem of deriving a demand function then becomes one of specifying the attribute variables, Z_k, that describe the traveler's choice situation, and the form of the utility function, U = U(Z). Utility maximization calculus is then applied to solve for travel, **X**, at the point at which the marginal costs of travel equal the marginal benefits from travel. (Continuous functions are assumed in the usual formulation, although discrete choice alternatives can be encompassed in programming solutions.) This results in some function of the scale values of the attributes.

$$\mathbf{X} = f(Z_k) \tag{23}$$

where **X** = quantity of travel.

Specifying Utility

The first step in deriving demand models analytically from utility assumptions involves specifying the utility function, $U = U(Z)$. Travel, according to prevailing thought, is a derived-demand commodity (34): "A trip is made because a household member wishes to purchase commodities or services, or obtain other satisfactions such as the purchase of food, a visit to the doctor, or obtaining of income (through work)."

Travel activity can be considered to consist of positively valued time foregone at the trip origin, time and money spent in travel, and positively valued benefits at the trip destinations. The quantity being maximized would therefore be some function of the benefits (utility) from the purpose(s) served by travel and the cost (disutility) of travel. The utility function, $U(Z)$, includes Z variables that describe characteristics of consumption activities, A_i, as well as transportation "activities," Z_k.

$$U = U(A_i, Z_k) \tag{24}$$

Models derived analytically from utility theory must include other than transportation variables, Z. Travel choices that are based on maximization of personal utility and that exclude positive utility from activities at the trip destinations will result in minimum quantities of travel, **X**. Such models omit or set equal to zero the relations between travel and the consumption activities resulting from travel.

CRA (10) includes the characteristics of the trip-making populations, s, in its characterization of utility, $U(Z)$. Some others do not (e.g., Golob and Beckmann, 21). On practical grounds, Stopher and Lavender (71) show that separate choice equations estimated for each population group (or "market segment") gave better fits than choice equations that included separate socioeconomic variables. On the other hand, the inclusion of s, the population characteristics, in the utility function avoids the necessity of stratifying the data by population group and thus allows all the data to be used in estimation when the data are limited. However, the penalty is to increase the number of variables and interaction terms in the utility function.

Analytically Deriving Travel Models from Utility Analysis

Several examples exist in the literature of models of travel demand derived analytically from assumptions of maximizing personal utility from travel. Excellent examples for purposes of illustration and clarity are provided by Golob and Beckmann (21).

Their derivations start out with the statement of the utility functions in the form of Eq. 24. That is, trips, X_k^m, by mode m to destination k, generate utility, Z^p, based on the achievement of purpose, p, equal to the sum of the achievements of p at all destinations, k, visited.

$$A_i = Z^p = \sum_{k,m} \alpha_k^p X_k^m \tag{25}$$

where α_k^p = degree to which purpose p is served at destination k; and the trips generate disutility, y^r, equal to the sum of the traveler's expenditures in terms of attributes, r, incurred on trips to all destinations visited.

$$Z_k = y^r = \sum_{k,m} \beta_k^{mr} X_k^m \tag{26}$$

where β_k^{mr} = perceived expenditure in terms of attribute r on a trip to destination k by mode m.

The utility function, therefore, includes both the utility derived from the trip and the disutility incurred in making the trip.

$$U = U(Z^1, Z^2, \ldots, Z^p, y^1, y^2, \ldots, y^r) \tag{27}$$

The equilibrium solution proceeds from the hypothesis that the traveler maximizes this function with respect to the decision variables (trips), X_k^m. If continuity and other conditions are satisfied, the necessary utility maximization calculus can be applied. The necessary condition for a maximum,

$$dU/dX_k^m = 0 \tag{28}$$

says that trips to destination k by mode m will be pushed to the point where the marginal net utility is zero (i.e., where the combined marginal utilities of the trip purposes equal the costs of the trips), while trip mode combinations that do not occur have a nonpositive initial marginal utility.

This particular approach assumes direct maximization of utility with no money or time expenditure constraints. Golob and Beckmann go on to derive a generalized gravity model that assumes purposes are identical with destinations, power form utility functions, $U = U(X^*)$, and separable (additive) utilities.

$$X_{ik} = \left(\frac{u_k}{C_{ik}}\right)^{\frac{1}{1-w}} \tag{29}$$

where

X = number of trips,
i = origin,
k = destination,
u_k = attraction of a destination,
C_{ik} = generalized trip cost (an empirically derived constant), and
w = constant varying between 0 and 1.

The authors also deduce other demand functions based on other assumed forms of U(z) (e.g., step functions). They conclude, "While a great number of demand functions can be deduced from corresponding utility functions, not necessarily every proposed demand function can be interpreted as the result of utility maximization."

In summary, travelers are assumed in a (static) partial equilibrium mode 1 to behave in such a way that their jointly derived satisfaction from both travel and the activities at the trip end(s) is maximized. Travel is assumed to increase until the marginal (dis)utility of the trip itself is equal to the additional marginal utility of the activity that can be engaged in. Thus, utility-based travel demand models, calibrated at (assumed static) equilibrium, reveal or show marginal rates of substitution among all the separate attributes associated with the travel decision.

Travel Models: A Review

Probability Models

An important accommodation to the practical difficulty (impossibility) of exactly specifying the worth (utility) of a particular travel choice to an individual traveler is to assume that the utilities from these choices are random variables. In these random-utility models, probabilistic behavior is assumed from the randomness of the utility function. Another class of probability models can arise from the assumption of constant utility and a probabilistic decision rule, that is, where the utility function is a fixed numerical function of the attributes of the choice alternatives and the response probabilities are some function of the scale values of the relevant alternatives (43). [According to Beckmann et al. (4), "Trip behavior is held to be rational, albeit with a random component... within the decision-maker's own value set. If the random component were greater than the rational component, then any attempt at prediction would have to be abandoned, at least at the individual level."]

CRA (10) derives analytically the multinomial logit model of probabilistic travel choice from considerations of maximizing personal utility from travel. That is, the utility-maximizing individual discussed previously will choose alternative travel choice i if

$$U(Z_i) > U(Z_j) \tag{30}$$

for $i \neq j$, $j = 1, \ldots, J$. The model is derived on the basis of attributing a random element to the worthy (utility) of outcomes. Full information on the outcomes is assumed available, and individuals are assumed to exhibit no bias in the valuation they attach to the worth of choice alternatives.

The utility, $U(Z)$, is taken as randomly varying because the vector of attributes of the choices "does not capture all of the factors influencing the formation of tastes or the perception (measurement of attributes) of alternatives" (10). There is a value of $U(Z)$ for each individual drawn from the population with the same observed characteristics and choice alternatives.

The utility of a travel choice can be written as the sum of a nonstochastic function, $V(Z_i)$, and a stochastic term ξ_i.

$$U(Z_i) = V(Z_i) + \xi_i \tag{31}$$

The deviations ξ_i are assumed to be independently distributed random variables containing the effects on utility of the choice-situation attributes that are unable to be measured.

The choices of individuals are then modeled in a probabilistic manner. That is, the probability of choice of option i is

$$P_i = \text{probability } [V(Z_i) + \xi_i > V(Z_j) + \xi_j] \tag{32}$$

$$P_i = \text{probability } [\xi_j - \xi_i < V(Z_i) - V(Z_j)] \tag{33}$$

for $i \neq j$ and $j = 1, \ldots, J$.

The specification of the probability function, P_i, requires an explicit functional form and probability distribution for each of the terms in the (probability) argument. CRA shows that, if the ξ_j are independently distributed with identical reciprocal exponential distributions,

$$\text{Prob } (\xi_j \leq w) = e^{-e^{-w}} \tag{34}$$

for the 2 (binary) choice case where $i = 1, 2$,

$$P_i = \text{Prob } (\xi_2 - \xi_1 < w) = \frac{1}{1 + e^{-w}} \tag{35}$$

and, from Eq. 33,

$$P_i = \frac{1}{1 + e^{-[V(Z_1) - V(Z_2)]}} \tag{36}$$

This is the logit function for the probability of choice of alternative 1, analytically derived from considerations of maximizing individual (personal) utility.

If the stochastic term $\xi_2 - \xi_1$ is bivariate normally distributed, then the standard binary-choice probit model is derived [assuming $V(Z)$ is linear in parameters]. And if the stochastic term is uniformly distributed over the feasible range (for which P_i varies between 0 and 1), a truncated linear ogive curve is the resulting probability model of binary travel choice.

In the multiple-choice case, the same assumption on the distribution of the random terms results in the multinomial logit formula:

$$P_i = \frac{e^{V(Z_{ki})}}{\sum_{j=1}^{J} e^{V(Z_{kj})}} \tag{37}$$

Equation 37, which is the same as Eq. 7, says that the probability that alternative i will be chosen is directly proportional to its utility, $V(Z_{ki})$ (a function of attributes, k, of the choice situation, i), and that the probabilities of choosing one alternative in the set of available alternatives, each with a nonzero probability of being chosen, must sum to one. This is the same "strict-utility" multinomial, multivariate choice model of Luce (41), which was presented before in Eq. 7. The strict-utility model is shown by Luce and Suppes (43) as being a (independent) random-utility model, but not all random-utility models are strict-utility models, In fact, only independently distributed reciprocal exponential distributions of the random utilities, or monotonic transformations thereof, result in this equivalence. According to Luce and Suppes (43), "It is conjectured that these are the only reasonably well behaved examples, but no proof has yet been devised." CRA also rejects multiple-choice generalizations of other random-utility models (which assume other probability distributions of the utility functions) as being analytically intractable or otherwise computationally impossible to work with. For example, the multivariate normal distribution of the utilities with a known covariance matrix, which would yield a multiple-choice generalization of the binary-choice probit model, is rejected on this basis. Thus, the binary logit model is the only binary probability model for which the multinomial extension is practical.

By a logarithmic transformation of the utilities, we can write Eq. 37 as follows (similar to Eq. 6):

$$P_i = \frac{h(Z_{ki})}{\sum_{j=1}^{J} h(Z_{kj})} \tag{38}$$

for j = 1, ..., i, j, ..., J.

And, if the utility function is in product form,

$$P_i = \frac{\theta_i X_{1i}^{a_{1i}} X_{2i}^{a_{2i}} \cdots}{\sum_{j=1}^{J} \theta_i \prod_k X_{kj}^{a_{kj}}} \tag{39}$$

where X is used instead of Z to represent choice variables.

Equation 39 is the McLynn and Woronka "market share" modal-split model (50). The derivation of this model, which proceeds from aggregate travel-behavior assumptions, is shown in an earlier section on travel behavior. If the parameters of Eq. 39 are not mode specific (i.e., do not contain subscripted j parameters), the equations are the same for all modes in a modal-choice model. This is the Mansod relative shares model (52). This model "approaches mode abstractness" (11) (but not "abstract mode" or complete choice abstractness because the constant term is assumed to capture the effects of the unmeasured attributes of any choice alternative in the context of the choices available.) The model was developed for the Northeast Corridor where the new-mode problem was of great concern, as discussed earlier.

The class of separable multiple-choice share models of which Eq. 6 or 33 is the general statement has been shown (and will later be shown) to be derived from many different assumptions. The CRA derivation from consideration of personal utility shows the consistency of utility theory with the independence axiom. More important, it provides an additional basis in behavior for interpreting strict utility and specifying appropriate choice-specific variables (attributes) that determine choice behavior within (assumed) separate choice situations.

For example, depending on the travel choice, revealed marginal utilities from equilibrium analysis will probably vary simply because marginal utilities are generally not constant according to well-known theories of diminishing marginal utility. [In the psychological literature this is expressed as follows: The weight or importance of any attribute will vary with the individual's level of satisfaction with respect to that attribute (27). Also, stated attitudes toward the importance of a particular attribute are a function of both the underlying strength of the human need and its present satisfaction level (8). This appears to reduce considerably the ability to transfer the utilities in a model based on attitude survey results (24) from one surveyed situation to another situation. Also, the direction of change of an attribute is thought to influence the weight attached to that attribute (66). A method for including directionality of effect of a change in the attribute in a travel demand model has been proposed by McLynn in his metric model (51).] Constant marginal utilities need not be assumed in travel choice models based on the independence axiom. However, constant relative marginal utilities must be assumed for the strict-utility function, $V(Z_{kt})$, i.e., constant marginal rates of substitution between travel-related attributes, such as the traveler's willingness to trade off time and money. Functional forms of strict-utility functions should be used that are plausible from the standpoint of prior understanding of travel behavior and not be solely based on goodness-of-fit considerations (i.e., which describe best, or discriminate best among, alternative choice situations within the data set used for model calibration).

Multiple-Choice and Direct Demand Models

Stopher and Lisco (70) propose a multiple-choice probability model as follows:

$$P = P_g P_d P_m P_r \tag{40}$$

where

- P = probability that an individual will make a trip to a specific destination by a given mode and route;
- P_g = probability that an individual will choose to make a trip;
- P_d = probability that an individual will accept a destination, d, given that he will make a trip;
- P_m = probability that an individual will choose a mode, m, given that he will make a trip to a particular destination; and
- P_r = probability that an individual will choose a route, r, given that he will make a trip to a particular destination by a specific mode.

In Manheim's (45) general share model, Eq. 11, the split fractions (shares) are separately modeled and must each sum to one for "internal consistency." This allows a probabilistic interpretation similar to the Stopher-Lisco model, Eq. 40. Both are multiplicative and, thus, assume separability of the travel choices. However, there is no guarantee that Eq. 5 will hold because of strict utilities having been estimated in accordance with an assumption of constant relative valuation of attributes throughout all travel choices in the hierarchy. Stopher and Lisco (70) address themselves to this point as follows: "The objective is to make sure that the behavioral relationships identified in one detailed disaggregate (choice) model still retain their basic identity in the more aggregate general ones. The aim is to see that the summed models are indeed the sum of their parts." And Manheim (44) states, "A desirable property of a sequential implicit system is that it be internally consistent."

The authors thus appear to be leaning heavily toward assuming constant relative valuation of attributes throughout the complete travel decision. In this case, their choice models can be only separately calibrated given the additive utility assumptions as discussed in the earlier section on the separability property.

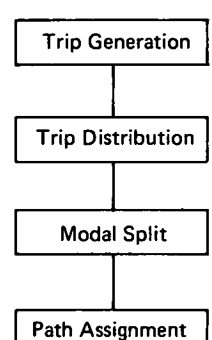
Manheim's (45) more detailed specification of his general share model, Eq. 11, is as a series of special product models. Each split fraction is in the form of Eq. 6, where the Z_{kt} include both activity and transportation system variables. For example, destination share, γ, is a share model in the form of Eq. 10, with A_j being activity system variables instead of trips attracted. Aggregation of costs (travel attributes) is carried out by simple summation. That is, the denominator in the (assumed) previous choice (e.g., trip distribution) is used to weight the (interzonal) travel attributes for input to trip generation. There is no (additional) relative frequency or probability weighting as there must be to preserve the basis in behavior of the additive utility and separability assumptions of the separately calibrated travel-choice models. The separability property can only be used to combine separately calibrated models on the basis of these assumptions.

Manheim (45) states, "Any explicit (direct) demand model can be expressed as a general share model." We note that it must be expressible as a multiplicative choice (share) model to be consistent with the basic travel-behavior assumption. However, none of the existing "1-stage" direct demand models is equivalent to the multiple-choice share model. It may be no accident that attempts to derive the present "standard" direct demand models analytically from considerations of maximizing personal utility have failed despite rather heroic attempts (10, 37). The possible reason the direct demand models cannot be analytically derived is that their causality premise (travel is a derived demand) results in a long-run demand (land use) model (7). Thus, short-run travel demand models, which are monotonic functions of scale variables describing the choice situation, can only be derived by resorting explicitly to the assumption that the (dis)utility of travel is additive to the utility from activities in place. The resulting models are probability share models of short-run travel choice.

McLynn and Woronka's composite analytic model (50) is a 2-stage (2-choice) aggregate demand model that incorporates the results of his separately estimated modal-choice share model (Eq. 39). The derivation of the model is only in terms of the shares themselves rather than the attributes (derivation is described below in the section on modal split). The method of aggregation of costs is similar to Manheim's method described above. Both models are in concept extensions of the gravity model, discussed in an earlier section; the gravity model is taken specifically as a starting "analogy" (50) in McLynn and Woronka's derivations.

In sum, the travel model that satisfies both the utility-maximizing (rational) travel behavior premise and the independence axiom is the (separable) multinomial probability or share model (Eq. 6). But to use the separability property of the independence axiom to reduce the number of choice alternatives and allow calibration of separate, less complex models that may later be combined requires the assumption of additive utilities from sequentially made travel choices to estimate sequential choice models that may later be combined into 1 multinomial, multivariate probability or share model.

Figure 6. Conventional UTP travel-forecasting chain.



Separate Travel-Choice Models

The conventional series of (aggregate) sequential choice travel forecasting models are usually chained, as shown in Figure 6. Current travel forecasting procedures that predict quantity of travel on transportation networks are based on the theory of equilibrium between supply and demand on the transportation network. That is, there should be an equality between the travel conditions, such as times and cost, on the loaded network and the travel conditions used as input to the prediction. As shown in Figure 6, the current conventional procedure is to model travel behavior as a series of sequential, independent choices of trip generation, trip distribution, modal split, and traffic (route) assignment. Land use forecasting precedes travel forecasting as a separate step. For each travel choice, the existing pattern of usage in the region at the prevailing equilibrium between supply and demand is related to a small set (often one) of independent variables. The trend or description is then assumed to hold in the future.

For example, trip distribution is modeled as a function of a description of the trip lengths that prevailed at the equilibrium between supply and demand represented in the base-date data file. Trip generation usually relates total trips in and out of a zone only to measures of the activities existing in the zone. The assumption is made that total travel, as measured by trip ends, varies only as development varies, not as conditions on the tested networks change.

In a single pass through the chain shown in Figure 6, the initial number of trips generated is kept constant, regardless of what happens later in the chain. Iteration is the conventional method of feeding back the effects of changes in travel conditions lower in the chain on forecasts made higher or earlier in the chain in order to equilibrate between supply and demand on the transportation network. The difficulties of introducing "lower down" choice attributes higher in the chain is well known, in part because of the incomplete and irregular specification of choice variables (e.g., transportation system attributes) in each step (7, 45). The way to overcome this problem is through parametric aggregation, as already discussed (assuming constant relative valuation of choice attributes throughout a complete travel decision).

The next sections discuss existing models of travel choice. Each is taken individually, except trip generation, which is discussed above and in the introduction. Issues of combining models into one demand model are not discussed.

Trip Distribution

The gravity model was shown earlier to be derivable from a general statement of proportionality to attributes of a constrained-choice situation, i.e., constrained in the sense that these attributes included the constraining (previously calculated and held constant) trips generated and attracted. The attributes can potentially include all the attributes of travel (disutility) between origins and destinations. Solving for the constant of proportionality results in the multinomial share model. This may be the simplest possible statement of the multinomial model as the logical result of assuming rational choice (transitive values) throughout the travel decision.

The gravity model was also shown earlier to be analytically derivable from considerations of maximizing personal utility (21), assuming destinations expressed the utility of the trip (purpose identical with destinations).

Wilson (81) derived the gravity model as the "most probable distribution of trips among zones" given the usual assumptions that the numbers (i.e., frequency) of trips generated from, and attracted to, each zone are fixed (constant) and the total "generalized" cost of travel is held constant. He later attempts to embellish this very interesting result by showing its consistency with maximizing entropy (82).

Loubal and Potts (40) derive a trip-distribution model that is equivalent to the exponential form of the gravity model and assumes that a "trip potential, giving an expected number of trips in the absence of resistance to travel can be combined with a correction term dependent on network constraints." Two of the initial assumptions made are the same as Wilson's (81); namely, trips to and from each zone are constant and known. However, Wilson's assumption that the total generalized cost of travel is constant is dropped. The model is derived on the basis of probability statements whose normalization properties allow the model to be applied with different zone configurations provided that "network parameters are adjusted with appropriate weight factors."

Wilson (81) also derives the intervening-opportunities model (63) by the same methods and from the same assumptions plus one, namely, that intervening opportunities are a proxy for cost. That is, "the number of opportunities passed (so far are) a measure of the cost of getting so far" (81). The total opportunities passed sum to total trip-end destinations, which are assumed fixed, as is total cost. The derivation provides an interesting equivalence statement between opportunities and cost of travel. If we assume that there is some utility derived from the purpose of travel, the statement says that the number of opportunities passed is minimized in order to maximize net benefit from travel. Thus maximizing net benefit from travel means minimizing the number of destinations passed. The L in the opportunity model, which is supposed to be a constant probability of accepting a given destination, can then be interpreted as a parameter, estimated on the basis of minimizing destinations passed, or trip cost, both of which are now considered equivalent. The (constant) parameter suggests that the value attached to trip cost (its "marginal utility") is constant.

Modal Split

There are numerous derivations from different "first principles" of the multiple-choice share model. Wilson (81) derived it (Eq. 7) in its aggregate form (P_i in Eq. 7 equals the split fraction on the ith mode) by using the method and assumptions for deriving the gravity model, adding the restriction that the cost of travel among all zones over all modes is fixed. He notes that the function (Eq. 7) is "identical in form to that derived from a statistical approach to modal split using discriminant analysis" (59). Warner (79) and later others (39, 69, 73) also use the probabilistic formulation (Eq. 7) or its equivalent as fitting functions to estimate the probability of mode choice in the so-called disaggregate probabalistic behavioral models, as noted previously. One of the principal interests of the latter group is to estimate the value of time from a binary-choice probit model or a strict-utility function, rather than to analytically derive a new demand model.

The next attempt to analytically derive the share formulation (Eq. 6) in modal split from some statement of first principles is by McLynn, Goldman, Meyers, and Watkins (49). Their model (Eq. 39) is derived analytically from assumptions only on the split fractions of each mode. The first assumption, or statement of behavior, is quite familiar: "The split fractions which define the share of the market are assumed to be functions of the choice influencing attributes of all the competing products." The split fractions, of course, express the aggregate result of modal-choice behavior. The choice influencing attributes are represented by a vector, X_{kj}, where j is the mode and k is the variables (attributes) describing the choice situation. According to the authors, "X_{11} need not have the same interpretation as X_{12}, and might even refer to some quality of (j = 1) that is meaningless for (j = 2)." The authors next define terms.

$$M_j = w_j M \tag{41}$$

where

M = total market size,
M_j = size of j's market share, and
w_j = j's fraction of the total market.

They then decompose Eq. 41 by differentiating in the usual fashion to derive the separate (additive) elasticities.

$$E_x(M_j) = E_x(M) + E_x(w_j) \tag{42}$$

where $E_x(\ )$ = elasticity with respect to the attribute x of the term in the argument ( ).

They then focus separately on the $E_x(w_j)$, the elasticity with respect to the X_{kj} of the market share or split fraction of mode j. The assumption that actually specifies the form of the model is that the elasticities of the split fractions, with respect to the attributes X_j, are a function only of the split fractions themselves. (This appears to be a general result, as well as a possible starting assumption, for multiple-choice share models—a result worth pondering.) That is,

$$E_{x_j}[w_j] = f(w_j) \tag{43}$$

$$E_{x_{i \neq j}}[w_j] = f(w_i) \tag{44}$$

The latter (Eq. 44) are the cross elasticities that depend only on the split fractions of the competing modes, not on the attributes X_j. This leads to possible nonmode specificity of the cross elasticity of the P_i or w_j (share) as discussed in the section on defining alternative travel choices.

After a lengthy and rigorous mathematical derivation, Eq. 39 results ($P_i = w_j$). The model is calibrated with aggregated data on market shares and mode-specific variables X_{kj}.

The derivation of the model from these simple assumptions is indeed an elegant piece of work. Unfortunately the assumptions offer no particular basis in "behavior" that is helpful in specifying appropriate X_{kj} variables. That is, the traveler is logically confronting situations described by the X_{kj}, not the direct and cross elasticities. One wishes to tie the assumptions back into statements of choice behavior that are more easily interpreted.

There follows a spate of additional derivations of the multiple-choice share model in transportation, and these should be mentioned. The first (48) follows from the Luce (41) independence-of-irrelevant-alternatives axiom. McFadden uses the statement of the independence condition (Eq. 4) to derive Eq. 7 by using the properties that the $h(Z_{ki})$ are proportional to the odds that i will be chosen and the sum over i of the $h(Z_{ki})$ must equal one. This is the same as McLynn's first quite general statement, in the form

$$h(Z_{ki}) = f(X_{kj}) \tag{45}$$

Townsend (74) derives the multiple-choice share model axiomatically from transitivity and continuity statements that are quite independent of, but analogous to, Luce (41). Mayberry (47) claims that Eq. 6 is "equivalent" to a statement that says that an increase in attractiveness of mode m (with other modes unchanged) will cause travel on m to increase and travel on all other modes to decrease. A decrease in attractiveness of mode m would cause the opposite behavior. This is, of course, nothing other than simple scalability. Mayberry worries aloud that his statement has entailed too large an assumption because Goldman pointed out to him the "problem" with Eq. 6: "The ratio of travel by one mode to travel by another depends only on the characteristics of those two modes, and not on the characteristics of any other mode!" (47). However, after worrying about the problem of not always being able to describe independently perceived modes within the abstract mode formulation (is a flying blue bus a bus or an airplane?), Mayberry is apparently satisfied that "homogeneous population groups" will make the distinction and continues his axiomatic development of Eq. 6.

Rassam, Ellis, and Bennett (60) derive independently an exponential-form multinomial logit model (Eq. 7) from 2 assumptions similar to those made before. Their first assumption (similar to that of Mayberry) is that, if the attractiveness of a mode is expressed by a disutility function, which includes transportation variables, "then the share of that mode decreases when any of its transportation variables increase and, ceteris paribus, those of the other modes will increase or remain stationary." The second assumption (similar to that of McLynn et al., 49) "structure(s) the relationship between modal split and the explanatory transportation variables, namely, that the ratio of a small change in modal split of a given mode to that of a given transportation variable is proportional to the modal split of this mode and to a linear function of the modal splits of all modes." This statement is expressed as

$$\frac{\partial w_m}{\partial X_{ij}} = w_m \sum_{k \neq m} \alpha_{ijmk} w_k \tag{46}$$

where X and w are as defined in Eqs. 43 and 44, i and j are origins and destinations, and k and m are modes. [This equation, which is in Rassam, Ellis, and Bennett's Eq. 5 (60), is the linear case of McLynn's et al. Eqs. 1.17 and 1.18 (49).] The usual set of assumptions and restrictions is made (choices are mutually exclusive and define the full alternative set, the sum of the shares, i.e., modal splits, equals one and so on) and the resulting system of differential equations is solvable as a series of exponential-form share equations (Eq. 7). The Z_{ki} in Eq. 7 is a linear function of the attributes X:

$$Z_{ki} = \sum_{i} \alpha_{ik} X_i + \alpha_k \tag{47}$$

The constant term, α_k, is again a mode (choice) specific constant that contains the effects of all attributes or purposes or both not considered or measured. Equation 7 has been used successfully in estimating the split among 4 modes to airports in the Washington, D.C., area.

Because the fundamental assumptions are the same as those of McLynn et al. (49), the same comments apply as to that model—namely, that, although the assumptions can be shown to result in a multinomial share model, they are not grounded in behavior in a way that is helpful in specifying variables. Utility analysis is much more helpful in this regard. However, because the models are the same as those derivable from utility analysis, consideration of personal utility from travel can be used in specifying variables to be used in this share model of modal choice. Specification of appropriate attributes in each choice situation is a critical issue in the aggregation of separately calibrated choice models.

Pratt and Deen (55) fitted a logit function to aggregate sub-modal-split data in Washington, D.C. (In this case, submodal split is intratransit-mode diversion from surface bus to rapid transit.) In that application, they state,

> The final equivalent time diversion curve was formulated by first applying regression analysis and then hand fitting a logistics curve to the data points. The resultant submodal split relationship can be expressed by
>
> $$y = \frac{100}{1 + e^{-0.3X}} \tag{48}$$
>
> where X is the equivalent time saving via rail (equivalence factor—weighting factor for out of vehicle time—of 2.5) and y is the percent using rail. Weighting each data point by the number of observations, the R^2 of the curve is 0.886. This R^2 value is computed by comparing predicted and actual percent submodal split on an interchange basis.

The aggregate "conventional" modal-split models familiar to us from the UTP process are excellently summarized by Fertal et al. (19) and Weiner (80). These models, whether or not they are post- or pre-distribution, fit the dependent variable (e.g., percentage of transit use) to some function of a set of variables describing the choice environment. The fitting is usually either eyeball smoothing of curves to plotted data (26) or linear regression fitting, necessitating the additional assumption that the effects of the independent variables on modal-split fractions in a linear regression equation are additive (20).

S-shaped hand-fitted modal-split curves often bear some resemblance to ogives (e.g., cumulative normal or logit curves), and the possible translation in concept to a probability model (e.g., logit) is clear. The linear regression equation can also be interpreted as a linear approximation to an ogive as long as it is appropriately bounded such that the dependent share, or probability of choice, is allowed only to vary between 0 and 1 (where probability is defined as the limit of the ratio of the number of outcomes of a given choice to all possible choices in a large number of trials, i.e., observations on individuals, in which the attributes of the choice situation are held constant).

Pratt (56) proposed a binary aggregate primary modal-choice model that was applied in Minneapolis-St. Paul (67). A disutility function is postulated for each modal alternative that transforms time, convenience, and dollar cost into a common unit of equivalent time (i.e., disutility). The differential weighting of various components of travel time was frequently used previously in coding transfer links in network analysis by Alan M. Voorhees and Associates (78) and probably by other organizations. Table 1 (67) gives the procedure. The weighting factors are drawn from a variety of previous modal-split and value-of-time studies. The disutility difference between automobile and transit is calculated for each interzonal pair, and the percentage of trips (between zones) using one mode, plotted as a function of this disutility, is assumed to follow the cumulative normal (probit) probability distribution. The resultant predictive curve is said to have its point of inflection at 50 percent probability on the y axis and zero-measured disutility difference on the x axis. This assumes, of course, that the choices are completely described by the disutility measures.

The problem with the model (i.e., the published versions) is that its analytic development appears to have stopped with the earlier Pratt and Deen (55) work. That is, the later work represents a conceptual, but not an analytic, translation in the concept of ogive resembling aggregate modal-split curves (26) to a probability model. Unfortunately, once the translation is made in concept, no effort is made to use the properties of the asserted normally distributed probability behavior in the calibration of a (probit) mathematical model, that is, a model with analytically estimated parameters (including a constant term that includes the effects of the left-out choice attributes) and significance tests on the variables and so on. The model continues to resemble the older hand-fitted diversion curves, but has the additional assumptions of additive and constant marginal rates of substitution of times and costs making up modal disutility.

Table 1. Variables and weighting factors in Twin-Cities marginal utility modal-choice model.

Variable	Symbol	Factor
Walk time to and from transit	T_a	2.5
Wait time for transit	T_w	2.5
Transit running time	T_r	1.0
Transit fare	F	1.0
Automobile terminal time	A_t	2.5
Automobile running time	A_r	1.0
Parking cost	P	0.5
Highway distance	D	4.0, 5.7[a]
Marginal utility[b]	U	–
Cost of time[c]	C	–

[a]Cost-per-mile factors rather than weights. For trips attracted to CBD, 5.7 cents/mile was used; for other trips, 4.0 cents/mile was used.
[b]Computation equation for marginal utility of automobile over transit for non-CBD trips: $U = 2.5(T_a + T_w - A_t) + (T_r - A_r) + (F - 0.5p - 4.0D)/C$.
[c]Computed as 25 percent of income: {[Annual income (cents/year)]/[(2,080 hours/year)(60 min/hour)]} x 0.25 = C.

Traffic Assignment

The first application of a multiple-choice share model in travel forecasting (aside from the gravity model, 77) appears to be by Traffic Research Corporation in route choice (traffic assignment). This route-choice model (29) was developed and applied in Toronto in the late 1950s.

$$(AF)_i = \frac{\left(\frac{1}{T_i}\right)^{a}}{\sum_{i}\left(\frac{1}{T_i}\right)^{a}} \tag{49}$$

where

AF_i = proportion of interzonal trips by mode assigned to route i, and
T_i = interzonal travel time on route i.

Time, T, only is used as the measure of route impedence (cost). The formula is in the form of Eq. 6, where the "assignment factor" for route i (AF_i) equals P_i in Eq. 6, and $h(Z_{ki}) = T_i^{a}$. The subscript k is dropped in Eq. 49 because there is only one (highway) mode being considered. T_i is the route (path) travel time for the i th route from the network. The parameter, α, was held constant over all paths. In effect, this is an "abstract-route" model, consistent with Eq. 39, where α_{ki} varies over modes k, and system variables (costs) i, both of which (k and i) equal 1 in this case. The value of the parameter was not mathematically fitted, but was selected on the basis of a reasonably proportional assignment to paths with Traffic Research Corporation's multipath capacity-restrained iterative assignment technique (29). Rapid settlement of volumes over (equilibration) iterations through trip generation, distribution, and so on was another fitting criterion.

The direct traffic-estimation method (64) is also a probability formulation. That is, the probability of a vehicle on a link finding a destination in the valid domain or set of destinations defined by the tree on which the link is located is inversely proportional to the further travel time (or impedance) to that destination. The derivation is similar to the previous gravity model derivation (Eqs. 8, 9, and 10). The resulting probability of accepting any destination, or destinations within a particular valid domain, is its fraction of the total domain integral. The domain integral, I_D, is defined as

$$I_D = \int_D F dV \tag{50}$$

where

I_D = domain integral;
F = some impedance function, e.g., $F = e^{-kt}$, where k = constant and t = travel time; and
V = set of destinations clustered around a point at which the function F has a definite value.

The probability of having a destination in a subregion R within the valid domain, n (e.g., north of the point on the link), is

$$P \text{ (destination in R)} = \frac{I_R}{I_n} \tag{51}$$

Only destinations within the valid domain have nonzero probabilities of being accepted, and the probabilities of accepting all destinations in the valid domain sum to one.

The probability expression, Eq. 51, is in the form of Eq. 6. The direct assignment technique calculates the appropriate domains for each point on each link of interest on the basis of shortest time (impedance) paths on the network and assigns traffic to links on the basis of Eq. 51 corrected for normalization and symmetry conditions. The direct traffic-estimation method uses practically the same inputs as conventional UTP models, namely, trip ends, an impedance function, and coded networks. It is advantageous in assigning travel to individual links. However, the method assumes complete symmetry in destination volumes and link and path loadings throughout the system, and capacity-constrained loadings are unavailable (22).

Dial (13) has developed a probabilistic multipath traffic-assignment model that uses Eq. 7 to calculate the probability of paths between origins and destinations. The model makes use of a 2-pass procedure that generates all "efficient" paths between origins and destinations and loads them simultaneously. Incremental loading in a capacity-restrained mode is allowed. Efficient paths are generally those that allow the traveler to make apparent progress toward his destination at every branch point (on the network). That is, that reduce the impedance between the traveler and his final destination.

The parallel with the Luce choice axiom is clear. Backtracking on the network in order to "come out ahead" is not ordinarily allowed. Such backtracking can be considered equivalent to decisions that are really "two or more intermediate decisions." These violate the necessary separability assumptions in the independence axiom because such decisions are not simple substitutes for other alternatives at that branch point (node). Such backtracking alternatives must somehow be combined in order that all relevant alternatives may be considered as substitutes for one another with nonzero probabilities of being chosen.

Dial's method appears to be completely general in the sense that any utility function may be used to calculate the probability of using any path (Eq. 7).

The model is a Markov model. At each node, the fraction (probability) of trips assigned to each alternate link (on an efficient path) is calculated based on the path impedance and the number of efficient paths through the link. The separability property of the multinomial formula (Eq. 6) is used (assumed) at every branch point. The use of Dial's method to apply multinomial, multivariate logit models to calculate the probability of any path through complete travel decision trees (e.g., Fig. 2) appears to have considerable promise. That is, the method could be used (applied as described earlier) to calculate the (path) probability of any (relevant) alternative combination of frequency, destination, mode, time of day, and route (having a nonzero probability of choice).