Resource Paper

EARL R. RUITER
Department of Civil Engineering
Massachusetts Institute of Technology

This paper has as its major goal the initial formulation of a research program on analytical structures for travel demand forecasting and the discussion of the motivations for this formulation.

By travel demand forecasting is meant the process of predicting the travel that will occur when a given transportation system is provided within a given activity system. (By activity system is meant all aspects of the world that are not parts of the transportation system, but that do have effects on that system.) This definition of travel demand assumes that we are looking at trip-making decisions only and, therefore, can ignore long-range changes in the activity system caused by travelers' changing their places of residence and work, except as those changes may be externally specified. The long-range changes in the activity system are left for the activity shift and land use modelers, although it is recognized that the transportation system is an important determinant of those long-range changes.

By analytical structure is meant 2 things: (a) primarily, the form of the travel demand forecasting function, whether it be a closed mathematical expression or an algorithm; and (b) to a lesser extent, the independent variables used in the forecasting process. More details and motivation for this definition are given later.

This paper is structured into 3 somewhat unequal sections. Section 1 includes extended definitions of demand models and analytical structures and a listing of some alternative structures that have been applied to the travel demand forecasting problem. Section 2 discusses the factors that must be considered in deciding on appropriate analytical structures for travel demand forecasting, and identifies a number of areas of necessary research. Section 3 brings all of these together as a concise initial formulation of a program of research in the area of analytical structures.

DEFINITIONS AND ALTERNATIVES

Analytic Definition

Because we are concerned with forecasting travel demand, it is useful to develop an analytic definition with a basis in consumer demand theory as it has been developed in the field of microeconomics (7). Beginning with the preferences of individual consumers, Henderson and Quandt postulate utility functions that state the level of utility associated with the purchase of quantities Q_i of a number of goods.

$$U(Q_1, Q_2, \ldots, Q_n) \tag{1}$$

Also, the consumer's budgetary limit is expressed as

$$\sum_{i=1}^{n} P_i Q_i \leq Y \tag{2}$$

where P_i is the price of the ith good, and Y is the total budget, or income, of the consumer. When U is maximized subject to the budgetary constraint, the following relations are obtained among the variables:

$$Q_i^\circ = D_i(P_1, P_2, \ldots, P_n, Y) \tag{3}$$

for all i, where Q_i° is the optimal quantity of good i purchased by a consumer with income Y. The functions $D_i(\cdot)$ are demand functions in the classical economic sense. They relate the quantity of a good consumed to the prices of all goods and to the income level of the consumer.

In theory all goods that contribute to the consumer's utility must be included in each demand function. Practically, however, it is impossible to find significant relations between the prices of many goods and the demand for others. We, therefore, group the subset of all prices that significantly affect the quantity of good i into a vector P. These prices include (a) the price of good i itself and (b) the prices of goods that are substitutes for good i.

Using the vector P, we can rewrite the demand function as follows:

$$Q_i^\circ = D_i(P, Y) \tag{4}$$

This equation represents the demand function for an individual. The summation of these functions to obtain total demand can be accomplished, at least theoretically, by assuming that individuals can be grouped into subsets of the total population with similar utility functions and income levels. Each subset can be described by socioeconomic variables, S, which include Y. This leads to the following functional form for total demand functions:

$$Q_i = D_i(P, S) \tag{5}$$

To adapt this general formulation to transportation demand, we must recognize that transportation is a good that is a complement to the demand for many other goods. Consumers travel to the corner to purchase bread; they travel downtown to purchase meals at restaurants; they travel to Florida to purchase sun in the winter; they travel to their working places to trade their labors for incomes. Transportation is therefore termed an intermediate good. Although it is a complement to many other goods, the quantity of transportation consumed does not contribute positively to the utility function, U. The demand for transportation is a derived demand: It is due to the demand for other goods rather than to its own contribution to the consumer's utility.

One approach to transportation demand forecasting, therefore, would be to model the demand for the final goods and services that result in transportation consumption. To date, however, this has proved to be too difficult. Instead, trips are typically classified according to trip purpose (class of final good), and the demand for transportation for each purpose is modeled separately. Also, an additional class of independent variables, measuring the attraction or intensity of the final activities, A, is added to the demand functions. Therefore, when the subscript i in Eq. 5 refers to a transportation good, V^n_{klm} (trips for purpose n from origin k to destination l by mode m), the general demand function becomes

$$V^n_{klm} = D^n_{klm}(P, S, A) \tag{6}$$

Another characteristic of transportation is that the traveler "pays" in a number of ways when he consumes transportation. There are a number of "prices" that include not only money paid but also time consumed, discomfort experienced, and risks endured. These and other prices can be classified together as level-of-service variables, L. The level-of-service variables have an added dimension not present in the prices, P. For each price, P_i, there exists a vector of level-of-service variables, $L_i = (P_i, t_i, c_i, s_i, \ldots)$ where P_i = price, t_i = travel time, c_i = comfort index, and s_i = safety index.

Our final general analytical expression of a travel demand function is obtained by substituting L for P:

$$V^n_{klm} = D^n_{klm}(L, S, A) \tag{7}$$

Equation 7 serves as the starting point for considerations of the analytical structure of travel demand forecasting techniques. It is useful to summarize the major ways in which this function differs from Eq. 5, the general demand formulation.

1. Because there are many costs associated with travel, monetary prices, P, are replaced by level-of-service variables, L.
2. Because transportation is a derived demand, travel must be predicted by trip purpose and must be a function of the activities, A, available at the destination.

The overall goal of this paper is to formulate a program of research that will lead to answers to the following questions:

1. What forms of the function D^n_{klm} are appropriate for various kinds of travel demand forecasting?
2. What variables belong in each of the sets of independent variables shown in Eq. 7?

Some Alternative Structures

Before discussing the factors that must be considered in answering the above questions, we should classify and list some of the major types of analytical structures for travel demand forecasting that have been developed to date. The purpose is not to include all existing forecasting procedures, but rather to illustrate each class of structures with a typical example. The general classes of procedures are sequential aggregate, direct aggregate, sequential disaggregate, and direct disaggregate. These classes are described in the sections that follow.

Sequential Aggregate

The urban transportation planning process (UTP) is a prime example of a set of sequential travel forecasting procedures. Because this process has been used so extensively for so many of the travel forecasts made for the past 15 years, it will be described very briefly here, with emphasis on the structural aspects.

Trip generation is the first sequential step, involving the prediction of total trips from an origin or to a destination by trip purpose (6). The independent variables are most commonly in the socioeconomic and activity classes used in Eq. 7. The functional form is usually linear. Symbolically,

$$T_i^n = \sum_1 b_1^n S_{i1} + k_1^n$$
$$T_j^n = \sum_1 c_1^n A_{j1} + k_2^n \qquad (8)$$

where

T_i^n = trips of purpose n generated in origin i,
T_j^n = trips of purpose n attracted to destination j, and
b_i^n, c_i^n, k_i^n = empirical parameters.

Typical socioeconomic variables used are average annual income, average number of automobiles owned, number of workers per household, and percentage of households having an income greater than a specified value. Typical activity-system variables used are zonal population, acres of land in various land use categories, and zonal employment.

The second sequential step is trip distribution, the prediction of trips from origin to destination. The independent variables are the trip ends resulting from the previous step plus level-of-service variables. Symbolically,

$$T_{ij}^n = f_n(T_i^n, T_j^n, L_{ij}) \qquad (9)$$

where

T_{ij}^n = trips of purpose n from origin i to destination j,
T_i^n, T_j^n = results of the trip generation step, and
L_{ij} = level-of-service variables between i and j.

The 2 most common functional forms are the gravity model and the opportunity model. A typical version of the gravity model is as follows:

$$T_{ij}^n = T_i^n \frac{T_j^n t_{ij}^{\beta_n}}{\sum_k T_k^n t_{ik}^{\beta_n}} \qquad (10)$$

where

t_{ij} = travel time from i to j, and
β_n = empirical parameter.

A typical version of the opportunity model is as follows:

$$T_{ij}^n = T_i^n e^{-L_n V_j^n}\left(1 - e^{-L_n T_j^n}\right) \qquad (11)$$

where

$V_j^n = \Sigma T_k^n$ = "subtended volume,"
k = all destinations for which $t_{ik} < t_{ij}$, and
L_n = empirical parameter.

These models are "share" models; they divide the total trips from i, T_i^n, among all destinations by using a fraction that, when summed over all destinations, equals 1. Travel time by a single mode, usually highway, is typically the only level-of-service variable used although, in some applications, a generalized cost has been used that is a linear combination of travel time, distance, and out-of-pocket costs. The level-of-service variable enters the opportunity model in an indirect way only. It affects the ranking of destinations from each origin, which in turn affects the subtended volumes that enter the model directly.

In some applications of both the gravity and the opportunity models, adjustments of T_j^n are made after initial application of Eq. 10 or 11 in an attempt to force the total trips to each destination ($T^n_{\cdot j} = \sum_i T^n_{ij}$) to equal the original T_j^n. This constraint, though logical, is not guaranteed by the functional form of either distribution model. Following adjustments of the original T_j^n, the equations are applied again. Iteration through application of the equations and adjustment of the original T_j^n continue until a desired level of correspondence between each T_j^n and $T^n_{\cdot j}$ is reached.

The third sequential step is modal split, the prediction of trips by mode from origin to destination. The independent variables are the trip interchanges resulting from the previous step plus modal level-of-service variables. Symbolically,

$$T^n_{ijk} = f_{nk}(T^n_{ij}, L_{ijm}, S_i, A_j) \tag{12}$$

where

T^n_{ijk} = trips of purpose n from origin i to destination j by mode k,
T^n_{ij} = results of the trip distribution step,
L_{ijm} = level-of-service variables for all modes m between i and j,
S_i = socioeconomic variables of travelers in i, and
A_j = activity-system variables in j.

Many approaches have been used to develop functional forms, f_{nk}, for modal-split models. The most commonly used prior to the past 3 or 4 years were regression or table look-up models based on the relative levels of service offered by each mode (4). Typically, origin zones have been classified by income level and automobile ownership, and for each subgroup linear equations or tables are developed that relate fraction of trips by automobile and transit to time and cost ratios or differences. Symbolically,

$$P^n_{ijk} = \frac{T^n_{ijk}}{T^n_{ij}} = f^m_{nk}\left(\frac{t_{ija}}{t_{ijt}}, \frac{c_{ija}}{c_{ijt}}\right) \tag{13}$$

or

$$P^n_{ijk} = \frac{T^n_{ijk}}{T^n_{ij}} = g^m_{nk}\,(t_{ijt} - t_{ija},\ c_{ijt} - c_{ija})$$

where

P^n_{ijk} = fraction of travel for purpose n between i and j by mode k,
t_{ija}, t_{ijt} = travel times by automobile and transit,
c_{ija}, c_{ijt} = costs by automobile and transit, and
m = income and automobile ownership group.

Various time and cost variables have been used, and often more than one of each has been used. Time has been divided into in-vehicle time, waiting time, and access time, for example. Cost has been divided into out-of-pocket cost, tolls, parking fees, fares, and total operating costs.

More recently, the following functional form has been used for f_{nk} (17, 20):

$$P^n_{ijt} = \frac{1}{1 + e^{h_k(L_{ijm})}} \tag{14}$$

and

$$h_k(L_{ijm}) = C_k + \sum_1 a^1_k(t^1_{ijt} - t^1_{ija}) + \sum_1 b^1_k(c^1_{ijt} - c^1_{ija})$$

Again, times and costs have been divided into various variables. The constant C_k, as well as the parameters a^1_k and b^1_k, allows the relative characteristics of modes not measured by times and costs (such as comfort, convenience, and modal "image") to be represented in the model. The function h_k can be interpreted as a difference in consumer utility between travel by transit and travel by automobile.

Direct Aggregate

In contrast to the sequential application of a number of models in the UTP process, direct aggregate procedures involve the prediction of travel demand by origin, destination, and mode with a single equation whose general form is given in Eq. 7. The original application of these procedures has been to the prediction of intercity trips between large zones, typically entire urban areas. More recently, application to urban areas has taken place. Functional forms that have been used for direct aggregate equations may be placed in the following major groups.

Independent Mode-Specific Equations

$$T^n_{ijk} = f^n_k(L_{ijm}, S_i, S_jA_i, A_j) \tag{15}$$

In the present models of this type, 3 forms of the function f^n_k are most common.

1. The product form (21) was applied to intercity travel for business and personal purposes.

$$T^n_{ijk} = a^n_{ok}\, P_i^{a^n_{1k}}\, P_j^{a^n_{2k}}\, Y_i^{a^n_{3k}}\, Y_j^{a^n_{4k}} \prod_m \left(c_{ijm}^{b^n_{mk}}\, t_{ijm}^{c^n_{mk}}\right) \tag{16}$$

where

P_i, P_j = populations,
Y_i, Y_j = average incomes,
c_{ijm} = travel costs by mode m, and
t_{ijm} = travel times by mode m.

2. The linear-log form (5) was applied to automobile work trips in a metropolitan area. The socioeconomic and activity-system variables are labor force at origin, employment at destination, median income at origin, and number of automobiles per person at origin. The level-of-service variables for both automobile and transit are in-vehicle travel time, out-of-vehicle travel time, line-haul cost, and out-of-pocket cost.

$$T^n_{ijk} = M^n_k(S_{io}A_{jo})\left(\sum_{m,l} a^n_{kml} L_{ijml} + \sum_{m,l} b^n_{kml} \ln L_{ijml} + \sum_l c^n_{kl} S_l + \sum_l d^n_{kl} \ln S_l\right) \tag{17}$$

where

M^n_k = constant term,
l = variable number,
S_{io}, A_{jo}, S_l = socioeconomic and activity-system variables, and
L_{ijkl} = level-of-service variables.

3. The product-exponential form (5) was applied to automobile shopping and transit work trips in a metropolitan area. The activity-system variables in the model for automobile shopping trips are number of households at origin, number of persons per household at origin, median income at origin, number of automobiles per person at origin, and density of retail trade employment at destination. The level-of-service variables for the automobile shopping-trip model include all listed for the linear-log

form, with the exception of out-of-vehicle travel time for the transit mode. For the transit work-trip model, the activity-system variables were the same as those used in the linear-log form. The level-of-service variables included no automobile model variables. The transit variables used were the same as those for the linear-log form.

$$T^n_{ijk} = M^n_k \prod_{m,l} L_{ijml}^{a^n_{kml}} e^{b^n_{kml} L_{ijml}} \cdot \prod_{l} S_l^{c^n_{kl}} e^{d^n_{kl} S_l} \tag{18}$$

where the variables are as defined for Eq. 17.

Independent Mode-Abstract Equations

$$T^n_{ijk} = f^n(L_{ijm}, Y_i, Y_j, A_i, A_j) \tag{19}$$

This general representation only differs from Eq. 13 in that the function f^n is independent of mode, k. The prime example of this model is the following form developed by Quandt and Baumol (18). Because it was developed for intercity travel for all purposes, no purpose superscript is used.

$$T_{ijk} = a_o P_i^{a_1} P_j^{a_2} Y_i^{a_3} Y_j^{a_4} c_{ijb}^{a_5} t_{ijb}^{a_6} f_{ijb}^{a_7} \left(\frac{c_{ijk}}{c_{ijb}}\right)^{a_8} \left(\frac{t_{ijk}}{t_{ijb}}\right)^{a_9} \left(\frac{f_{ijk}}{f_{ijb}}\right)^{a_{10}} \tag{20}$$

where f_{ijk} is the frequency of service; and the new variables, c_{ijb}, t_{ijb}, and f_{ijb}, are the cost, time, and frequency for the "best" mode with respect to each parameter: the cheapest cost, the fastest time, and the most frequent service.

A distinct advantage of a mode-abstract direct demand equation is its ability to predict the demand for new modes without changing the functional form of the model or its parameters.

Modal Share Models

$$T^n_{ijk} = f^n_2(A_i A_j Y_i Y_j L_{ijm}) \frac{f^n_{1k}(L_{ijk})}{\sum_m f^n_{1m}(L_{ijm})} \tag{21}$$

As the general form of this model indicates, these models include 2 separable functions: one to predict total trips from i to j(f^n_2) and a second to predict the share of these trips that will use mode k(f^n_{1k}). Therefore, this model can be classed as a direct aggregate model or as a partially sequential model.

The prime example of this model is McLynn's composite analytic model developed for intercity travel for all purposes (15). In that model, the function f_{k1} and f_2 are as follows:

$$f_{1k} = a_{ok} c_{ijk}^{a_{1k}} t_{ijk}^{a_{2k}} f_{ijk}^{a_{3k}} \tag{22}$$

$$f_2 = b_o P_i^{b_1} P_j^{b_2} Y_i^{b_3} Y_j^{b_4} \left(\sum_m f_{1m}\right)^{b_5} \tag{23}$$

The 2 functions are typically estimated sequentially: First the f_{1m} functions are estimated, and then their sum is obtained as a variable to be used in the estimation of f_2.

Both of the analytical structures discussed above have been developed and applied to aggregated travel data: data for entire zones whose sizes range from fractions of square miles for urban applications to entire metropolitan areas for intercity applications. Modeling at either of these levels of aggregation smoothes out most of the variations of the individuals who actually make the travel decisions being modeled. For this reason, much of the recent demand modeling effort has addressed the problem of predicting the travel decisions of individual travelers. Initially, these studies were concerned only with the mode-choice decision. The models developed were individual traveler applications of the utility model form shown in Eq. 14 (9, 10, 22, 23). When applied to individuals, the dependent variable can only take on the values 0 or 1, requiring a different set of estimation procedures to be used. In the initial models of this type, only 2 modes were included, leading to a binary-choice situation. More recently, multiple-choice models have been developed (19).

Building on the earlier work in modeling the individual mode-choice decision, researchers have developed equations to model not only mode choice but also destination choice and the choice of whether to make a trip.

Charles River Associates (3) developed a sequence of individual choice models based on the assumption that travelers first choose whether to travel, then where to travel, then what time to travel, and finally what mode to use. Because of this assumed sequence of choices and the use of inclusive prices, the models must be calibrated in the reverse order of the assumed order of choice. They are presented in that order here.

1. The modal-choice submodel is based on a binary choice between automobile and transit.

$$\frac{P^n_{ija}}{1 - P^n_{ija}} = \exp\left[a^n + \sum_1 b^n_1(L_{ija1} - L_{ijt1}) + \sum_1 c^n_1 S_{i1}\right] \tag{24}$$

where

P^n_{ija} = fraction of trips by purpose n (work or shopping) by household i to destination j made by automobile rather than transit,
L_{ija1}, L_{ijt1} = automobile and transit level-of-service variables, and
S_{i1} = socioeconomic variables.

The socioeconomic variables are automobiles per worker in the household, indicator for race, and indicator for occupation. The level-of-service variables are waiting time (assumed to be 0 for automobile trips), in-vehicle travel time, and operating, parking, and fare costs.

2. The time-of-day-choice submodel is based on a binary choice between traveling in both directions during off-peak hours for shopping or traveling in at least one direction during a peak hour. The shopping purpose is the only one modeled.

$$\frac{P_{ijo}}{1 - P_{ijo}} = \exp\left[a + b(IP_{ijo} - IP_{ijp}) + \sum_m c_m S_{im}\right] \tag{25}$$

where

P_{ijo} = fraction of shopping trips made by household i to destination j completely during off-peak periods,
S_{im} = socioeconomic variables,
IP_{ijo}, IP_{ijp} = inclusive prices for off-peak and peak shopping trips,
IP_{ijo} = $\sum_1 b^n_1 L_{ijk1o}$,
b^n_1 = parameters from Eq. 24, and
L_{ijk1o} = level-of-service variables for the mode used during off-peak travel.

IP_{ijp} is similarly defined for peak-hour shopping trips. The socioeconomic variables used are indicator for sex of the head of household, number of workers per number of residents in the household, and number of preschool children in the household.

3. The destination-choice submodel is based on a multiple-option choice of traveling to each of a number of destinations for shopping. The shopping purpose is the only one modeled.

$$\frac{P_{ij}}{P_{im}} = \exp[a_1(IP_j - IP_m) + a_2(A_j - A_m) + a_3(IP_j \cdot S_i - IP_1 \cdot S_i)] \tag{26}$$

where

P_{ij}, P_{im} = fraction of shopping trips to destinations j and m by household i,
A_j, A_m = activity-system variables for destinations j and m,
S_i = socioeconomic variable for origin i,
IP_j, IP_m = inclusive prices for shopping trips to j and m,

$$IP_j = \sum_1 b_1^a L_{ija1},$$

b_1^a = parameters from Eq. 24, and
L_{ija1} = level-of-service variables for automobile trips to destination j.

IP_m is similarly defined for trips to m. The activity system variables are the fraction of total retail employment occurring in each destination. The socioeconomic variable, used with the inclusive price in the interaction term, is the number of preschool children in household i. No level-of-service variables for transit trips were used.

4. The trip-frequency-choice submodel is based on a binary choice between making 0 or 1 shopping trip per day. The shopping purpose is the only one modeled.

$$\frac{P_i}{1 - P_i} = \exp(a_1 IP_i + a_2 IE_i + a_3 Y_i) \tag{27}$$

where

P_i = probability that household i will make a shopping trip,
Y_i = family income of household i,
IP_i = inclusive price to household i = $\sum_j IP_j P_{ij}$, and

IE_i = average shopping opportunity = $\sum_j A_j P_{ij}$.

IP_j, P_{ij}, A_j, and P_{ij} are obtained from Eq. 26.

Direct Disaggregate

The set of equations presented above is the disaggregated analog of the UTP sequential process. A disaggregated analog of the direct aggregate models also has been postulated and calibrated (2). The functional form of this model is as follows:

$$\frac{P_{ijk}}{P_{ij'k'}} = \exp\left[\sum_1 a_1(A_{j1} - A_{j'1}) + \sum_1 b_1(M_{k1}^1 - M_{k'1}^1) + \sum_1 c_1 Y_i(M_{k1}^2 - M_{k'1}^2) + \sum_1 d_1(L_{ijk1}^1 - L_{ij'k'1}^1) + \sum_1 \frac{e_1}{Y_i}(L_{ijk1}^2 - L_{ij'k'1}^2)\right] \tag{28}$$

where

P_{ijk}, $P_{ij'k'}$ = fraction of total trips from household i going to destinations j and j by modes k and k′ (either j and j′ or k and k′ may be the same, but not both),

A_{j1}, $A_{j'1}$ = activity-system variables,
M^n_{k1}, $M^n_{k'1}$ = modal variables,
L^n_{ijk1}, $L^n_{ij'k'1}$ = level-of-service variables, and
Y_1 = household income variable.

As estimated by Ben-Akiva, the following variables were used:

1. Activity-system variables, A_{j1}—number of jobs in wholesale and retail establishments in the zone of destination j and indicator for CBD destinations;
2. Modal variable in separate term, M^1_{k1}—indicator for automobile usage;
3. Modal variable in interaction term with income, M^2_{k1}—indicator for automobile usage;
4. Level-of-service variables in separate terms, L^1_{ijk1}—out-of-vehicle travel time and in-vehicle travel time; and
5. Level-of-service variable in interaction term, L^2_{ijk1}—out-of-pocket cost.

This model was calibrated for automobile and transit trips for the shopping purpose only and does not deal with trip-making or time-of-day choices. It, therefore, represents a model that can be used to divide total shopping trips from a household among the available modes and destinations.

This concludes a brief survey of the major classes of analytical structures that have been applied to travel demand forecasting or proposed for application. In later sections, I will refer to these structures to illustrate the issues involved in the choice of an appropriate analytical structure for a given travel forecasting problem.

FACTORS AFFECTING ANALYTICAL STRUCTURE

Two questions were posed as the overall goal of a program of research to be developed by this workshop:

1. What forms of the function D^n_{k1m} are appropriate for various kinds of travel demand forecasting?
2. What variables belong in each of the sets of independent variables shown in Eq. 7?

The factors discussed below must be considered in answering these questions.

Travel Demand Theories

Theoretical constructs that can be applied to travel demand are available in 2 general fields: economics and psychology. We have drawn on classical demand theory to develop a starting point for our definition of the analytical structure of travel demand forecasting. This discussion includes not only the basics of classical theory but also the adjustments and extensions that make possible its application to travel demand.

Other theoretical developments can be analyzed in the same way. This is done in this section for the alternative approach to consumer theory developed by the economist Lancaster and for the behavioral theory of choice developed in psychology. (The resource paper for Workshop 5 should be referred to for a more complete discussion of the theories underlying travel demand forecasting.)

As stated by Lancaster (8), the following assumptions, each of which differs from the classical theory, are the essence of his approach:

> 1. The good, per se, does not give utility to the consumer; it possesses characteristics, and these characteristics give rise to utility.
> 2. In general, a good will possess more than one characteristic, and many characteristics will be shared by more than one good.
> 3. Goods in combination may possess characteristics different than those pertaining to the goods separately.

When the nature of transportation as a derived demand with many "prices" is considered, the relevance of Lancaster's approach to travel demand becomes evident. Transportation is a good with a number of characteristics that give rise to disutility, but is nevertheless consumed in combination with other goods because it makes possible the consumption of those goods. The other goods have 0 utility until they can be reached; then they provide utility that exceeds the disutility of transportation.

Without going any deeper into Lancaster's approach than the 3 assumptions quoted above, I shall provide a theoretical basis for expanding the single-valued price of classical economics to a vector of characteristics—the level-of-service variables—and for including measures of the activity system. This can be shown by developing the analog of Eqs. 1, 2, and 3, which arise from Lancaster's approach.

Utility functions now state the level of utility associated with the purchase of the quantities Z_i of a number of characteristics.

$$U(Z_1, Z_2, \ldots, Z_n) \tag{29}$$

These characteristics are obtained by engaging in a number of activities, j, each at level W_j. The relation between the vector of characteristic quantities, Z, and the vector of activity levels, W, is

$$Z = BW \tag{30}$$

where B is a matrix of elements b_{ij}, each of which is the amount of characteristic i provided per unit of activity j.

The amount of each good, k, consumed is Q_k, which depends on the consumption of goods in each activity, as represented by the following relation between the vector of goods consumed, Q, and W:

$$Q = AW \tag{31}$$

where A is a matrix of elements a_{kj}, each of which is the amount of good k consumed per unit of activity j.

As in the classical theory, a budget constraint exists. In matrix notation,

$$PQ \le Y \tag{32}$$

If U could be maximized subject to the constraints shown in Eqs. 30, 31, and 32, the following relations would be expected:

$$Q_k^\circ = D_k(P, Y, W, A, B) \tag{33}$$

Although Lancaster provides no general solution in terms of forms of the demand function $D_k(\cdot)$, he does discuss a number of implications of his approach. As an example, Eq. 33 provides a theoretical base for including measures of each of the following in demand functions in general and in travel demand functions in particular:

P = prices of goods,
Y = income level of the consumer,
W = activity levels of the consumer,
A = consumption of goods per unit of activity, and
B = provision of characteristics per unit of activity.

A second implication occurs when a new good, such as a new mode of transportation, is considered. In the classical theory, this situation requires the reformulation of the utility function, U, in an additional dimension before estimates can be made of the effects of this new good on the former equilibrium state. Before the new good is available, there is no way to estimate the changes to the utility function. Because in Lancaster's approach the utility function is dimensioned by characteristics rather than goods, it remains unchanged when new goods are added. To revise the demand functions, therefore, if no new activities are expected, requires only adding to the dimensions of Q, A, and P. Because Q and P are variables, only a new row of coefficients of A must be determined, based on the amount of the new good that is consumed in each of the activities. This is a much more straightforward task than formulating a new utility function based on consumers' responses to a situation that does not yet exist.

In many cases, a new good may result in new activities. This can also be represented by expanding the dimensions of A, B, and W. New columns must be added to A and B to represent the consumption of goods and production of characteristics of these new activities. This also can be done much easier than adding a dimension to the utility function.

In summary then, Lancaster's approach provides a number of bases for travel demand forecasting that are not provided by the classical theory. This added power has been recognized by a number of travel demand model developers. Others have gone beyond classical theory in ways that can only be supported by Lancaster's approach. His approach, therefore, can probably be profitably explored further by demand model developers.

One attempt to explore this approach has sought to formulate a general equilibrium model that adapts Eqs. 29, 30, 31, 32, and 33 to transportation (3). This is done by concentrating on the following classes of goods: transportation, consumer goods with fixed locations in the short run (work, home), and consumer goods available at many alternate locations (groceries, entertainment).

Although no tractable solution has been obtained with this formulation, 3 types of further work may be warranted.

1. Continue searching for a utility function form that results in a closed-form solution in terms of demand functions, $D_k(\cdot)$, for the transportation variables;
2. Continue exploring the existing formulation, as far as it has been developed, for its implications on suitable analytic structures; and
3. Search for realistic revisions of the formulation that will result in useful demand functions.

Both in the classical theory of the consumer and in Lancaster's formulation, only monetary prices are considered. Lancaster deals with multiple characteristics, but only price has a budget limit. In transportation demand work, it is often useful to consider time as a price also and to recognize that each traveler has a limited budget of time available for transportation or, in general, for the consumption of all goods. It is desirable, therefore, to expand Eqs. 2 and 32 to include a time budget that must be greater than or equal to the time used in consuming each good or in carrying out each activity. This added constraint can be expected to be more important for transportation demand analyses, where alternatives can have significant time variations, than for general demand modeling.

In the area of psychology, a theory of rational choice behavior has been developed (11). Its basic assumptions are that a decision-maker can rank possible alternatives in order of preference and will always choose from the available alternatives the option that he considers most desirable. These assumptions lead to the specification of utility functions that measure the desirability of an alternative, i, to a decision-maker with characteristics S_j.

$$U(Z_i, S_j) \tag{34}$$

where

Z_i = vector of attributes of alternative i, and
S_j = vector of characteristics of decision-maker j.

The decision-maker maximizes his utility by choosing the alternative with the highest value of the function; or, in the case of random variables, the decision-maker chooses the alternative for which his utility is maximized with some probability, P_i.

To make probabilistic choice models tractable, an axiom on choice behavior developed by Luce is often used. Termed the independence-of-irrelevant-alternatives axiom, it requires that the relative odds of 2 alternatives being chosen be independent of the presence or absence of third alternatives. Symbolically, if i and k are 2 alternatives, both of which are chosen part of the time, and if there exists another set of alternatives n_1, n_2, . . ., then

$$P_i/P_k = f(Z_i, Z_k, S_j) \tag{35}$$

and this function is not affected by the presence or absence of any of the alternatives n_1, n_2, ...

This is a critical axiom to accept because it has important benefits and costs. One benefit is that, in the modal-choice case, for example, it allows demand to be predicted for new modes before they are built, if all of the Z variables are based solely on generic attributes of the modes, such as travel time and cost. On the other hand, an important cost is that, when such a new mode is introduced, the reduction in usage of all existing modes will be a constant percentage. These characteristics do not exist when some of the Z variables are mode-specific (for example, a dummy variable that is 1 for the transit mode and 0 otherwise). This, however, is equivalent to replacing Z_i and Z_k in Eq. 35 with Z_{in} and Z_{kn}, which implies rejection of the independence-of-irrelevant-alternatives axiom.

The theory of rational choice behavior provides a powerful tool for the development of disaggregated demand models. It is not, however, a perfect tool. Additional development of the theory of rational choice behavior, with the goal of providing a more realistic model for travel demand forecasting, appears to be a worthwhile effort.

Data for Travel Demand Forecasting

The effects of data availability on the analytical structure of travel demand forecasting procedures can be described in terms of the data limitations that now exist, the present needs for new data types and new survey procedures, and the problems caused by the use of the available data when present estimation procedures are applied.

The major source of data for travel demand model development continues to be the home interview survey, which has been conducted in every major city of the United States. The data obtained from this survey are deficient for all kinds of demand modeling work for a number of reasons, including these two.

1. The data have been collected by sampling large metropolitan areas with relatively low sampling rates—typically 2 to 10 percent. Any subdivision of the results into a large number of cells (by origin, destination, mode, and purpose, for example) results in a large number of observations of either 0 or 1 trip. These surveyed trips must be factored to represent 0 or 10 to 50 trips, and the factored trips are much too "lumpy" for advantageous use in model development.
2. The tedious process of interviewing, filling out forms, coding, and keypunching can only be done for large surveys by relatively untrained people who must work fast. The net result is that many of the data that result are inaccurate and often are not complete because of the inability of the interviewee to remember all of the details requested.

Additional problems occur when these surveys are used for behavioral disaggregate demand modeling.

1. Home interview surveys only produce data on the trips actually made. Information on the use of alternate modes must be reconstructed from other sources, after the fact, in order to use the data in the development of disaggregated models. Similarly, information on potential trips for households that did not make trips of various kinds may be required, but are not available from the data.
2. Accurate disaggregate modeling at the household level often requires ignoring the

machine-readable data obtained from surveys in favor of returning to coding forms, which include more precise location information (street address versus traffic zone, for example). This greatly increases the costs of disaggregated modeling.

3. The definition of a trip in home interview surveys is an arbitrary one requiring a single mode and purpose. This definition is then modified somewhat by forming new "linked" trips. Often, however, what is desired in behavioral modeling is a "tour" composed of a number of trips that take a traveler from home to one or more destinations and then back home. To obtain such tours often requires a return to coding-form analysis.

Another important source of data for demand modeling work is the U.S. census, which collects a wide range of income, activity-system, and some trip-making data. Because these data must be aggregated to some geographical unit greater than the household to meet confidentiality requirements, they are mainly useful in aggregate rather than disaggregate model development. Expanded data on work trips are available from the 1970 census, and it is possible to consider the development of an aggregate work-trip model based on census data and network data only. Drawbacks remain, however: The degree of aggregation, especially of destinations, often is high, and the data are collected only every 10 years.

The paragraphs above imply a number of needs for new kinds of travel data and for new data collection methods. When disaggregated demand modeling is contemplated, a number of the limitations of existing home interview data can be overcome by designing surveys better suited to these models. Because it is not necessary to have data obtained from entire metropolitan areas to develop these models, surveys can be designed with high sampling rates in relatively small areas. Data recording can be modified to preserve as much locational information as necessary and to represent tours rather than arbitrarily defined trips. Information on alternative modes and destinations can be requested explicitly. Better trained and higher paid interviewers can be used to help improve the reliability of the data. These changes will remove a number of limitations of present travel data, but will only make the obtaining of accurate data more critical. Research aimed toward the improvement of survey data accuracy should be undertaken. Also, methods of integrating survey data with engineering information, such as travel times on highway and transit facilities, should be improved.

With regard to the use of travel data to develop travel demand models, a number of problems can be identified. These problems depend not only on the use of the data but also on the estimation procedures.

As pointed out, there are definite advantages in developing mode-independent demand functions. Such functions require, however, that each alternative mode be described by using the same variables. This raises the problem of developing a set of variables that are meaningful for all modes. The major problem arises when one attempts to describe automobile transportation in terms of variables such as frequency and cost; the variables are relatively straightforward for common-carrier modes. Should automobile cost be out-of-pocket cost only or out-of-pocket cost plus operating cost or both of these plus depreciation, insurance, and other fixed costs? These problems often make the use of mode-independent models impractical.

A second data-estimation problem is multicollinearity among 2 or more variables. As an example, for any mode, both travel time and fare will be strongly related to distance and, therefore, to each other. How can a model be developed that includes both time and cost variables when the estimation procedure cannot accurately determine their parameters because of multicollinearity? Often, this question can only be answered by conducting special experiments or studies to determine the relative effects of 2 or more collinear variables.

A third data-estimation problem is the choice of accurate proxy variables to take the place of ones that theoretically belong in a demand formulation but that are not available. As examples, retail employment may be used as a proxy for shopping opportunities or occupation indicator as a proxy for income. The model developer must analyze the suitability of each proposed proxy variable before accepting it as a potential variable.

In summary, the analyst who must develop demand forecasting procedures by using available data must choose his analytical structure carefully to ensure that he will not be defeated by a lack of the proper data. Also, the analyst who is asked to specify his data needs before a survey strategy is developed should be able to recommend survey procedures and questions that will provide a maximum of data useful for demand modeling.

Demand Estimation Methods

The estimation methods discussed in this section are the distribution model calibration procedures, linear regression, nonlinear regression, and simultaneous equation estimation.

Distribution Model Calibration Procedures

For both the gravity model and the opportunity model (Eqs. 10 and 11), specialized calibration procedures have been developed. In the case of the gravity model, t^{β} is replaced by a generalized distance function f(t), and the values of this function for each value of t are determined such that the actual distribution of trip lengths is matched. In the case of the opportunity model, the parameter L is determined such that the actual average trip length is matched. In both cases, the actual observations, T_{ij}, are not used in the calibration, but instead more aggregate characteristics are matched. Each of these procedures is limited to the particular analytical structure of the corresponding trip distribution model.

Linear Regression

This general parameter-estimation procedure requires that the functional form of the model, or a transform of it, be linear in the parameters. This limits the use of linear regression to functional forms of the following types:

$$Y = a_o + \sum_i a_i x_i \tag{36a}$$

$$Y = a_o + \sum_i a_i \ln x_i \tag{36b}$$

$$Y = a_o + \sum_i (a_i x_i + b_i \ln x_i) \tag{36c}$$

$$\ln Y = a_o + \sum_i a_i x_i \tag{36d}$$

$$\ln Y = a_o + \sum_i a_i \ln x_i \tag{36e}$$

$$\ln Y = a_o + \sum_i (a_i \ln x_i + b_i x_i) \tag{36f}$$

where

Y = either trips, T, or a probability variable $P/(1 - P)$ or P_i/P_j, where P_i is the probability of making a specified trip;
a_i = coefficients to be estimated; and
x_i = independent variables.

The untransformed versions of Eqs. 36d, e, and f are

$$Y = \exp\left(a_o + \prod_i a_i x_i\right) \tag{37a}$$

$$Y = e^{a_o} \prod_i x_i^{a_i} \tag{37b}$$

$$Y = e^{a_o} \prod_i x_i^{a_i} e^{b_i x_i} \tag{37c}$$

Each of the models presented in Eqs. 8, 14, 16, 17, 18, 20, and 22 through 28 can be expressed in one of the forms shown in Eq. 36. However, because of limitations on the independent variables in disaggregated models, linear regression was not used to estimate the equations.

Linear regression is based on the minimization of the sum of the squares of a linear error term. When the dependent variable is transformed, as in Eq. 36, the untransformed error term is no longer linear. In Eq. 37, if U is the transformed error term, then the untransformed error term is e^U, and in each case it has a multiplicative effect on Y. Often this effect is not desirable and, therefore, linear regression is not applicable to the calibration of models such as those of the form of Eq. 36.

A number of modifications of simple linear regression, or ordinary least squares procedures, have been developed. Some of these are

1. Generalized least squares, where observations or error terms or both are weighted to take account of the variation in reliability among observations; and
2. Constrained regression, where some parameters are constrained to equal prespecified values (more flexible constraints are discussed below).

These modifications do not significantly affect the cost of using linear regression and often prove to be useful in travel demand estimation.

Nonlinear Regression

A number of nonlinear regression procedures exist. They overcome the restriction that the model to be calibrated, or a transform of it, be linear in the parameters. However, this requires that the solution method be an iterative programming or direct search procedure, and these procedures are significantly more costly than ordinary least squares. Some of the available features of these procedures are

1. Replacement of the additive (in the linear transform) error term of linear regression with a general error term, depending on the model formulation;
2. Inclusion of constraints on the coefficients, including inequality constraints involving either single coefficients or functions involving both coefficients and independent variables (these constraints can represent theoretical considerations such as the proper signs for the coefficients of price and socioeconomic variables); and
3. Incorporation of procedures to determine maximum likelihood coefficient estimates such as those typically used in multiple logit models (Eq. 26).

Simultaneous Equation Estimation

These methods are essentially methods of determining the best parameters for systems of simultaneous equations usually based on 2-stage least squares procedures. They allow model calibration in the situation where supply and demand functions are shifting simultaneously, as they do over time and across zones. Because few time series data sets or models exist in travel demand forecasting and because demand functions are usually assumed to be fixed in cross-sectional models, little use has been made of simultaneous equation estimation methods.

The most common statistical estimation procedure, ordinary least squares, severely limits the number of functional forms available for travel demand forecasting. Many functional forms cannot be estimated by using this procedure, and, in addition, the number of independent variables is usually limited because of multicollinearity. Only by using more costly procedures, and by developing specialized procedures, can these limitations be overcome.

Structural Characteristics

Three critical structural characteristics of demand forecasting procedures are summations, elasticities, and zonal aggregations. Early demand forecasting procedures stressed the summations of demand by mode, by mode and destination, and by mode, destination, and origin as quantities over which the analyst should have significant control. More recently, the influence of economics has been felt, and the elasticity of trip-making with respect to activity system and level-of-service variables has become more important to the analyst. The effects of aggregation on demand procedures have always been important to the transportation analyst. In this section, each of these terms is formally defined, and their theoretical ranges are stated. The nature of these measures for a number of the analytical structures discussed above is then displayed.

1. The following summations of predicted trips by origin, destination, and mode (T_{ijk}) are of concern to the transportation analyst:

$$T_{ij\cdot} = \sum_k T_{ijk} = \text{trips by zone pair} \tag{38a}$$

$$T_{i\cdot\cdot} = \sum_j \sum_k T_{ijk} = \text{trips by origin} \tag{38b}$$

$$T_{\cdot j\cdot} = \sum_i \sum_k T_{ijk} = \text{trips by destination} \tag{38c}$$

$$T_t = \sum_i \sum_j \sum_k T_{ijk} = \text{total trips} \tag{38d}$$

In the UTP models, these summations are typically predicted in reverse to the order shown above, and an important part of each sequential step is to ensure that the previous predictions, taken as "control totals," are preserved.

2. The formal definition of the elasticity of trip-making from i to j by mode k, with respect to any independent variable, w, is

$$e(T_{ijk}:w) = \frac{\partial T_{ijk}}{\partial w} \cdot \frac{w}{T_{ijk}} \tag{39}$$

Elasticity is a dimensionless number that represents the percentage of change in trip-making from i to j by mode k (T_{ijk}) for each percentage of change in the independent variable w. For a number of independent variables, a more specific name is given. These are indicated below:

$e(T_{ijk}:t_{ijk})$ = direct time elasticity,
$e(T_{ijk}:t_{lmn})$ = time cross elasticity (in this case, only one of subscripts l, m, n need be different from i, j, k), and
$e(T_{ijk}:Y_i)$ = income elasticity.

Similarly, specific names can be given for the elasticities of other level-of-service and activity-system variables.

Economic theory leads to the following statements of the ranges within which the various elasticities can be expected to occur: (a) Direct level-of-service elasticities are less than or equal to 0; (b) level-of-service cross elasticities are greater than or equal to 0; (c) income and similar activity-system elasticities are greater than or equal to 0, unless T_{ijk} represents an inferior good. Equation 39 can also be generalized to apply to the summations shown in Eq. 38, resulting in the elasticity of trips by zone pair, origin, destination, or total trips with respect to any independent variable.

3. A critical question to be answered for each alternative travel demand forecasting procedure is the range of zone sizes for which the procedure is valid. Because of the analytical structure and the magnitude of the coefficients of the socioeconomic and activity-system variables in many models, they are limited to the range of zone sizes for which they were calibrated. If the zone sizes are to be changed greatly, the model will require recalibration.

To explore the conditions that will require recalibration, we must divide both socioeconomic and activity-system variables into 2 classes: (a) scaling variables, such as zonal population and employment, which express the "size" of the zones; and (b) rate variables, such as automobiles per household and dollars of sales per square foot of retail store area. In the remainder of this discussion, we can limit ourselves to the scaling variables, for these are the critical ones in zonal aggregation considerations.

A useful index for any demand model is the sum of the exponents of all scaling variables that are multiplied together. For example, we may have a multiplicative model that predicts T^n_{ijk} by using the following scaling variables and coefficients: (origin population)$^{0.8}$ and (destination employment)$^{0.7}$. In this case, our index is 1.5, which suggests that, for each 1 percent change in zone size, trips will change by 1.5 percent.

As this index begins to vary significantly from 1 for models that predict T^n_{ijk}, we will expect changes in zone size to require recalibration. We will term this aggregation index the AI.

When these summations, elasticities, and aggregation indexes are obtained for the models discussed previously in this paper, the following characteristics of the models are discovered.

Urban Transportation Process

Trip Generation (Eq. 8)

Equation 40c is the major deficiency of the standard trip generation approach: Total trip-making for a zone does not change as level-of-service variables change. The equations, are, however, usually insensitive to zone size.

$$T^n_{i\cdot\cdot} \text{ (obtained directly)} \tag{40a}$$

$$T^n_{\cdot j\cdot} \text{ (obtained directly)} \tag{40b}$$

$$e(T^n_{i\cdot\cdot} : L_{lmn}) = 0 \tag{40c}$$

for all subscript values.

$$e(T^n_{i\cdot\cdot} : S_{i1}) = \frac{c^n_1 S_{i1}}{T^n_{i\cdot\cdot}} \tag{40d}$$

$$AI = 1.0 \tag{40e}$$

Trip Distribution (Eq. 9)

Equation 41d shows that in the gravity model (Eq. 10) a change in level of service from i to any destination affects the number of trips to all destinations. Usually, only the L_{im} for the automobile mode is used. The elasticity for other modes is 0 if this is done. Equation 41e indicates that the level-of-service variables for all other origins are irrelevant. Equation 41f shows that the activity system has no effect on trip distribution beyond its effect on T_i^n and T_j^n, as represented in the trip generation step.

$$T_{ij\cdot}^n \text{ (obtained directly)} \tag{41a}$$

$$T_{i\cdot\cdot}^n \text{ (constrained to equal } T_i^n) \tag{41b}$$

$$T_{\cdot j\cdot}^n \text{ (sometimes constrained to approximate } T_j^n) \tag{41c}$$

$$e(T_{ij\cdot}^n : L_{im}) = \beta_n\left(\delta_{jm} - \frac{T_{im\cdot}^n}{T_{i\cdot\cdot}^n}\right) \tag{41d}$$

where $\delta_{jm} = 1$ if $j = m$ and 0 if $j \neq m$.

$$e(T_{ij\cdot}^n : L_{1mk}) = 0 \tag{41e}$$

when $1 \neq i$.

$$e(T_{ij\cdot}^n : A_1) = 0 \tag{41f}$$

for all values of 1.

$$AI = 1.0 \tag{41g}$$

In addition, Eqs. 41a, b, c, e, and f also hold for the opportunity model (Eq. 11). The differential in Eq. 42a is 0 except when m = j and the ranking of destinations from i changes because of the change in t_{ij} (the differential is positive in this case) and when $m \neq j$ and the ranking of j changes, which will only occur when $|t_{ij} - t_{im}| \leq |dt_{im}|$ (the differential is negative in this case). These conditions imply that the elasticities of trips to all but a few destinations are zero.

$$e(T_{ij\cdot}^n : L_{im}) = -L_{im}L_n \frac{dV_j^n}{dt_{im}} \tag{42a}$$

$$AI = 1.0 \tag{42b}$$

Modal Split-Binary Choice (Eq. 14)

Equations 43b and c indicate the symmetrical nature of the binary-choice model. Equations 43d and e point out that only the travel variables for the various modes connecting i and j have an effect on T_{ijk}.

$$T_{ij\cdot}^n \text{ (constrained to equal } T_{ij}^n) \tag{43a}$$

$$e(T_{ijk}^n : L_{ijk}) = \frac{-a_k L_{ijk}}{1 + e^{-h_k}} \tag{43b}$$

$$e(T_{ijk}^n : L_{ijn}) = \frac{a_k L_{ijn}}{1 + e^{-h_k}} \tag{43c}$$

$$e(T^n_{ijk}:L_{imk}) = 0 \tag{43d}$$

$$e(T^n_{ijk}:L_{imn}) = 0 \tag{43e}$$

$$e(T^n_{ijk}:S_i) = 0 \tag{43f}$$

$$AI = 1.0 \tag{43g}$$

Direct Aggregate Procedures

Product Form (Eq. 16)

In Eq. 44, travel time is used as a typical level-of-service variable. All elasticities and cross elasticities for T_{ijk} are constants and are 0 for level-of-service and activity-system variables not associated with zones i and j. The elasticities of the various summations all have a form similar to Eq. 44e; the simple elasticities are weighted by the appropriate trip share ($T_{ijk}/T_{ij.}$ in the equation shown). Because the simple elasticities are both positive and negative, it is possible that the elasticities of the summations with respect to level-of-service variables will be positive, which is contrary to economic theory. The use of constrained regression to prevent this is infeasible because of the large number of constraint equations required (one for each i-j pair) and cannot ensure that predictions will have the proper summation elasticity, because the shares will change in the future. Zonal aggregation can cause a problem if the coefficients in Eq. 44g sum to a number significantly different from 1.

$$e(T_{ijk}:t_{ijn}) = c_{nk} \tag{44a}$$

$$e(T_{ijk}:t_{imn}) = 0 \tag{44b}$$

$$e(T_{ijk}:S_i) = a_{1k},\ a_{3k} \tag{44c}$$

$$e(T_{ijk}:A_i) = 0 \tag{44d}$$

$$e(T_{ij.}:t_{ijn}) = \frac{\sum_k c_{nk} T_{ijk}}{T_{ij.}} \tag{44e}$$

$$e(T_{ij.}:t_{ijn}) = 0 \tag{44f}$$

$$AI = a^n_{1k} + a^n_{2k} \tag{44g}$$

Linear-Log Form (Eq. 17)

All elasticities and cross elasticities for T^n_{ijk} are linear functions of the respective independent variables, inversely proportional to T^n_{ijk}. Zero elasticities occur whenever the independent variable of concern is not associated with the i-j zone pair. The elasticities of the "scaling" activity-system variables (S_{io}, A_{jo}) are both unity, resulting in an aggregation index of 2. The elasticities of summations all take on a form similar to Eq. 45d. Because a^n_{kki} and b^n_{kki} can be expected to be negative and the remaining parameters can be expected to be positive, but small in magnitude when compared with the direct parameters, these elasticities will normally have the proper sign. It is possible to ensure that this will be the case by using constrained regression.

$$e(T^n_{ijk}:L_{ijml}) = \frac{M^n_k S_{io} A_{jo}}{T^n_{ijk}} (a^n_{kml} L_{ijml} + b^n_{kml}) \tag{45a}$$

$$e(T^n_{ijk}:S_{io}) = 1 \tag{45b}$$

$$e(T^n_{ijk}:S_l) = \frac{M^n_k S_{io} A_{jo}}{T^n_{ijk}} (c^n_{kl} S_l + d^n_{kn}) \tag{45c}$$

$$e(T^n_{ij\cdot}:L_{ijml}) = \frac{S_{io} A_{jo}}{T^n_{ij\cdot}} \sum_k M^n_k (a^n_{kml} L_{ijml} + b^n_{kml}) \tag{45d}$$

$$AI = 2.0 \tag{45e}$$

Product-Exponential Form (Eq. 18)

All elasticities and cross elasticities for T^n_{ijk} are linear functions of the respective independent variables, independent of the level of T^n_{ijk}. Zero elasticities occur whenever the independent variable of concern is not associated with the i-j zone pair.

$$e(T^n_{ijk}:L_{ijml}) = a^n_{kml} + b^n_{kml} L_{ijml} \tag{46a}$$

$$e(T^n_{ijk}:S_l) = c^n_{kl} + d^n_{kl} S_l \tag{46b}$$

$$e(T^n_{ij\cdot}:L_{ijml}) = \sum_k \frac{T^n_{ijk}}{T_{ij\cdot}} (a^n_{kml} + b^n_{kml} L_{ijml}) \tag{46c}$$

$$AI = 1.0 \tag{46d}$$

Independent Abstract Mode Procedures (Eq. 20)

Equation 47 indicates significant discontinuities for the elasticities of "best" modes and other modes. The 0 cross-elasticity of Eq. 47c when $m \neq b$ is especially troublesome. Equation 47 indicates that the elasticities and cross elasticities of this model are independent of the mode of trips, k, as would be expected in an abstract mode model.

$$e(T^n_{ijk}:P_l) = a_1 \tag{47a}$$

$$e(T^n_{ijk}:t_{ijk}) = \begin{cases} a_9 \text{ when } k \neq b \\ a_6 \text{ when } k = b \end{cases} \tag{47b}$$

$$e(T^n_{ijk}:t_{ijm}) = \begin{cases} 0 \text{ when } m \neq b \\ a_6 - a_9 \text{ when } m = b \\ a_6 \text{ when } k = b \end{cases} \tag{47c}$$

$$e(T^n_{ij\cdot}:t_{ijl}) = \begin{cases} a_9 \dfrac{T_{ijl}}{T_{ij\cdot}} \text{ when } l \neq b \\ a_6 - a_9\left(1 - \dfrac{T_{ijl}}{T_{ij\cdot}}\right) \text{ when } l = b \end{cases} \tag{47d}$$

$$AI = a_1 + a_2 \tag{47e}$$

Modal Share Models (Eqs. 21, 22, and 23)

Equation 48 indicates that the elasticities and cross elasticities with respect to travel time by a given mode are directly related to the share of trips using that mode. The parameter b_5 should be in the range of 0 to 1, with a value near 0 expected. If it is 0, the elasticity of total trips by zone pair (Eq. 48e) will be 0. If it is 1, the direct time elasticity (Eq. 48b) will be simply a_{2k}, and the cross elasticities (Eq. 48c) will be 0, as in the product form of Eq. 16.

$$e(T^n_{ijk}:P_i) = b_1 \tag{48a}$$

$$e(T^n_{ijk}:t_{ijk}) = a_{2k}\left[\frac{T_{ijk}}{T_{ij\cdot}}(b_5 - 1) + 1\right] \tag{48b}$$

$$e(T^n_{ijk}:t_{ijm}) = a_{2m}\frac{T_{ijm}}{T_{ij\cdot}}(b_5 - 1) \tag{48c}$$

$$T^n_{ij\cdot} = f_2 \tag{48d}$$

$$e(T^n_{ij\cdot}:t_{ijk}) = a_{2k}b_5\frac{T_{ijk}}{T_{ij\cdot}} \tag{48e}$$

$$AI = b_1 + b_2 \tag{48f}$$

Disaggregate Separable Decision Models

Modal Choice (Eq. 24)

In a similar fashion to Equations 43a, b, c, d, and e, Eq. 49 indicates that the elasticities of travel by a given mode with respect to the independent variables are directly proportional to the value of the independent variables, the value of their coefficient, and the fraction of traffic not using the given mode. The elasticities of travel with respect to variables not associated with origin i or destination j are all 0. Also, as expressed in the independence-of-irrelevant-alternatives axiom, the elasticity of travel by any mode with respect to level-of-service variable of any second mode does not depend on the characteristics of any mode except the second. Let

$$h^n_{ij} = a^n + \sum_1 b^n_1(L_{ija1} - L_{ijt1}) + \sum_1 c^n_1 S_{i1} \tag{49a}$$

Then Eq. 24 becomes

$$\frac{P^n_{ija}}{1 - P^n_{ija}} = \exp(h^n_{ij}) \tag{49b}$$

$$e(T^n_{ija}:S_{i1}) = c^n_1 S_{i1}(1 - P^n_{ija}) \tag{49c}$$

$$e(T^n_{ija}:L_{ija1}) = b^n_1 L_{ija1}(1 - P^n_{ija}) \tag{49d}$$

$$e(T^n_{ija}:L_{ijt1}) = -b^n_1 L_{ijt1}(1 - P^n_{ija}) \tag{49e}$$

$$e(T^n_{ija}:L_{mnk1}) = 0 \tag{49f}$$

$$T^n_{ij\cdot} \text{ (constrained to equal } T^n_{ij}) \tag{49g}$$

$$AI = 1.0 \tag{49h}$$

Other Decisions (Eqs. 25, 26, and 27)

The remaining decisions—time of day, destination, and trip frequency—all have basically the same structure as the modal-choice structure of Eq. 24. Their elasticities and summations, therefore, also have the same characteristics.

Direct Decision Model (Eq. 28)

Equation 28 also has the same structure as Eq. 24 and, therefore, its elasticities have the same characteristics. However, because it is not a sequential model, the elasticities of trip summations are expressed differently.

$$e(T_{ij.}:L_{ijkl}) = d_l L_{ijkl}\left(\frac{T_{ijk}}{T_{ij.}} - \frac{T_{ijk}}{T_{i..}}\right) \tag{50}$$

This equation indicates that the elasticity of trips by all modes from i to j with respect to an independent variable is directly proportional to that variable, its coefficient, and the difference between trips by the mode of that variable as a fraction of total trips between i and j and the same trips as a fraction of total trips from i.

This concludes a summary of the structural characteristics for the set of currently used demand forecasting procedures described in an earlier section. It is obvious that these procedures have a wide range of characteristics and that in some cases the analytical structure itself does not ensure that all characteristics will agree with economic and travel behavioral theory. When these procedures are used, the analyst must investigate carefully the resulting characteristics, to be sure that all aspects of his model are realistic.

After determining the characteristics of a number of forecasting procedures, we can list a number of desirable characteristics. Research can then be done to search for analytical structures that satisfy those desires. This approach to the development of improved analytical structures for travel forecasting has, to some extent, influenced past developments in the field (14, 16, 18, 24). Some of the kinds of desirable characteristics are as follows:

1. The mathematical form of critical elasticities and cross elasticities should be as specified,
2. The effects of the aggregation of traffic zones on model predictions should be as specified,
3. The variation in competition between pairs of modes should be reflected in the model, and
4. The effects of adding new modes on summations of trips should be as specified.

Integration Into Analysis Systems

A number of desirable characteristics of transportation analysis systems place critical constraints on demand forecasting procedures and create requirements for a number of specialized kinds of procedures. Four examples of these characteristics are discussed in this section.

Consistent Estimation of Network Equilibrium

Manheim (12) has discussed the need for transportation analysis systems that use a consistent set of level-of-service variables, consistent both with the demand procedure and with the supply procedure. He points out that this requirement is violated in the UTP procedures when final values of level-of-service variables are not used during the trip distribution and modal-split phases. As a result, demand is erroneously estimated, and the final level-of-service variables are incorrect.

To modify present transportation analysis systems so that they can be consistent, less cumbersome demand procedures than those now used are desirable. Because of their structures, direct demand models have been seen as logical candidates to meet this requirement. In large measure, this accounts for their use in DODOTRANS, one of the first transportation analysis systems that explicitly attempts to estimate network equilibrium in a consistent manner (13). As discussed in the previous section, however, present direct demand models have structural characteristics that are not satisfactory. Therefore, improved models are needed—ones that have the ease of application of the direct demand models and are as controllable as the present UTP procedures. Manheim has proposed a family of analytical formulations to meet these objectives. These models, the general share models, can be expressed either as a sequential set of models or as a direct model.

Pivot-Point Procedures

Often, the analyst is faced with the following situation: The details of the existing travel pattern in an analysis area are known (all interzonal trips and level-of-service variables by mode), and the effects on the transportation system of relatively small changes on this travel pattern are desired. Usually the analyst has a number of choices. The first is to manually estimate the effects. The second is to perform a complete analysis from trip distribution through traffic assignment. The remaining choices fall somewhere in between, involving only partial use of the UTP, based on assumptions that trip distribution or that modal split will not change. Regardless of the choice made, very little of the existing information will be used and, therefore, the resulting estimates may differ from the existing situation more because of calibration errors than of the proposed changes.

Pivot-point procedures have been designed to improve the analyst's forecasts when he is faced with the situation just described. They allow changes in travel to be estimated, based on changes in the transportation system. These procedures minimize the calibration problem by using the existing data and by specifying the elasticities of travel-making with respect to the available level-of-service data. The equation used for estimating changes, based on the total differential of a function, is the following:

$$\Delta T^n_{ijk} = T^n_{oijk}\left[\sum_{1} e(T^n_{ijk}:S_l)\frac{\Delta S_l}{S_{ol}} + \sum_{1} e(T^n_{ijk}:A_l)\frac{\Delta A_l}{A_{ol}} + \sum_{m} e(T^n_{ijk}:L_m)\frac{\Delta L_m}{L_{om}}\right] \tag{51}$$

where

ΔT^n_{ijk} = change in trips from i to j by mode k for purpose n,
o = old or former value,
S_l = socioeconomic variable,
A_l = activity-system variable, and
L_m = level-of-service variable.

Regardless of what the demand model structure is, the elasticities can be assumed to be constant for small changes. Equation 51, therefore, becomes generally applicable for predicting the effects of small changes. For larger changes, explicit functional forms of the elasticities (arc elasticities) can be used.

The most significant impact of pivot-point procedures is on the design of analysis systems. They also, however, have an effect on demand modeling. They imply that much effort should be put into obtaining good estimates of elasticities, for these alone are needed to use Eq. 51. Because elasticities can best be estimated when a change is observed, this implies that many careful before-and-after studies of transportation should be carried out.

Dynamic Transportation Analysis

As stated in the introduction, we assumed that transportation demand can be divided into short-run and long-run phenomena, and we will concentrate on modeling the short-run situation. Actually, however, there is a continuous variation in effects over time from short run to long run. To reflect this continuity in our models, we must construct dynamic systems by using variables that have a range of lag times, as discussed by Ben-Akiva (2). Such a system would incorporate both land use models and travel prediction models into a set of demand models that would provide predictions of both the long- and the short-range effects of transportation.

Although such an approach is useful as a method to incorporate the time dimension into travel forecasting, it will generate new problems in the areas of empirical estimation, data collection, and convergence of the solution. Work should begin on a study of these problems so that in the future dynamic transportation modeling can be started.

Aggregation of Disaggregate Procedures

To incorporate disaggregate travel demand forecasting procedures into analysis systems, methods of interfacing these procedures with aggregate zonal data must be developed. If the models are applied directly to zonal averages of socioeconomic, activity-system, and level-of-service variables, the major advantage of disaggregated procedures will be lost. Some way must, therefore, be found to incorporate the distributions of zonal variables into the application of the procedures.

One approach that has been suggested is the sampling from these distributions by using Monte Carlo simulation techniques to obtain observations of the independent variables required to predict trips. For some models, it may be possible to analytically obtain the expected value of trips, based on incorporating all of the relevant distributions of variables. This is an area in which research should begin, both to look for alternate approaches and to test the various proposed methods to determine their usefulness and accuracy.

RECOMMENDED PROGRAM OF RESEARCH

In this part of the paper, all of the suggestions for further research included in the previous part will be brought together as a unified program of research in the area of the analytical structure of travel forecasting procedures. Each recommended area of research will be given a priority rating and a recommended time frame for carrying out the research.

Travel Demand Theories

1. Lancaster's approach to consumer utility and demand should be expanded to be applied directly to travel demand. The implications of this approach to estimating the demand for transportation as a part of activities that have utility to the consumer should be explored with a view toward developing additional theoretical guidelines to the travel demand model developer. The priority is medium, and the time frame is 3 to 8 years.
2. Work should be continued on the development of a general equilibrium model that concentrates on transportation demand prediction. The work done to date (3) should be continued in the following areas: (a) searching for a utility function form that results in demand functions with a closed form, (b) exploring the existing formulation for its implications on suitable analytic structures, and (c) searching for realistic revisions of the formulation that will result in useful demand functions. The priority is medium, and the time frame is 3 to 8 years.
3. Work should be begun on the incorporation of the total travel time constraint into economic theories of the consumer because of the importance of travel time as a deter-

minant of travel demand. The priority is medium, and the time frame is 1 to 5 years.

4. The theory of rational choice behavior, as developed in psychology, should be developed further, with a view to its application to travel behavior in particular. The goal should be to develop a framework that can be used to construct more realistic models for travel demand forecasting. The priority is high, and the time frame is 1 to 5 years.

5. Work should continue on the testing of alternative assumed sequences of traveler choice. Because these sequences are so crucial to both aggregate and disaggregate sequential models, the effects of alternative assumptions on model accuracy should be determined for a number of classifications of trips, including urban work and shopping trips and intercity business and pleasure trips. The priority is medium, and the time frame is 1 to 5 years.

Data for Travel Demand Forecasting

1. Work should begin on developing travel survey methods that will provide the data needed for disaggregated demand modeling in the most accurate and efficient manner possible. This work should proceed from the development of alternative designs through the conducting of prototypical surveys, the use of the data obtained in model estimation, and the evaluation of the methods for future use. The priority is high, and the time frame is 1 to 3 years.

2. Research into methods of improving the accuracy of survey data should be carried out, including alternative methods of monitoring and recording travel data and of integrating survey data with engineering information. This is an area where the usefulness of new technology, such as automatic vehicle (and perhaps people) locator systems, should be explored. The priority is medium, and the time frame is 3 to 8 years.

3. Specialized surveys and studies should be designed and conducted to help provide answers to questions not answered by present demand procedures because problems of multicollinearity prevented all relevant variables from being included. For example, careful before-and-after studies and controlled experiments should be conducted to learn more about the responses of travelers to fare, time, and frequency changes. The priority is high, and the time frame is 1 to 5 years.

Demand Estimation Methods

Research should be carried out by statisticians to develop accurate and unbiased estimation procedures for use in travel demand model development. The concentration should be placed on analytic structures that have theoretical appeal but have not been used to date because it has not been possible to estimate their parameters. The priority is medium, and the time frame is 3 to 8 years.

Structural Characteristics

1. The various analytical structures that have been developed or proposed should be studied carefully to determine their characteristics: elasticities, cross elasticities, aggregability, summations, and ability to balance trip origins and destinations by zone. Characteristics that can be, or are always, contrary to theory should be pointed out, and changes to the structures should be proposed to prevent such characteristics from occurring. The priority is high, and the time frame is 1 to 5 years.

2. As proposed analytical structures are found that have promising characteristics, work should be done to calibrate them to determine their applicability to actual travel phenomena. Alternative structures should be compared by using criteria based on goodness-of-fit measures, ease of calibration, and constancy of parameters. The priority is high, and the time frame is 1 to 10 years.

3. Research should be conducted to proceed from alternative specifications of the

structural requirements of demand models to the determination of analytical structures that satisfy these requirements. The alternative sets of specifications should be generated with particular demand estimation problems in mind, such as predicting the demand for a new mode by a particular market segment or predicting the effects of relatively minor changes in operating policies. The priority is medium, and the time frame is 3 to 8 years.

Integration Into Analysis Systems

1. Research should be carried out to determine methods by which the existing analysis systems can be modified to provide for the consistent estimation of travel demand, both by modifying the structure of those systems minimally and keeping the present demand procedures and by incorporating new procedures better suited to the consistent estimation of network equilibrium. The priority is high, and the time frame is 1 to 3 years.
2. Research should be carried out to develop new analysis systems that will incorporate a wide range of demand procedures in an efficient system that consistently estimates network equilibrium. The limitations placed on demand procedures by these systems should be determined and removed if necessary to provide for the realistic estimation of travel demand. The priority is medium, and the time frame is 3 to 5 years.
3. Research should be carried out to develop demand models that will be efficient for use in consistent network equilibrium prediction systems. The general share models should be examined in this light, and recommendations should be made on their further development or on alternative directions of improvement. The priority is medium, and the time frame is 3 to 5 years.
4. Research should be conducted to develop pivot-point procedures as integral parts of transportation analysis systems and to develop the demand models and data needed to make these procedures useful for a wide range of small-scale transportation prediction problems. The priority is high, and the time frame is 1 to 3 years.
5. The feasibility of developing a dynamic system of models to incorporate short-term demand estimation and long-term land use predictions should be studied. Such a study should address the data requirements that this approach will generate, the estimation problems, and the convergence problems. The result should be a program of work to provide the necessary data and tools to allow the calibration of such a model in the future. The priority is medium, and the time frame is 3 to 8 years.
6. Methods to interface disaggregate demand models with aggregate zonal data in analysis systems should be developed and tested. Also, the possibility of eliminating the zonal aggregation of the data needed for demand models should be explored, taking advantage of the data directly available from home interview surveys and from the census. These research tasks should be addressed both to the use of disaggregated models with existing and with predicted future socioeconomic, activity-system, and level-of-service data. The priority is high, and the time frame is 1 to 3 years.

ACKNOWLEDGMENTS

I am indebted to my colleagues, Marvin Manheim, Moshe Ben-Akiva, and Wayne Pecknold, who in a number of discussions generated many of the concepts presented here. Daniel Brand should be acknowledged for his preparation of the guidelines for the paper and for his helpful review of a previous draft of the paper. I have drawn from the work of many others, and I hope they are all acknowledged in the references. Finally, I take full responsibility for all errors made in pulling together their work in this paper.

REFERENCES

1. Ben-Akiva, M. E. PhD diss., Dept. of Civil Eng., M.I.T., Cambridge, 1972.
2. Ben-Akiva, M. E. Passenger Transportation Demand: Theory and Models. Dept. of Civil Eng., M.I.T., Cambridge, 1972.
3. A Disaggregated Behavioral Model of Urban Travel Demand. Charles River Associates, Inc., 1972.
4. Deen, T., Mertz, W., and Irwin, N. Application of a Modal Split Model to Travel Estimates for the Washington Area. Highway Research Record 38, 1963, pp. 97-123.
5. 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.
6. Fleet, C. R., and Robertson, S. R. Trip Generation in the Transportation Planning Process. Highway Research Record 240, 1968, pp. 11-31.
7. Henderson, J. M., and Quandt, R. E. Microeconomic Theory. McGraw-Hill, New York, 1958.
8. Lancaster, K. J. A New Approach to Consumer Theory. Jour. of Polit. Econ., Vol. 74, April 1966.
9. Lave, C. A. A Behavioral Approach to Modal Split Forecasting. Transp. Res., Vol. 3, No. 4, 1969.
10. Lisco, T. E. The Value of Commuters' Travel Time: A Study in Urban Transportation. PhD diss., Univ. of Chicago, 1967.
11. Luce, R. D. Individual Choice Behavior. John Wiley, New York, 1959.
12. Manheim, M. L. Practical Implications of Some Fundamental Properties of Travel Demand Models. Dept. of Civil Eng., M.I.T., Cambridge, Res. Rept. R72-63, 1972.
13. Manheim, M. L., and Ruiter, E. R. DODOTRANS I: A Decision-Oriented Computer Language for Analysis of Multimode Transportation Systems. Highway Research Record 314, 1970, pp. 135-163.
14. Mayberry, J. K. Structural Requirements for Abstract Mode Models of Passenger Transportation. In The Demand for Travel: Theory and Measurement (Quandt, R. E., ed.), Heath-Lexington, 1970, Ch. 5.
15. McLynn, J. M., and Woronka, T. Passenger Demand and Modal Split Models. Northeast Corridor Transp. Proj., U.S. Dept. of Transp., Rept. 230, 1969.
16. Parks, R. W. Systems of Demand Equations: An Empirical Comparison of All Functional Forms. Econometrica, 1969.
17. Pratt, R. H., and Deen, T. B. Estimation of Sub-Modal Split Within the Transit Mode. Highway Research Record 205, 1967, pp. 20-30.
18. Quandt, R. E., and Baumol, W. J. The Demand for Abstract Transport Modes: Theory and Measurement. Jour. of Reg. Sci., Vol. 6, No. 2, 1968.
19. 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.
20. Shunk, G. A., and Bouchard, R. J. An Application of Marginal Utility to Travel Model Choice. Highway Research Record 322, 1970, pp. 30-39.
21. Demand for Intercity Passenger Travel in the Washington-Boston Corridor. Systems Analysis and Research Corp., 1963.
22. Talvitie, A. P. Econometric Model for Downtown Work Trips. PhD diss., Northwestern Univ., Evanston, Ill., 1971.
23. Warner, S. L. Stochastic Choice of Mode in Urban Travel: A Study in Binary Choice. Northwestern Univ. Press, Evanston, Ill., 1962.
24. Wilson, A. G. The Use of Entropy Maximizing Models. Jour. of Transp. Econ. and Policy, Vol. 3, No. 1, 1969.