PRACTICAL IMPLICATIONS OF SOME FUNDAMENTAL PROPERTIES OF TRAVEL-DEMAND MODELS

Marvin L. Manheim, Department of Civil Engineering,
Massachusetts Institute of Technology

The 2 major approaches to finding equilibrium between supply and demand in a network are the indirect, which uses a sequential demand model, and the direct, which uses an explicit demand model. This paper describes the general share model that can be used in explicit form in a direct approach to computing equilibrium or in the form of a sequence of equations in an indirect approach. A number of practical implications follow from these theoretical results. A new model system should be developed without the serious limitations and internal inconsistencies of the urban transportation model system, which uses the indirect, sequential approach. Such a new system can be very general and designed to compute equilibrium with the GSM in a direct approach. Options can be provided with a rich variety of specific demand models, as special cases of the general share model, but within the same general structure. Efficient procedures for computing equilibrium may be developed by exploiting the elasticity properties of the general share model. As an immediate and practical step, present computer programs should be modified to compute a valid equilibrium, to include level of service explicitly and consistently at each step, especially trip generation, and to incorporate the special product models.

• THE OBJECTIVE of this paper is to present some recent results in the theory of transportation systems analysis. These results deal with properties of demand models. Although the results are largely theoretical, they have immediate practical application. The present paper stresses these practical implications and only summarizes the key aspects of the theoretical results; the theoretical material is presented more extensively elsewhere (1).

Several factors motivated the research that is summarized here:

1. The development of the theory of transportation systems analysis;
2. The need to have travel-demand models appropriate to the transportation issues with which we are now concerned, especially in urban transportation planning;
3. The significant weaknesses of the conventional approach to travel-demand forecasting used in urban transportation studies; and
4. The emergence of alternative approaches—direct demand models of an aggregate nature and behavioral models of a disaggregate nature.

THEORY

The theory of transportation systems analysis has emerged from several sources (2, 3, 4, 5). In outline, the problem of predicting the flows in a transportation system is a simple application of economic theory: The flows that will result from a particular transportation system T and the pattern of socioeconomic activities A can be determined by finding the resulting equilibrium in the transportation market. If V = volume of flow,

Sponsored by Committee on Transportation Systems Design.

L = level of service experienced by that volume, and F = (V,L) = flow pattern, then we find equilibrium by establishing a supply function S and a demand function D and by solving for the equilibrium flows F_o consistent with both relations (1):

$$\begin{bmatrix} L = S(V,T) \\ V = D(L,A) \end{bmatrix} \longrightarrow \begin{bmatrix} F_o = (V_o,L_o) \end{bmatrix} \tag{1}$$

This is shown in Figure 1 with travel time t as the level of service L.

Although simple in outline, the application of the theory becomes complex in practice for several reasons:

1. The consumer considers many service attributes of the transport system when making a choice (e.g., line-haul travel time, transfer time, walk distance, out-of-pocket cost, and privacy), and thus L must be a vector with many components;
2. Determining the demand functions (as well as other elements) to use is difficult; and
3. The equilibrium occurs in a network, where flows from many origins to many different destinations interact and compete for the capacity of the network, and the form of these interactions is affected by the topology of the network.

Thus, fairly elaborate computational schemes are required to actually determine the equilibrium flows F_o for a particular (T,A).

In the case of a multimodal network, the symbol V represents an array of volumes

$$V = \{V_{klmp}\} \tag{2}$$

for every k, l, m, and p, where V_{klmp} is the volume flowing from origin zone k to destination zone l via mode m and path p of that mode and where the braces indicate a set of elements V_{klmp}. Ideally, once we have established our demand and supply functions, we would then like to be able to turn directly to an equilibrium-calculating procedure to solve the 2 sets of relations to find the equilibrium flow pattern. The result of this computation would be the 2 arrays comprising that flow pattern:

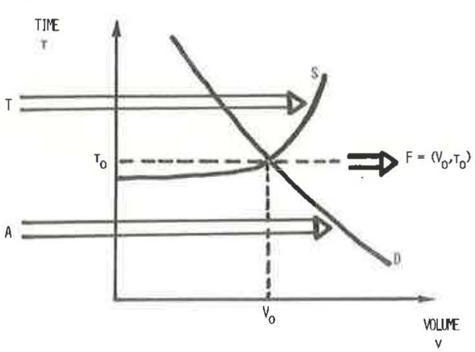
$$F_o = (V_o,L_o) \tag{3}$$

where

$V_o = \{V_{klmp}\}$ for every k, l, m, and p; and
$L_o = \{L_{klmp}\}$ for every k, l, m, and p.

In words, we should get out of our equilibrium procedure the volumes, and the levels of service experienced by those volumes, from k to l by mode m and path p.

Figure 1. Basic theory.



Unfortunately, at this state of the science of transportation modeling, although several systems of transportation models exist, there is not even one operational model that solves for these equilibrium flows exactly and directly. Each of the available systems of models represents a different operational approach to computing equilibrium in transport networks; the differences in these approaches are reflected both in the computational algorithms and in the structure of the demand models that are used. None of these produces an exact equilibrium.

Alternative Approaches

We deal here with only one of Wardrop's "principles"; we will not discuss the other, concerned with global optimization of the flow pattern, as this is inapplicable to urban transportation flow prediction.

One particular computational scheme is that used in urban transportation planning studies. In this indirect approach, the equilibrium flows are estimated in a sequence of steps, commonly called trip generation, trip distribution, modal split, and traffic assignment (9, 10).

Correspondingly, the demand function D is represented as a sequence of functions: trip generation (and attraction) equations, trip distribution procedures (including the friction factor transformation of L), modal-split equations, and minimum-path rules of the traffic assignment procedures. We will refer to this approach as the urban transportation model system (UTMS).

More recently, alternative approaches have been developed. One approach uses explicit demand models to estimate the equilibrium flows in a direct approach, i.e., in a single step instead of in a sequence of steps as in the UTMS approach. Thus, such explicit demand models combine the functions of generation, distribution, and modal split (and, potentially, route choice) into a single process. The first such models were developed for forecasting intercity passenger travel for the Northeast Corridor Project of the U. S. Department of Transportation. They began with the Kraft-SARC model and were followed by the work of McLynn and others (3, 11, 12, 13, 14). Later work extended these models to urban travel (15, 16). These types of explicit demand models were first used for transportation network analysis in the simulation studies for the Northeast Corridor Project. In the DODOTRANS system of computer models, a number of these models are available for use in computing equilibrium flows in networks (17).

Other major directions of work include the disaggregate approach of the so-called behavioral models (18, 19), the development of aggregate models from entropy considerations (20), and the Harvard model system for national transportation planning (21).

Careful appraisal of all of these approaches indicates that, viewed simply as a computational problem, the task of computing the equilibrium of supply and demand in a network remains a difficult one. The UTMS approach has significant limitations, as we shall see shortly; the Northeast Corridor and Harvard models make special assumptions that essentially make their approaches not generally applicable. The DODOTRANS system is the most theoretically acceptable of the currently operational and practical approaches; however, the convergence and uniqueness properties of its computational procedures are not wholly satisfactory either. Several promising approaches are under development (22, 23, 24).

Thus, although the theory of what should be done is clear, at present there are a number of alternative approaches that are or can be taken to predict flows in networks as the equilibrium of supply and demand. Each approach involves specific assumptions, both explicit and implicit, in the choice of demand models and of the computational procedures for determining equilibrium. Many alternative assumptions and computational procedures are possible. Very serious biases may occur in the computed flow patterns, as compared with the "true" equilibrium, if the assumptions and computational approaches are not carefully considered. Therefore, it is essential that the transportation analyst have a sound understanding of the role of demand models in the equilibrium calculation so that he can appraise possible biases.

APPRAISAL OF THE URBAN TRANSPORTATION MODEL SYSTEM

The UTMS is the most widely used transportation systems analysis approach. It has been applied in more than 200 cities in the United States and in many other cities around the world. The development and the institutionalization of this approach during the past 15 years are major accomplishments; it is the first large-scale application of modern systems analysis techniques to problems of the civil sector.

In the UTMS, the travel-demand models—and the equilibrium computations—are structured into a sequence of 4 major steps: trip generation, trip distribution, modal split, and traffic assignment. Essentially, this amounts to estimating V_{klmp} in a series of successive approximations: first V_k, then V_{kl}, then V_{klm}, and finally V_{klmp}.

It is useful to examine the UTMS critically from the perspective of equilibrium theory and the challenge of today's urban transportation problems (1). It seems obvious that the following conditions should be met by any set of demand models and equilibrium-calculating procedures.

1. Level of service L should enter into every step, including trip generation (unless an analysis of the data indicates in a specific situation that trip generation is, in fact, independent of level of service for all market segments over the full range of levels of service to be studied).
2. The level of service attributes used should be as complete as necessary to adequately predict traveler behavior. For example, time reliability, number of transfers, and privacy should be included if empirical evidence indicates these are important.
3. The same attributes of service level should influence each step (unless the data indicate otherwise). For example, transit fares, automobile parking charges, walking distances, and service frequencies should influence not only modal split but also assignment, generation, and distribution.
4. The process should calculate a valid equilibrium of supply and demand; the same values of each of the level-of-service variables should influence each step. For example, the travel times that are inputs for modal split, distribution, and even generation should be the same as those that are outputs of the assignment. If necessary, iteration from assignment back to generation, distribution, and so on should be done to get this equilibrium.
5. The levels of service of every mode should influence demand. Congestion, limited capacity (e.g., parking lots), and fares of each mode should (in general) affect not only its own demand but also the demand for other modes at all steps (generation, distribution, modal split, and assignment). That is, there should be provision for explicit cross elasticities.
6. The several demand functions for each step should be internally consistent (in the sense defined in the next section).
7. The estimation procedures should be statistically valid and reproducible.

Careful examination of the UTMS indicates that it violates each of these conditions (1). As a consequence, serious questions can be raised about the biases and limitations of the flow predictions resulting from use of the UTMS. Although the UTMS does have important advantages, these do not outweigh its very serious liabilities.

What is desirable is an improved approach that overcomes these limitations by meeting the conditions listed. The results described here do this and, at the same time, preserve the advantages of the indirect approach to equilibrium used in the UTMS.

BASIC DEFINITIONS

Notation

The following notation is used in this paper:

V_{klmp} = volume of trips going from zone k to zone l by mode m and path p;
V_{klm} = volume of trips going from zone k to zone l by mode m;
V_{kl} = volume of trips going from zone k to zone l;
V_k = volume of trips originating in zone k;
V_T = volume of trips (interzonal) in the region;
$\underline{A}$ = vector of variables describing the socioeconomic activity system;
$\underline{a}$ = vector of parameters applying to $\underline{A}$;
$\underline{X}_{klmp}$ = vector of S level-of-service variables (i = 1, 2, . . . , S) describing the transportation system characteristics as experienced by trips going from zone k to zone l by mode m and path p;
$\underline{X}$ = $\{\underline{X}_{klmp}\}$ = set of all level-of-service characteristics for all paths p of all modes m between all origins k and all destinations l;
$\underline{w}$ = vector of parameters applying to $\underline{X}$;
R_{klmp} = $f(\underline{X},\underline{w})$ = combined effect of all level-of-service characteristics of all modes as they influence trips going from zone k to zone l by mode m and path p;

$R_{klmq,p}$ = combined effect of all level-of-service characteristics of mode q as they influence trips going from zone k to zone l by mode m and path p;
Z = $f(\underline{A},\underline{a})$ = combined effect of all activity-system characteristics;
Y = $f(Z,R)$ = combined effect of all activity-system and level-of-service characteristics; and
ψ = demand function.

V_{klmp}, V_{klm}, V_{kl}, and V_T are in general not the same as the various partial sums obtained by aggregating over all values of one or more subscripts. For example, we will write $V_{klm.}$ and $V_{kl..}$ when we do mean

$$\left.\begin{aligned} V_{klm.} &= \sum_p V_{klmp} \\ V_{kl..} &= \sum_m V_{klm.} \end{aligned}\right\} \quad (4)$$

and so on. In general,

$$\left.\begin{aligned} V_{klm.} &\neq V_{klm} \\ V_{kl..} &\neq V_{kl} \end{aligned}\right\} \quad (5)$$

except where otherwise indicated.

The definition of $R_{klmq,p}$ represents a cross-elasticity type of effect where the service attributes of mode q affect the demand for path p of mode m.

Those definitions are illustrated as follows (we assume each mode m has only 1 path, and therefore we drop the subscript p):

$$V_{klm} = a_1 P_k^{a_2} E_l^{a_3} t_{klm}^{a_4} c_{klm}^{a_5} \quad (6)$$

In this simple product form of demand model, we have

$\underline{A}$ = (P_k, E_l) = population at origin k, employment at destination l;
$\underline{X}$ = (t_{klm}, c_{klm}) = time, cost;
$\underline{a}$ = (a_1, a_2, a_3);
$\underline{w}$ = (a_4, a_5);
Z_{kl} = $a_1 P_k^{a_2} E_l^{a_3}$;
R_{klm} = $t_{klm}^{a_4} c_{klm}^{a_5}$;
Y_{klm} = $Z_{kl} \cdot R_{klm}$; and
V_{klm} = $\psi_{klm}(Y) = Z_{kl} \cdot R_{klm}$.

Types of Models

The general demand model system is

$$V = \psi(V,Y) \quad (7)$$

This is implicit in that V appears on the right side as well as on the left of the equation. Thus, even if we know all elements of Y, including the level of service $\underline{X}$, we would still generally have difficulty computing the value(s) of V to satisfy the equation (for example, the iteration of the modified gravity model to balance productions and attractions in the UTMS).

The implicit system may be several equations:

$$\left.\begin{aligned} V_1 &= \psi_1(V,Y) \\ V_2 &= \psi_2(V,Y) \\ &\cdots \\ V_N &= \psi_N(V,Y) \end{aligned}\right\} \quad (8)$$

or it may be a single equation:

$$V_{klmp} = \psi(V_{klmp}, Y) \tag{9}$$

One special case of the general implicit system is the explicit system of demand models:

$$V_{klmp} = \psi(Y) \tag{10}$$

in which V appears only on the left side of the equation. The McLynn, Kraft-SARC, and similar models are explicit models.

Another special case of the general implicit system is the sequential implicit:

$$\left.\begin{aligned} V_T &= \sigma_1(Y) \\ V_k &= \sigma_2(V_T;Y) \\ V_{kl} &= \sigma_3(V_k;Y) \\ V_{klm} &= \sigma_4(V_{kl};Y) \\ V_{klmp} &= \sigma_5(V_{klm};Y) \end{aligned}\right\} \tag{11}$$

In principle, the sequential form can be transformed into an explicit form:

$$V_{klmp} = f(Y) \tag{12}$$

where $f = f(\sigma_1, \sigma_2, \sigma_3, \sigma_4, \sigma_5)$.

The explicit form can be used in the direct approach to computing equilibrium. The sequential form requires an indirect approach in a series of steps; when transformed into an explicit form as in Eq. 12, the direct approach can be used.

The forecasting method used in urban transportation studies, the UTMS, is based on a sequential implicit system of demand models used in an indirect approach to equilibrium: σ_1 and σ_2 correspond to the trip generation equations, σ_3 to trip distribution, σ_4 to modal split, and σ_5 to traffic assignment.

Consistency and Beta Conditions

The question that naturally arises is, Under what conditions will both direct and indirect approaches give the same results? Although the question cannot be answered in general, one necessary (but not sufficient) condition can be identified: The sequential implicit system must be internally consistent, defined as follows.

We would like the set of volumes V produced by a sequential implicit system (Eq. 11) to meet this obvious condition:

$$\left.\begin{aligned} V_T &= \sum_k V_k \\ V_k &= \sum_l V_{kl} \\ V_{kl} &= \sum_m V_{klm} \\ V_{klm} &= \sum_p V_{klmp} \end{aligned}\right\} \tag{13}$$

for all Y.

If a sequential implicit system $V = \psi(V,Y)$ produces volumes V that meet the conditions in Eq. 13 for all Y, we say it is internally consistent.

By Eq. 11, this leads immediately to the following necessary and sufficient beta conditions for a sequential implicit system to be internally consistent:

$$\left.\begin{aligned}
\sigma_1(Y) &= \sum_k \sigma_2(V_T, Y) \\
\sigma_2(V_T, Y) &= \sum_l \sigma_3(V_k, Y) \\
\sigma_3(V_k, Y) &= \sum_m \sigma_4(V_{kl}, Y) \\
\sigma_4(V_{kl}, Y) &= \sum_p \sigma_5(V_{klm}, Y)
\end{aligned}\right\} \quad (14)$$

These conditions seem reasonable by themselves. They are also necessary conditions if the explicit and sequential implicit forms of a model are to give the same equilibrium flow-pattern predictions.

GENERAL SHARE MODEL

The general share model (GSM) is defined as

$$V_{klmp} = \alpha(Y) \cdot \beta_k(Y) \cdot \gamma_{kl}(Y) \cdot \delta_{klm}(Y) \cdot \omega_{klmp}(Y) \quad (15)$$

where $Y = f(R,Z) = f(\underline{A}, \underline{a}, \underline{X}, \underline{w})$ as in the notation defined earlier, and α, β, γ, δ, ω are functions that meet the following range conditions for all values of Y:

$$\left.\begin{aligned}
&0 \le \alpha(Y) \\
&0 \le \beta_k \le 1, \sum_k \beta_k(Y) = 1 \\
&0 \le \gamma_{kl}(Y) \le 1, \sum_l \gamma_{kl}(Y) = 1 \text{ for every } k \\
&0 \le \delta_{klm}(Y) \le 1, \sum_m \delta_{klm}(Y) = 1 \text{ for every } k, l \\
&0 \le \omega_{klmp}(Y) \le 1, \sum_p \omega_{klmp}(Y) = 1 \text{ for every } k, l, m
\end{aligned}\right\} \quad (16)$$

The name "share model" is used because each of the terms α, β, γ, δ, and ω "splits" the flow volume into successive "shares." (These relations can be derived by summing Eq. 15 over the various subscripts p, m, l, and then k and using Eq. 16.) Hence,

1. The total level of travel in the region is given by $\alpha(Y)$:

$$V_T = \alpha(Y) \quad (17)$$

2. Of this total travel, the fraction that originates in zone k is given by $\beta_k(Y)$:

$$V_k = \beta_k \cdot V_T \quad (18)$$

3. The fraction that originates in zone k and has zone l as its destination is given by $\gamma_{kl}(Y)$:

$$V_{kl} = \gamma_{kl} \cdot V_k \quad (19)$$

4. The fraction that goes from zone k to zone l and uses mode m (the modal split) is given by $\delta_{klm}(Y)$:

$$V_{klm} = \delta_{klm} \cdot V_{kl} \quad (20)$$

5. Finally, the fraction that goes from zone k to zone l and uses path p of mode m is given by $\omega_{klmp}(Y)$:

$$V_{klmp} = \omega_{klmp} \cdot V_{klm} \tag{21}$$

These relations can be derived from the basic definition of the GSM.

Thus, although the GSM is itself explicit, we can also write the GSM (Eq. 15) in an alternative sequential implicit form:

$$\left.\begin{aligned} V_T &= \alpha(Y) \\ V_k &= \beta_k(Y) \cdot V_T(Y) \\ V_{kl} &= \gamma_{kl}(Y) \cdot V_k(Y) \\ V_{klm} &= \delta_{klm}(Y) \cdot V_{kl}(Y) \\ V_{klmp} &= \omega_{klmp}(Y) \cdot V_{klm}(Y) \end{aligned}\right\} \tag{22}$$

with conditions (Eq. 16) as before.

This is analogous to the sequential implicit demand model used in the UTMS indirect approach to equilibrium, as follows:

$$\left.\begin{aligned} V_k &= \beta_k \cdot \alpha \\ V_{kl} &= \gamma_{kl} \cdot V_k \\ V_{klm} &= \delta_{klm} \cdot V_{kl} \\ V_{klmp} &= \omega_{klmp} \cdot V_{klm} \end{aligned}\right\} \tag{23}$$

for trip generation, trip distribution, modal split, and traffic assignment respectively. In a later section, we show how this analogy enables us to overcome the biases and limitations of the UTMS as currently implemented.

PROPERTIES OF THE GENERAL SHARE MODEL

The following important properties of the GSM can be demonstrated. Details are given in another paper (1).

Theorem 1: The GSM can be expressed in both explicit and sequential implicit forms; the sequential implicit form is internally consistent.

Theorem 2: Any explicit demand model system can be expressed as a GSM.

Theorem 3: Any internally consistent, sequential implicit demand system can be expressed as a GSM.

Theorem 4: For every explicit demand system, there is a corresponding internally consistent sequential implicit system, and conversely.

The proof of theorem 1 follows from Eqs. 15, 22, and 16 and the beta conditions (Eq. 14).

The proof of theorem 2, although simple, is interesting because it offers a constructive method for getting the GSM for an explicit system:

1. Begin with the explicit model,

$$V_{klmp} = f(Y) \tag{24}$$

2. Derive the various partial sums,

$$V_{klm.},\ V_{kl..},\ V_{k...},\ V_{....} \tag{25}$$

by

$$V_{klm.} = \sum_{p} V_{klmp} \text{ and so on} \tag{26}$$

3. Derive the "split fractions,"

$$\left.\begin{aligned}
\omega_{klmp} &= \frac{V_{klmp}(Y)}{V_{klm.}(Y)} \\
\delta_{klm} &= \frac{V_{klm.}(Y)}{V_{kl..}(Y)} \\
\gamma_{kl} &= \frac{V_{kl..}(Y)}{V_{k...}(Y)} \\
\beta_k &= \frac{V_{k...}(Y)}{V_{....}(Y)} \\
\alpha &= V_{....}(Y)
\end{aligned}\right\} \quad (27)$$

4. Construct the GSM as the product of these split fractions,

$$V_{klmp} = \alpha \cdot \beta_k \cdot \gamma_{kl} \cdot \delta_{klm} \cdot \omega_{klmp} \quad (28)$$

In particular, all of the existing explicit demand models, such as the Kraft-SARC, McLynn, and Baumol-Quandt, are special cases of the GSM.

The proof of theorem 3 is also constructive and follows from the definition of a sequential implicit model (Eq. 11):

$$\left.\begin{aligned}
\alpha &\equiv V_T = \sigma_1(Y) \\
\beta_k &\equiv V_k/V_T = \frac{\sigma_2(V_T, Y)}{\sigma_1(Y)} \\
\gamma_{kl} &\equiv V_{kl}/V_k = \frac{\sigma_3(V_k, Y)}{\sigma_2(V_T, Y)} \\
\delta_{klm} &\equiv V_{klm}/V_{kl} = \frac{\sigma_4(V_{kl}, Y)}{\sigma_3(V_k, Y)} \\
\omega_{klmp} &\equiv V_{klmp}/V_{klm} = \frac{\sigma_5(V_{klm}, Y)}{\sigma_4(V_{kl}, Y)}
\end{aligned}\right\} \quad (29)$$

Because the sequential model is specified as internally consistent, we know that the beta conditions (Eq. 14) will apply, and, therefore, Eq. 29 will meet the range conditions on the GSM (Eq. 16).

The proof of theorem 4 follows from the preceding theorems. Given any explicit system, by theorem 2 we can get a corresponding GSM. By theorem 1, this GSM has a corresponding sequential implicit form that is internally consistent, completing the first part of the proof. To show the converse, given any internally consistent sequential implicit system, by theorem 3 we find a corresponding GSM, and then by theorem 1 we have a corresponding explicit form.

These results, especially theorem 4, have very important practical implications, as described in the next section.

Additional properties of the GSM are described elsewhere (1). To summarize: The elasticities of the GSM and its components take a particularly useful form, which suggests directions for development of efficient equilibrium algorithms. The form of the GSM can be given a probabilistic interpretation, based on the range conditions (Eq. 16); this can lead to an explicit bridge between disaggregate stochastic models and aggregate models. The travel behavior of different market segments can be explicitly represented in different special cases of the GSM.

IMPLICATIONS

Because of space limitations, only a few of the major implications of the theoretical results can be presented here.

Families of Demand Models

The theorems given in the preceding section indicate the very general nature of the GSM. It is particularly interesting to explore various families of specific demand models that arise as special cases of the GSM and to see the relation of existing models to these.

In this discussion, for simplicity, we will ignore the complications introduced by assignment of flows to paths (i.e., network assignment) by assuming only 1 path p of each mode m and dropping the subscript p. This assumption can easily be relaxed.

Recall that α, β, γ, and δ are all functions of $Y = f(Z,R)$, where $Z = f(\underline{A},\underline{a})$ and $R = f(\underline{X},\underline{w})$. These are very general and provide for a great deal of flexibility in designing specific demand models.

For example, consider first the level of service (we drop the subscripts k, l, and m for the moment) $L = \underline{X} = (x_1, x_2, \ldots, x_i, \ldots, x_s)$ where perhaps we have the following specific level-of-service variables: x_1 = travel time, in-vehicle portion; x_2 = travel time, out-of-vehicle portion; x_3 = out-of-pocket cost; x_4 = frequency of service; and so on.

Alternative forms of R are as follows:

1. $R = w_0 + w_1x_1 + w_2x_2 + w_3x_3$;
2. $R = w_0e^U$, where $U = w_1x_1 + w_2x_2 + w_3x_3$;
3. $R = f(x_1)$;
4. $R = w_0x_1^{w_1}x_2^{w_2}x_3^{w_3}x_4^{w_4}$;
5. $R_m = \dfrac{w_{m0}\,x_{m1}^{w_1}x_{m2}^{w_2}}{\sum_q w_{q0}\,x_{q1}^{w_1}x_{q2}^{w_2}}$; and
6. $R_m = w_{m0}\,(x_{m1}^{w_{m1}}x_{m2}^{w_{m2}}) \prod_{q \neq m} R_{mq}$, where $R_{mq} = x_{q1}^{w_{mq1}}x_{q2}^{w_{mq2}}$.

Form 1 shows a value-of-time formulation, where w_2/w_1 expresses the relative values placed on in-vehicle and out-of-vehicle times, and w_1/w_3 expresses the value-of-time equivalency in cents per minute. Form 2 shows an exponential transform of a linear cost as used in the Twin Cities modal-split model (25). In form 3, f corresponds to a friction-factor transformation of travel time as used in typical gravity model applications. Form 4 is a general product form, and form 5 is that used in the McLynn model. Form 6 is the form used in the Kraft-SARC model, where w_{mi} is a direct elasticity, and w_{mqi} is a cross elasticity, reflecting the effect of mode q's level of service on travel by mode m.

The generality of the GSM should be clear from these examples and from theorem 4. As shown in Figure 2, the GSM includes as special cases all of the explicit demand models developed to date—for example, the Baumol-Quandt, McLynn, and Kraft-SARC are all special cases of the general direct demand model (GDDM), which is in turn a special case of the GSM. Further, another sequence of models can be formulated: the special product models (SPM), of which a modified form of the UTMS is a special case.

UTMS Models: Special Product Models

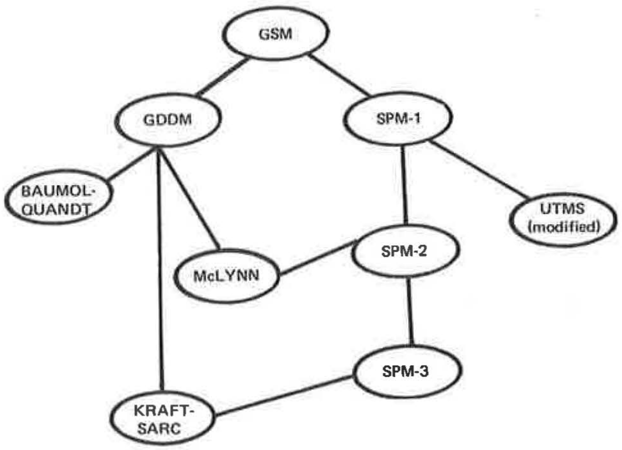
We begin with a very general form, the special product model 1 (SPM-1), defined as

$$V_{klm} = \left(\frac{R_{klm}}{R_{kl.}}\right)\left(\frac{Z_1 \cdot R_{kl.}^{\delta_{10}}}{\sum_1 Z_1 \cdot R_{kl.}^{\delta_{10}}}\right)\left[\frac{Z_k\left(\sum_1 Z_1R_{kl.}^{\delta_{11}}\right)^{\delta_{20}}}{\sum_k Z_k\left(\sum_1 Z_1R_{kl.}^{\delta_{11}}\right)^{\delta_{20}}}\right]\left\{\left[\sum_k Z_k\left(\sum_1 Z_1R_{kl.}^{\delta_{12}}\right)^{\delta_{21}}\right]^{\delta_{30}}\right\} \quad (30)$$

In sequential form, SPM-1 is as follows: trip generation, total,

$$V_T = \left[\sum_k Z_k\left(\sum_1 Z_1R_{kl.}^{\delta_{12}}\right)^{\delta_{21}}\right]^{\delta_{30}} \quad (31)$$

Figure 2. Families of models.



trip generation, zonal,

$$V_k = V_T \left[\frac{Z_k \left(\sum_1 Z_1 R_{kl.}^{\delta_{11}} \right)^{\delta_{20}}}{\sum_k Z_k \left(\sum_1 Z_1 R_{kl.}^{\delta_{11}} \right)^{\delta_{20}}} \right] \tag{32}$$

trip distribution,

$$V_{kl} = V_k \left(\frac{Z_1 R_{kl.}^{\delta_{10}}}{\sum_1 Z_1 R_{kl.}^{\delta_{10}}} \right) \tag{33}$$

and modal split,

$$V_{klm} = V_{kl} \left(\frac{R_{klm}}{R_{kl.}} \right) \tag{34}$$

If there is more than 1 path in each mode, traffic assignment can be added,

$$V_{klmp} = \left(\frac{R_{klmp}}{R_{klm.}} \right) V_{klm}$$

and corresponding changes made in Eq. 30.

From Eq. 34, the resistance R_{klm} should be a positive conductance measure; e.g., $R_{klm} = t_{klm}^{-a}$. Thus, the signs of δ_{10}, δ_{11}, and δ_{12} should be positive also.

To see the relation to the UTMS, we can examine a special case of SPM-1, obtained by setting the parameters as follows:

$$\delta_{10} = \delta_{11} = \delta_{12} \equiv \delta_1 \tag{35}$$

$$\delta_{20} = \delta_{21} \equiv \delta_2 \tag{36}$$

$$\delta_{30} \equiv \delta_3 \tag{37}$$

This leads to special product model 2 (SPM-2):

$$V_{klm} = \left(\frac{R_{klm}}{R_{kl.}}\right)\left(\frac{Z_l R_{kl.}^{\delta_1}}{\sum_l Z_l R_{kl.}^{\delta_1}}\right)\left[\frac{Z_k\left(\sum_l Z_l R_{kl.}^{\delta_1}\right)^{\delta_2}}{\sum_k Z_k\left(\sum_l Z_l R_{kl.}^{\delta_1}\right)^{\delta_2}}\right]\left\{\left[\sum_k Z_k\left(\sum_l Z_l R_{kl.}^{\delta_1}\right)^{\delta_2}\right]^{\delta_3}\right\} \tag{38}$$

SPM-2 can also be written as follows:

$$V_{klm} = R_{klm} \cdot Z_k \cdot Z_l \left\{ R_{kl.}^{\delta_1 - 1} \left(\sum_l Z_l R_{kl.}^{\delta_1}\right)^{\delta_2 - 1} \left[\sum_k Z_k\left(\sum_l Z_l R_{kl.}^{\delta_1}\right)^{\delta_2}\right]^{\delta_3 - 1} \right\} \tag{39}$$

Now we introduce one further set of special conditions:

$$\delta_1 = \delta_2 = \delta_3 \equiv 1 \tag{40}$$

which leads to special product model 3 (SPM-3):

$$V_{klm} = R_{klm} \cdot Z_k \cdot Z_l \tag{41}$$

To see the relationship with the standard UTMS, we consider the sequential form of SPM-2 as follows: trip generation, total,

$$V_T = \left[\sum_k Z_k\left(\sum_l Z_l R_{kl.}^{\delta_1}\right)^{\delta_2}\right]^{\delta_3} \tag{42}$$

trip generation, zonal,

$$V_k = V_T \left[\frac{Z_k\left(\sum_l Z_l R_{kl.}^{\delta_1}\right)^{\delta_2}}{\sum_k Z_k\left(\sum_l Z_l R_{kl.}^{\delta_1}\right)^{\delta_2}}\right] \tag{43}$$

trip distribution,

$$V_{kl} = V_k \left(\frac{Z_l R_{kl.}^{\delta_1}}{\sum_l Z_l R_{kl.}^{\delta_1}}\right) \tag{44}$$

and modal split,

$$V_{klm} = V_{kl}\left(\frac{R_{klm}}{R_{kl.}}\right) \tag{45}$$

In its first form, the classical gravity model of early transportation modeling was simply

$$V_{kl} = Z_k \cdot Z_l \cdot R_{kl} \tag{46}$$

where Z_k was the population at $k(B_k)$; Z_l was the population or employment at $l(B_l)$; and R_{kl} was distance or travel time to the power θ, where θ was approximately -2:

$$V_{kl} = \frac{B_k \cdot B_l}{t_{kl}^{-\theta}} \tag{47}$$

SPM-3 (Eq. 41) is simply a generalized gravity model. In later forms, the gravity model was normalized in this way:

$$
\begin{aligned}
V_{kl} &= \left(\frac{Z_l \cdot R_{kl}}{\sum_l Z_l \cdot R_{kl}} \right) Z_k \\
&= B_k \left(\frac{B_l t_{kl}^{\theta}}{\sum_l B_l t_{kl}^{\theta}} \right)
\end{aligned}
\tag{48}
$$

In this form, the term

$$P_{kl} = \frac{B_l}{t_{kl}^{-\theta}} \tag{49}$$

is referred to as the "potential" of zone k. The term

$$P_{k.} = \sum_l \left(\frac{B_l}{t_{kl}^{-\theta}} \right) \tag{50}$$

is called the "accessibility," reflecting a weighted average of the attractiveness of the various destinations as measured by B_l, weighted by the difficulty of access to those destinations measured by $t_{kl}^{-\theta}$.

Now, to be more general, we can define a generalized potential as

$$P_{kl} \equiv Z_l \cdot R_{kl}^{\delta_1}. \tag{51}$$

and a generalized accessibility as

$$P_{k.} = \sum_l Z_l R_{kl}^{\delta_1}. \tag{52}$$

where we allow a number of variables to enter the Z and R terms; and in particular, we have cross-elasticity terms $R_{klmq,p}$ in R (as in form 6 given in an earlier section).

Thus, the distribution stage of SPM-2 (Eq. 44) is a generalized gravity model, with population B_k replaced by the more general measure of the intensity of the activity system $Z = f(\underline{A}_k, \underline{a})$ and time t_{kl}^{θ} replaced by a more general measure of the resistance of the transportation system R_{kl}.

Now let us examine the generation stage. We substitute P_{kl} and $P_{k.}$ and get

$$V_T = \left(\sum_k Z_k \cdot P_{k.}^{\delta_2} \right)^{\delta_3} \tag{53}$$

$$V_k = V_T \left(\frac{Z_k \cdot P_{k.}^{\delta_2}}{\sum_k Z_k \cdot P_{k.}^{\delta_2}} \right) \tag{54}$$

Thus, to get total trips in the region, we calculate a trip-generating potential,

$$G_k \equiv Z_k \cdot P_k^{\delta_2} \tag{55}$$

This potential reflects both the level of the activity system at k and Z_k and the influence of $P_{k.}$, the accessibility of k (an average over all destinations). The exponent δ_2 scales the effect of accessibility on trip generation (and can be zero when appropriate). The total trips generated in the region V_T depends on G., the sum of these trip-generating potentials G_k for all zones; the exponent δ_3 establishes the extent to which V_T is sensitive to G. The trips generated in any zone k are proportional to its share of this sum of potentials:

$$V_k = V_T \left(\frac{G_k}{G.}\right), \quad V_T = G.^{\delta_3} \tag{56}$$

In a sense, then, the generation stage of SPM-2 is also a generalization of the classical gravity model. However, instead of assuming that V_k is a constant, independent of level of service, as in the UTMS, we are establishing V_k as a function of level of service consistent with the gravity-model-like approach to distribution.

Thus, as this sequential form shows, SPM-2 (Eq. 38) is a more general version of the UTMS, but with the following properties:

1. Any desired level-of-service variables, with any desired direct and cross elasticities, can be incorporated in the resistance R_{klm} used to characterize a particular mode.
2. The same resistance term enters every step of the process—generation and distribution as well as modal split—in a consistent way, either as R_{klm} or as part of the sum $R_{kl.}$.
3. To compute a valid equilibrium, we can use the explicit form as in Eq. 38 to accomplish the calculations in one step, using a direct approach. If there should be some reason why we want to compute equilibrium in an indirect manner, we can use the sequential form (Eqs. 42, 43, 44, and 45), which is internally consistent.
4. The various parameters δ establish the influence of level of service on various aspects of travel behavior and may have different values to reflect the behavior of different segments of the urban travel market (1). (SPM-1 is even more general and provides even more flexibility for the analyst in developing alternative special cases for different conditions.)
5. The model can be estimated by using standard statistical techniques (26). Therefore, we have met all of the conditions outlined earlier in this paper.

CONCLUSIONS AND RECOMMENDATIONS

Although the results presented here are relatively theoretical and abstract, they have practical implications. These implications are summarized in the following sections in the form of recommendations.

Summary of the Theoretical Results

Before specific recommendations are described, it will be useful to summarize the major features of the theoretical results that have been presented and their implications:

1. A desirable property of a sequential implicit system is that it be internally consistent (Eq. 13) for which the beta conditions (Eq. 14) are necessary and sufficient.
2. The GSM (Eq. 15) has both explicit and sequential implicit forms; the sequential implicit form is internally consistent. Equilibrium can be computed in the direct approach with the explicit form or in the indirect approach with the sequential implicit form.
3. Any explicit demand model can be expressed as a GSM. For example, the Northeast Corridor and the urban models (Kraft-Domencich-Vallette and Plourde) are all special cases of the GSM.
4. Any internally consistent, sequential implicit demand system can be expressed as a GSM.
5. For every explicit demand model, there is a corresponding internally consistent sequential implicit form, and conversely.
6. As a consequence,

   a. We are completely free to choose whether to compute equilibrium via the direct or indirect approaches (if direct, we use the explicit form, or, if indirect, the sequential implicit form);

   b. The UTMS is a sequential implicit system, can be modified to be internally consistent, and in modified form can be expressed as a GSM (therefore, the shortcomings of the indirect approach to computing equilibrium can be overcome by using the explicit form of the modified UTMS in a direct approach to computing equilibrium); and

c. Any explicit demand model can be expressed in its corresponding sequential implicit form and used in an indirect approach to equilibrium (thus, the analyst can have the capability, if desired, to control each step of a travel prediction with an explicit model in the same way each step of the UTMS can be controlled).

7. The GSM suggests a variety of possible demand models. In particular, a series of special product models (SPM-1, SPM-2, and SPM-3) provide a family of model forms (Eqs. 30 to 45), which are similar to the UTMS, the classical gravity model, and explicit models such as the Kraft-SARC and the McLynn, but provide a rich variety of options. These options can be used to represent the travel behavior patterns of different market segments. (Another attractive feature of the GSM arises from the properties of its elasticities. These suggest the development of efficient approximation techniques to explore the effects on flow volumes of small changes in the transportation system.)

Attitudinal Change

The UTMS was a major accomplishment for its time. The profession and the governmental transportation agencies (federal, state, and local) must recognize that the UTMS is no longer satisfactory; it is neither relevant to the practical issues that must be addressed in the urban transportation studies of today (1, 27) nor acceptable when viewed from a theoretical perspective. The UTMS should be neither completely discarded nor allowed to remain unchanged as the basic working tool of urban transportation analysis.

A new generation of transportation analysis tools is required. Development of new systems should build on the several directions of current research, as well as the practical experience gained from the UTMS. The recommendation is that we begin by asking not whether but how.

A Direction

We cannot here lay out the preliminary design of a new generation of urban transportation models. However, the theoretical results presented in earlier sections suggest one possible approach. Recommendations are as follows:

1. A model system should treat the transportation system of a region as a single multimodal system, taking each trip from door to door through any possible mix of transport facilities.
2. A model system should allow explicit treatment of any number of market segments and should allow each market segment to have different behavior patterns. These differences may be expressed not only in the values of the parameters of the demand functions (e.g., the values of direct and cross elasticities) but also in the structural forms of the functions (e.g., as represented by the different cases of SPM-1 and SPM-2).
3. A system should have capability for including explicitly any desired set of level-of-service variables.
4. A system should have a valid procedure for computing equilibrium of supply and demand within the network, considering the interaction of all market segments.
5. To implement recommendation 4 will probably require a direct approach using explicit demand models. It should be implemented in a single integrated system of computer programs. (Although DODOTRANS is one candidate for this, more efficient and satisfactory procedures can certainly be developed. Ultimately, a technique may be developed for computing a valid equilibrium with the indirect approach. Until that time, the direct approach should be used.)
6. The GSM theorems (especially theorem 4) suggest that a single model system could be developed to compute equilibrium in the direct approach by using the explicit form of the GSM:

a. Any of a large variety of specific functional forms for functions such as α and β can be provided; the user can select those he wishes to use (and the corresponding sets of parameter values) when he makes his run to analyze a particular transportation plan. The same equilibrium computational procedures would be used for every form. (Ideally, the computational procedures should allow simultaneous use of several different forms corresponding to different market segments.)

b. As operational experience with specific functional forms is gained, additional special-purpose algorithms can be developed and added to the system for more efficient computation of equilibrium for specific forms of demand models.

c. The availability of alternative forms in a single model system would allow the analyst to do sensitivity studies of alternative models as well as of alternative parameter values.

7. In the same model system, options should be provided to compute equilibrium with an indirect approach by using the sequential implicit form of the GSM:

a. As a minimum, the modified UTMS forms (SPM-1, SPM-2, and SPM-3) should be provided in the indirect approach (obviously, levels of service would thus enter into every step of the indirect approach).

b. As an alternative, the direct approach can be used, and the parameters δ of a general form such as SPM-1 can be used by the analyst to control intermediate totals (e.g., V_{kl} and V_k).

8. The same model system should also have the pivot-point capabilities suggested by the elasticity properties of the GSM (1, 28, 29).

9. A major program of research should be mounted to develop specific demand models for a variety of different market segments and under different conditions of urban area life style and transportation system. Theoretical research should explore the properties of various families of specific models. Empirical research should attempt to get results that can be generalized across many urban areas through careful research design. The GSM may be considered as one hypothesis (of many) to be tested and can serve as a framework for development of specific models.

10. Research should be undertaken to develop efficient procedures for computing equilibrium in networks. Priority should be given to direct approaches.

Immediate Actions

The recommendations presented in the preceding section will likely take a few years to accomplish, even if a decision by those in control of the resources were taken today. The theoretical results also suggest specific practical steps that can be taken almost immediately:

1. Exploration should be undertaken of immediate modifications that can be made to existing model systems (the FHWA-UTMS packages, DODOTRANS, and others) to implement capabilities of analyzing a single multimodal network, handling a range of level-of-service variables, iterating all steps in the indirect approach to achieve convergence to equilibrium, having level of service influence each step—including generation and distribution as well as modal split and assignment—in a consistent manner, and representing the behavior of different market segments.

2. Modifications of existing systems should be developed to implement SPM-3, SPM-2, and SPM-1 to obtain acceptable forms of UTMS-like capabilities.

3. Immediate efforts should be initiated to develop pilot or experimental model systems that compute equilibrium in a direct approach with the GSM.

CLOSING

This presentation has ranged from some fairly abstract to some relatively practical issues. We invite discussion and debate on both the theoretical results and the policy recommendations arising from those results.

ACKNOWLEDGMENTS

The intellectual roots of this paper are many. Martin Wohl first convinced me of the need to view transportation systems analysis from the perspective of economic theory; and it was some observations on the properties of elasticities in a paper by Alan Goldman and James McLynn that first stimulated my thinking along the path described here. Discussions with Earl Ruiter, Daniel Brand, Len Sherman, Alain Bieber, and especially Moshe Ben-Akiva fueled this development at particular points. I am particularly grateful to Britton Harris, Gerald Kraft, Damian Kulash, Peter Stopher, and Alan Wilson for comments on an earlier draft. The Israel Ministry of Transport gave me the stimulus and opportunity to focus this research during several periods of work in Israel for which I am indebted to Uri Ben-Efraim and Gideon Hashimshony. To A. Scheffer Lang and C. L. Miller of M.I.T., I acknowledge with deep gratitude the guidance, freedom, and support they have given me during the years. I would be pleased to acknowledge research sponsors if there were any to acknowledge. Someday, perhaps, there will be support available explicitly for fundamental research in transportation systems analysis.

REFERENCES

1. Manheim, M. L. Fundamental Properties of Systems of Travel Demand Models. In press.
2. Beckman, M., McQuire, C. B., and Winston, C. B. Studies in the Economics of Transportation. Yale Univ. Press, New Haven, 1956.
3. Kraft, G., and Wohl, M. New Directions for Passenger Demand Analysis and Forecasting. Transportation Research, Vol. 1, 1967, pp. 205-230.
4. Wohl, M., and Martin, B. V. Traffic Systems Analysis. McGraw-Hill, New York, 1967, Ch. 5.
5. Manheim, M. L. Search and Choice in Transport Systems Planning. Highway Research Record 293, 1969, pp. 54-82.
6. Wardrop, J. G. Some Theoretical Aspects of Road Traffic Research. Institute of Civil Engineers, London, Road Paper 36, 1952.
7. Beckmann, M. On the Theory of Traffic Flow in Networks. Traffic Quarterly, Jan. 1967.
8. Jewell, W. S. Models for Traffic Assignment. Transportation Research, Vol. 1, No. 1, May 1967.
9. Martin, B. V., Memmott, F. W., and Bone, A. J. Principles and Techniques of Predicting Future Urban Area Transportation. M.I.T. Press, Cambridge, 1965.
10. U. S. Bureau of Public Roads. Various publications.
11. Demand for Intercity Passenger Travel in the Washington-Boston Corridor. Systems Analysis and Research Corp., 1963.
12. Quandt, R. E., and Baumol, W. J. Abstract Mode Model: Theory and Measurement. Jour. of Regional Science, Vol. 6, No. 2, 1966.
13. McLynn, J. M., et al. A Family of Demand Modal Split Models. Arthur Young and Co., April 1969.
14. McLynn, J. M., and Watkins, R. H. Multimode Assignment Model. National Bureau of Standards, 1965.
15. Plourde, R. Consumer Preference and the Abstract Mode Model: Boston Metropolitan Area. Dept. of Civil Eng., M.I.T., Cambridge, Res. Rept. R68-51, 1968.
16. Domencich, T. A., Kraft, G., and Vallette, J. Estimation of Urban Passenger Travel Behavior: An Economic Demand Model. Highway Research Record 238, 1968, pp. 64-78.
17. 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.
18. Stopher, P. R., and Lisco, T. E. Modelling Travel Demand: A Disaggregate Behavioral Approach—Issues and Application. In Proc., Transportation Research Forum, Richard B. Cross Co., Oxford, Ind., 1970.
19. Reichman, S., and Stopher, P. R. Disaggregate Stochastic Models of Travel-Mode Choice. Highway Research Record 369, 1971, pp. 91-103.
20. Wilson, A. G. The Use of Entropy Maximizing Models. Jour. of Transport Economics and Policy, Vol. 3, No. 1, Jan. 1969, pp. 108-126.

21. Kresge, D. T., and Roberts, P. O. Techniques of Transport Planning: Systems Analysis and Simulation Models. Brookings Institution, Washington, D.C., 1971.
22. Wilkie, D. F., and Stefanek, R. G. Precise Determination of Equilibrium in Travel Forecasting Problems Using Numerical Optimization Techniques. Highway Research Record 369, 1971, pp. 239-252.
23. Dafermos, S. C. An Extended Traffic Assignment Model With Applications to Two-Way Traffic. Transportation Science, Vol. 5, No. 4, Nov. 1971, pp. 366-389.
24. Kulash, D. C. A Transportation Equilibrium Model. The Urban Institute, Washington, D.C., Rept. 708-45, Sept. 1971.
25. Pratt, R. H., and Deen, T. Estimation of Sub-Modal Split Within the Transit Mode. Highway Research Record 205, 1967, pp. 20-30.
26. Ben-Akiva, M. E. Dept. of Civil Eng., M.I.T., Cambridge, PhD dissertation.
27. Transportation Analysis: Past and Prospects. Highway Research Record 309, 1970, 44 pp.
28. Loubal, P. S. A Network Evaluation Procedure. Highway Research Record 205, 1967, pp. 96-109.
29. Ridley, T. M. An Investment Policy to Reduce the Travel Time in a Transportation Network. Transportation Research, Vol. 2, No. 4, Dec. 1968, pp. 409-424.