TRANSPORTATION RESEARCH RECORD 728

Passenger Travel Forecasting

TRANSPORTATION RESEARCH BOARD

COMMISSION ON SOCIOTECHNICAL SYSTEMS
NATIONAL RESEARCH COUNCIL

NATIONAL ACADEMY OF SCIENCES
WASHINGTON, D.C. 1979

Transportation Research Record 728
Price $5.00
Edited for TRB by Susan Singer-Bart

modes
1 highway transportation
2 public transit

subject areas
12 planning
13 forecasting

Transportation Research Board publications are available by ordering directly from TRB. They may also be obtained on a regular basis through organizational or individual affiliation with TRB; affiliates or library subscribers are eligible for substantial discounts. For further information, write to the Transportation Research Board, National Academy of Sciences, 2101 Constitution Avenue, N.W., Washington, DC 20418.

Notice
The papers in this Record have been reviewed by and accepted for publication by knowledgeable persons other than the authors according to procedures approved by a Report Review Committee consisting of members of the National Academy of Sciences, the National Academy of Engineering, and the Institute of Medicine.

The views expressed in these papers are those of the authors and do not necessarily reflect those of the sponsoring committee, the Transportation Research Board, the National Academy of Sciences, or the sponsors of TRB activities.

Library of Congress Cataloging in Publication Data
National Research Council. Transportation Research Board.
Passenger travel forecasting.

(Transportation research record; 728)
1. Transportation–Passenger traffic–Addresses, essays, lectures.
2. Urban transportation–Addresses, essays, lectures. 3. Choice of transportation–Addresses, essays, lectures. I. Title. II. Series.
TE7.H5 no. 728 [HE151] 380.5s [388.4'1321]
ISBN 0-309-02981-3 80-607015
ISSN 0361-1981

Sponsorship of the Papers in This Transportation Research Record

GROUP 1–TRANSPORTATION SYSTEMS PLANNING AND ADMINISTRATION
Leon M. Cole, Library of Congress, chairman

Transportation Forecasting Section
George V. Wickstrom, Metropolitan Washington Council of Governments, chairman

Committee on Passenger Travel Demand Forecasting
Daniel Brand, Massachusetts Executive Office of Transportation and Construction, chairman
Moshe Ben-Akiva, Robert T. Dunphy, Raymond H. Ellis, Robert E. Gall, David S. Gendell, Michael B. Godfrey, Thomas F. Golob, Walter G. Hansen, David T. Hartgen, Kevin E. Heanue, Thomas J. Hillegass, Joel L. Horowitz, Gerald Kraft, Eugene J. Lessieu, Robert E. Paaswell, Frederick A. Reid, Martin G. Richards, Peter R. Stopher, Vergil G. Stover, Anthony R. Tomazinis, Edward Weiner, Charles C. Whitmarsh, Martin Wohl

Committee on Traveler Behavior and Values
David T. Hartgen, New York State Department of Transportation, chairman
M. E. Beesley, Nils Anders Bruzelius, Katherine Patricia Burnett, Melvyn Cheslow, Ricardo DePaul Dobson, Thomas F. Golob, David A. Hensher, Frank S. Koppelman, Peter S. Liou, Jordan J. Louviere, Daniel McFadden, Arnim H. Meyburg, Richard M. Michaels, Robert E. Paaswell, Shalom Reichman, Earl R. Ruiter, Geoffrey A. C. Searle, Peter R. Stopher, Antti P. Talvitie, Peter L. Watson

James A. Scott, Transportation Research Board staff

Sponsorship is indicated by a footnote at the end of each report. The organizational units and officers and members are as of December 31, 1978.

Contents

Equilibrium Trip Assignment: Advantages and Implications for Practice

R. W. Eash, Chicago Area Transportation Study
B. N. Janson and D. E. Boyce, University of Illinois, Urbana-Champaign

During the past 10 years the problem of assignment of vehicles to large, congested urban transportation networks according to the principle of equal travel times has been solved and an efficient, convergent computer algorithm devised. Although the algorithm is available in the Urban Transportation Planning System, many practitioners continue to use the heuristic trip-assignment algorithms devised in the early 1960s. As in many other cases, this slow implementation of a new, improved algorithm appears to come from (a) a lack of understanding of its basic concepts, (b) an unfamiliarity with the computer program for applying the algorithm, and (c) a lack of evidence concerning the new algorithm's performance in large-scale applications. These three issues are addressed in this paper. Based on the experience with its implementation on a large network, it is recommended that equilibrium trip assignment should always be used instead of iterative assignment. Better results, as judged by the criterion of equalizing travel times for alternative paths between each origin-destination pair, will always be obtained with the equilibrium algorithm for any given amount of computational effort. Which method best replicates the observed vehicle flows may depend on the detail of the network, the adequacy of the capacity-restraint functions, and the time period of the assignment (24 h or peak period).

Assignment of vehicles to large, congested urban transportation networks has been a problem of interest to transportation planners and researchers for over two decades. Initially, heuristic or approximate solution techniques were developed for the problem. Later, several convergent algorithms were devised and some were tested, culminating in an International Symposium on Traffic Equilibrium Methods at the University of Montreal in 1974 (1-3).

Despite these theoretical and practical developments, relatively few applications are being made of equilibrium assignment, despite its availability in the Urban Transportation Planning System (UTPS) (4) and its desirable attributes. Two reasons are apparent for this situation.

1. Practitioners have experienced difficulty in understanding the formulation of the equilibrium-assignment problem and the algorithms devised to solve it, and
2. Practitioners were uncertain about whether the algorithms were superior to competing algorithms, such as iterative and incremental assignment, for large networks.

This paper will explain the equilibrium-assignment problem and the algorithm in terms that are familiar to practitioners and report on a large-scale, prototype implementation of the model. The implementation provides convincing evidence that equilibrium assignment is the method of choice for congested networks. The shortcomings of existing capacity-restraint functions and the weaknesses of 24-h assignments are evident from this application.

The problem of trip assignment in the sequential urban travel-forecasting process is how to assign (or allocate) a specified number of vehicles (or persons) to the paths taken from each origin to each destination. The path chosen by each traveler is generally assumed to be the path that minimizes his or her journey time, or some combination of time and cost. All travelers are assumed to have identical perceptions of travel time and cost. If the network is congested, that is, if each link's travel time depends on the flow of vehicles on that link, then the following equilibrium problem results: Find the assignment of vehicles to links such that no traveler can reduce his or her travel time from origin to destination by switching to another path. These equilibrium conditions were stated by Wardrop (5) and are commonly referred to as the Wardrop conditions.

The user-equilibrium problem has been stated mathematically in several forms: the conceptually simplest form is stated below. Let

v_a = number of vehicles per unit time on link a of the network;
$S_a(v_a)$ = generalized travel time on link a, which increases with flow v (a typical congestion function is $t_a[1 + 0.15(v_a/c_a)^4]$ where t_a is the travel time with zero flow, and c_a is a measure of the capacity per unit time of link a);
X_{ij}^r = number of vehicles of i to j on path r; and
δ_{ij}^{ar} = 1 if link a belongs to path r from i to j, 0 otherwise.

If the trip matrix (T_{ij}) is given, then the equilibrium assignment of trips to links may be found by solving the following nonlinear programming problem:

$$\min \sum_a \int_0^{v_a} S_a(x)dx \tag{1}$$

subject to

$$v_a = \sum_i \sum_j \sum_r \delta_{ij}^{ar} X_{ij}^r \tag{2}$$

$$\sum_r X_{ij}^r = T_{ij} \tag{3}$$

$$X_{ij}^r \geq 0 \tag{4}$$

For all links a in the network; i = 1, . . . N; j = 1, . . . N; and N = number of zones.

This is a nonlinear programming problem with a convex objective function subject to two sets of linear constraints and two sets of nonnegativity conditions. Constraint set Equation 2 states that the flow of vehicles v_a on link a is equal to the sum of the flows from all zones i to all zones j that use that link. Constraint set Equation 3 states that the number of vehicles from zone i to zone j over each path used must sum to the specified number of trips (T_{ij}). Constraint set Equation 4 ensures that no flow is negative.

Now consider the objective function (Equation 1). $S_a(x)$ is the link-congestion or capacity-restraint function for link a. The integral term is the area under the link-congestion function from zero flow to flow v_a. In Figure 1, S_a is the average travel time. The area under curve S_a has no (known) interpretation. Why, then, should we be interested in minimizing the sum of these areas over all links? The answer to this question is conceptually simple. The link flows for which this objective function achieves its minimum value are those that satisfy the equilibrium conditions stated by Wardrop.

This point can be readily grasped if we consider a highly simplified example (6). Let A and B be two links that connect node 1 to node 2, as shown in Figure 2a. A total of 8000 vehicles travel from node 1 to node 2.

To assign these vehicles to the two links, plot the congestion for link A, mark off the required flow (8000), and plot the second function in the reverse direction. The intersection of these two functions gives the equilibrium travel time of 63.3; the equilibrium flows are 2153 vehicles on link A and 5847 vehicles on link B.

This graphical solution may be stated mathematically as follows:

Figure 1. Congestion function for a given link.

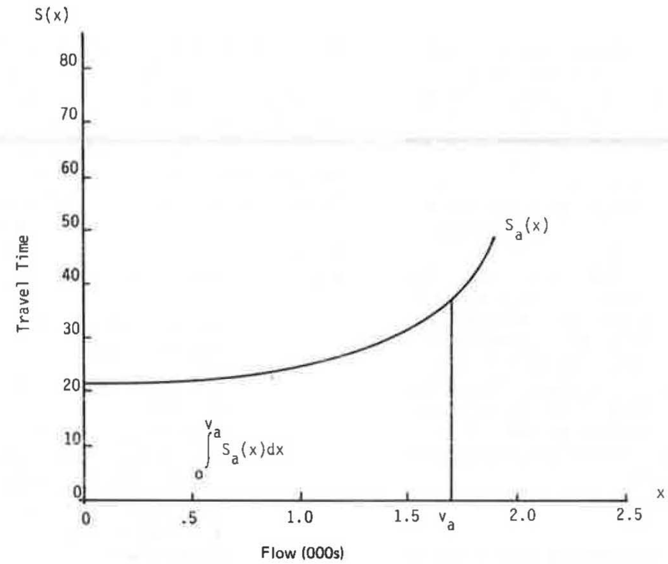
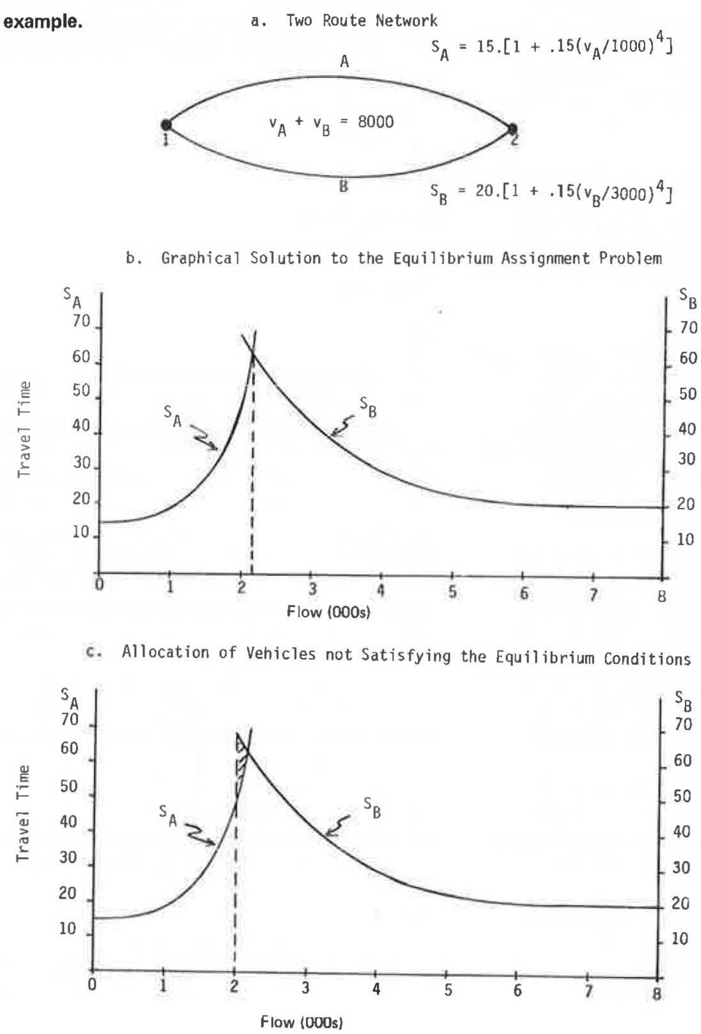


Figure 2. A two-link example.



$$\min \int_0^{v_A} S_A(x)dx + \int_0^{v_B} S_B(x)dx \qquad (5)$$

subject to

$$v_A + v_B = 8000 \qquad (6)$$

$$v_A, v_B \geq 0 \qquad (7)$$

Note that the area under the congestion functions in Figure 2b is equal to 220 674, which is the value of the objective function.

Now, consider any other solution than the one given by the intersection of S_A and S_B, say $v_A = 2000$ (see Figure 2c). The area under the two congestion functions for this solution is the same as in 2b plus the small triangular-shaped area that lies between 2000 and 2153, which has an area of 1326. Thus all solutions other than the equilibrium solution have a larger value of the objective function than does the equilibrium solution. Hence, the solution that minimizes the sum of the integrals of the congestion functions for all of the links is the equilibrium solution.

ALGORITHM FOR EQUILIBRIUM ASSIGNMENT

Next, consider how we solve the equilibrium-assignment problem for large networks. The equilibrium-assignment algorithm, which is commonly used, has a structure somewhat similar to the version of the iterative assignment in the Federal Highway Administration (FHWA) PLANPAC computer programs (7). To illustrate these similarities and differences each of three algorithms is outlined, and a simple three-link example is solved.

Equilibrium-Assignment Algorithm

Given (a) a network with congestion functions for each link, (b) a trip matrix to be assigned, and (c) a current solution for the link loadings (v_a), perform the following steps:

1. Compute the travel time on each link $S_a(v_a)$ that corresponds to the flow v_a in the current solution;
2. Trace minimum path trees from each origin to all destinations by using the travel times from step 1;
3. Assign all trips from each origin to each destination to the minimum path (all-or-nothing assignment); call this link loading (w_a);
4. Combine the current solution (v_a) and the new assignment (w_a) to obtain a new current solution (v'_a) by using a value λ selected so as to minimize the following objective function:

$$\sum_a \int_0^{v'_a} S_a(x)dx \qquad (8)$$

where $v'_a = (1 - \lambda)\, v_a + \lambda w_a$; and

5. If the solution has converged sufficiently, stop; otherwise return to step 1.

Initially, a current solution can be obtained by performing an all-or-nothing assignment based on free-flow times. This initial assignment is then used to compute revised travel times to perform another all-or-nothing assignment (steps 1-3). The two assignments are then combined by using a weight λ selected so as to give a new solution that minimizes the objective function of the nonlinear programming problem. This parameter can be readily determined by use of a one-dimensional search technique.

The change in the value of the objective function provides a measure of the convergence of the algorithm. As the change approaches zero, so does the value of the parameter λ. Thus, the equilibrium assignment is a weighted combination of a sequence of all-or-nothing assignments. The algorithm is not heuristic, that is, a method found to give good solutions. Rather, it is the Frank-Wolfe method for solving nonlinear programming problems applied to the equilibrium-assignment problem. LeBlanc (3) gives a rigorous derivation of the algorithm.

Now, consider a very simple example of the use of the algorithm. A three-link network is defined by adding link C to the network in Figure 2:

$$S_c = 21[1 + 0.15(v_c/1500)]^4 \qquad (9)$$

Even this simple problem cannot be solved graphically.

The results of applying the algorithm to this problem are given in Table 1. Five iterations are given after an initial solution. For each iteration, the all-or-nothing assignment is given on the first line followed by the new solution on the second line. The travel times given for each link are the values of the congestion functions for the link flows shown. The values of the objective function and λ were given on the right-hand side of the table.

The initial solution assigns all 8000 vehicles to link A. In the first iteration, all vehicles are assigned to link B, which results in the same combined solution shown in Figure 2. Next, all vehicles are assigned to link C, which results in the first good approximation of the equilibrium solution and has an objective function of 174 807. Iterations 3-5 refine this solution by making small adjustments on the order of 1 percent or less. One could effectively stop the algorithm after iteration 3 since a very small decrease in the objective function and a small value for λ were found. Iterations 4 and 5 are given only to indicate how the algorithm continues to converge.

Table 1. Equilibrium assignment.

		Link A		Link B		Link C		Equilibrium	
Iteration	Step	Flow	Time	Flow	Time	Flow	Time	Objective Function	λ
Initial solution		8000	9231.0	0	20.0	0	21.0	14 864 600	
1	3	0		8000		0			
	4	2153	63.3	5847	63.3	0	21.0	220 674	0.731
2	3	0		0		8000			
	4	1598	29.7	4341	33.2	2060	32.2	174 807	0.258
3	3	8000		0		0			
	4	1666	32.3	4296	32.6	2039	31.8	174 697	0.011
4	3	0		0		8000			
	4	1659	32.0	4277	32.4	2065	32.3	174 687	0.004
5	3	8000		0		0			
	4	1666	32.3	4273	32.3	2062	32.2	174 686	0.001

Iterative Assignment

As a further basis for understanding the equilibrium-assignment algorithm, the FHWA version of iterative assignment is now sketched (7, pp. 189-193). The algorithm requires the same input information as does equilibrium assignment. To execute the algorithm, perform four iterations of the following sequence and compute the mean of the four all-or-nothing assignments.

1. Compute the travel time on each link $S_a(v_a)$ corresponding to the flow v_a in the current solution;
2. Compute a weighted mean travel time (S_a''), which consists of the current travel time $[S_a(v_a)]$ and the travel time (S_a') from the previous iteration:

$$S_a'' = 0.75 S_a' + 0.25 S_a(v_a) \tag{10}$$

3. Trace minimum path trees from each origin to all destinations by using the weighted travel times S_a'' from step 2;
4. Assign all trips from each origin to each destination to the minimum path (all-or-nothing assignment); call this link loading v_a'; and
5. Return to step 1 and replace v_a with v_a'.

The use of a weighted mean travel time is an attempt to prevent the method from oscillating widely in computing minimum paths. Note, however, that the link loadings are not averaged until the final step, although the link travel times reflect implicitly the all-or-nothing assignments at each iteration.

The same three-link example is solved by using this algorithm in Table 2. The new travel times are given in each iteration as a basis for determining the next assignment. Following four all-or-nothing assignments, the mean flow is computed. The objective function of the equilibrium-assignment problem is computed for each iteration and for the final solution. This function provides a useful measure for comparison of the equilibrium and iterative assignments. The final value of the objective function for the iterative assignment has a somewhat higher value than for the equilibrium assignment. Thus the iterative assignment is not as close to true equilibrium. This conclusion can also be drawn by comparing the travel times that correspond to the final link loading in Tables 1 and 2. At equilibrium, these travel times should be equal.

Another weakness of the iterative-assignment algorithm is that there is no reliable rule about how many iterations to perform or what weights to use in computing the mean travel times. Had one more iteration of the algorithm been performed (or one less), the result would have been much different. With the equilibrium procedure, the overall result always improves with each iteration; the number of iterations depends only on how much improvement is desired.

Incremental Assignment

Another heuristic assignment procedure that has been widely used is incremental assignment. There are two types of incremental loading of a network. In the first type each origin-destination flow is divided into n equal parts, typically four. Each part is assigned by using all-or-nothing assignment; the link-loading and travel times are updated following the assignment of each increment. Following the assignment of the nth part, the link loadings are summed to determine the final loading. An alternate method developed by the Chicago Area Transportation Study (CATS) is the tree-by-tree method. In this case each row of the trip table is assigned completely by all-or-nothing assignment; the travel times are updated following each assignment.

Table 3 gives the results of the first incremental method applied to the three-link example. Four increments are used. By coincidence the final result happens to be the same as that given by the iterative method. The objective function value applies only to the final solution in this case. As with iterative assignment, the number of increments is an important determinant of the quality of the solution. In this case, however, the quality tends to improve as the number of increments increases.

In all three methods, a similar number of all-or-nothing assignments are performed to obtain a solution. No conclusions should be drawn about the relative quality of the solutions among the three methods, since such a small example could be quite misleading. The purpose here is only to educate and to compare the actual calcu-

Table 2. FHWA iterative assignment.

Iteration	Step	Link A Flow	Link A Time	Link B Flow	Link B Time	Link C Flow	Link C Time	Equilibrium Objective Function
Initial solution		8000	9231.0	0	20.0	0	21.0	14 864 000
1	2		2319.0		20.0		21.0	
	4	0	15.0	8000	171.7	0	21.0	402 726
2	2		1743.0		57.9		21.0	
	4	0	15.0	0	20.0	8000	2570.0	4 245 796
3	2		1311.0		48.4		658.3	
	4	0	15.0	8000	171.7	0	21.0	402 726
Mean flows and corresponding travel times		2000	51.0	4000	29.5	2000	31.0	177 967

Table 3. Incremental assignment.

Increment	Step	Link A Flow	Link A Time	Link B Flow	Link B Time	Link C Flow	Link C Time	Equilibrium Objective Function
1	Assignment	2000		0		0		
	Sum, time	2000	51.0	0	20.0	0	21.0	
2	Assignment	0		2000		0		
	Sum, time	2000	51.0	2000	20.6	0	21.0	
3	Assignment	0		2000		0		
	Sum, time	2000	51.0	4000	29.5	0	21.0	
4	Assignment	0		0		2000		
	Sum, time	2000	51.0	4000	29.5	2000	31.0	177 967

lations performed in each case.

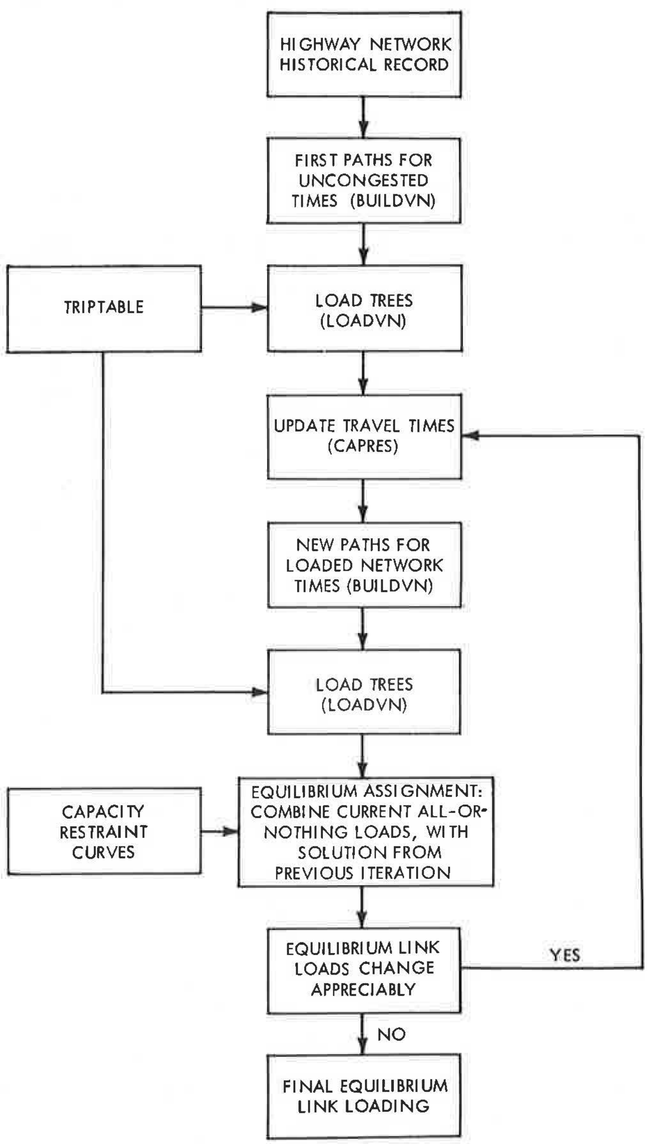
APPLICATION OF AN EQUILIBRIUM-ASSIGNMENT ALGORITHM

This section presents an application of equilibrium assignment to a large-scale trip table and network by CATS. The only other report of such an application was made by Florian and Nguyen (8) for a medium-sized network for Winnipeg. Applications have also been made by the Los Angeles Regional Transportation Study, but no results have been published.

The Equilibrium-Assignment Program

A program to perform equilibrium assignment was developed cooperatively by CATS and the University of Illinois, Urbana-Champaign. This program uses modules of the FHWA System-370 PLANPAC program battery, including programs for tree building, network loading, and network travel-time updating. The equilibrium-assignment program, which incorporates the existing PLANPAC programs, is illustrated in Figure 3. Analysts familiar with PLANPAC will recognize that the sequence of program steps shown in this figure differs only slightly from the usual application of the PLANPAC programs. The equilibrium-assignment program simply replaces the program VOLAVG, which is used to average loadings from separate assignments. But whereas the analyst must arbitrarily select how the two sets of link volumes are to be weighted in VOLAVG, the equilibrium-assignment routine internally determines the weighting of the link loadings that most nearly results in an equilibrium assignment.

Figure 3. Equilibrium assignment combined with PLANPAC programs.



Note: PLANPAC program names in parenthesis.

There is one feature of the PLANPAC programs that greatly simplifies the use of the equilibrium-assignment algorithm—this is the format of the highway network file. In the PLANPAC battery, highway network files are maintained in a binary file, called the network historical record. For each iteration the new link volume and recomputed link travel time are successively added at the end of a link record. Thus, all of the information needed for the calculation of a new equilibrium link volume (except λ) can be stored in one link record. New equilibrium link volumes can then be tagged at the end of the historical record (just like any other link volume) and passed directly into the CAPRES program to recompute link travel times.

The program to compute the equilibrium assignment has an uncomplicated linear structure. Logic of this program is as follows:

1. The capacity-restraint curves are read into memory;
2. The control card that identifies the location in the network historical record of the current solution and the current all-or-nothing assignment is read;
3. The network historical record is read, and the link capacities and both sets of link volumes and times are loaded into arrays;
4. A one-dimensional search procedure is executed to find the value of λ that minimizes the objective function computed from the current solution and the current all-or-nothing loading; and
5. The historical record is reread and a new historical record is written, which contains the new current solution.

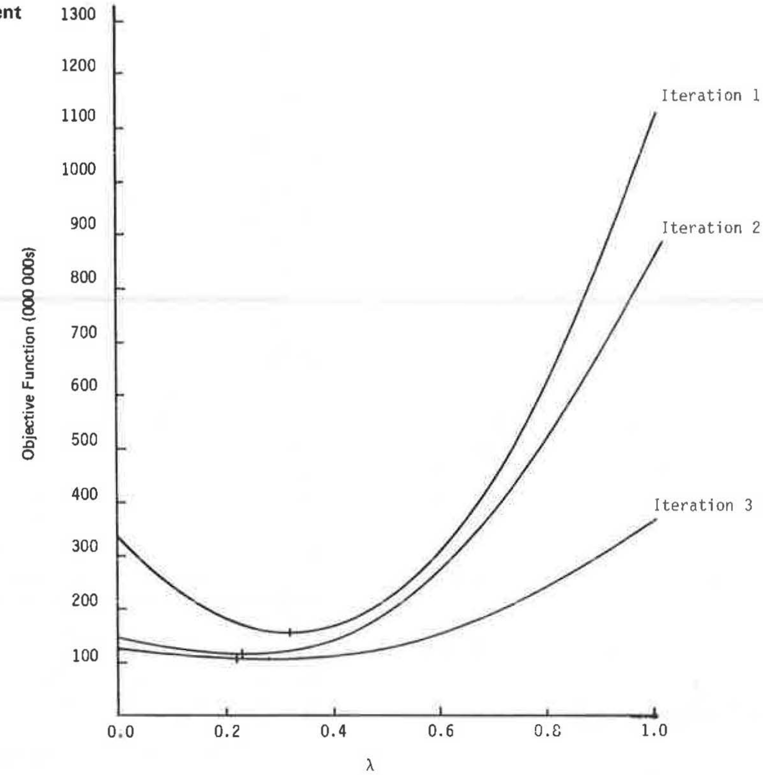
BASE DATA

Network and trip-table data for the application were obtained from a subarea transportation study for DuPage County, Illinois. This is a suburban county in northeastern Illinois, which covers an area of approximately 900 km^2 (350 miles^2) directly west of Cook County and the city of Chicago. The eastern half of the county is quite developed and has several major retail and employment centers. Current county population is about 500 000 persons; county employment is about 250 000 jobs. A wide range of traffic conditions can be observed in the county, including congestion and delay on many arterials.

Although the DuPage County network is for a subarea study, the assignment network is still quite large. There are nearly 29 000 one-way links and 9400 nodes in the 1975 network. Approximately one-third of the network is contained within the primary study area of DuPage County and a 10-km (6-mile) wide collar around the county. The network in this area is detailed and includes all roads except minor local streets. Outside of the primary study area the network is more aggregated, but it still contains all major and minor arterials.

The zone system has 906 zones in DuPage County plus

Figure 4. The equilibrium assignment objective function versus λ.



an additional 93 zones for the remainder of the northeastern Illinois region.

Definition of Capacity-Restraint Functions

Three different sets of capacity-restraint functions were used to determine their effect on the algorithm's performance: (a) CATS original capacity-restraint curves, (b) the standard FHWA capacity-restraint curves, and (c) a revised set of FHWA capacity curves. Instead of using the actual functions, the curves are entered into the program as a set of data points. The function is then approximated by chords connecting these points.

The CATS capacity-restraint curve used in this application is

$$S = t_0(2^{v/c} + 1)/2 \tag{11}$$

The standard FHWA capacity function is

$$S = t_0[1 + 0.15(v/c)^4] \tag{12}$$

where t_0 = free-flow link travel time and v/c = link flow to capacity ratio.

Algorithm Convergence Toward Equilibrium

One of the first questions raised in dealing with the equilibrium-assignment algorithm is, How quickly does the assignment converge to equilibrium? Figure 4 shows how the objective function varies with different values of λ through three iterations by using the CATS' capacity-restraint function. In the first iteration, the objective function is strongly concave and has a minimum at λ = 0.34. The objective functions for the next two iterations flatten out considerably; by the third iteration the optimal value of the objective function differs from the objective function at λ equal to zero by less than 10 percent. Nearly identical results were obtained by use of the standard FHWA capacity-restraint function.

This experience from the DuPage study suggests that, for all practical purposes, equilibrium is reached after four iterations of the equilibrium-assignment algorithm. This corresponds to the building and loading of five sets of minimum-time-path trees since one additional all-or-nothing assignment is needed to find an initial solution. The building of the minimum-time paths is the most expensive operation in each iteration. The one-dimensional search does not significantly increase the computation time as compared with an FHWA iterative assignment. Execution of the BUILDVN program for the DuPage network requires 10 min of central processing unit (CPU) time on an IBM 370/168 computer.

Further documentation of how λ converges is tabulated in Table 4, which lists the values of the objective function for separate runs of four iterations on each of the two capacity-restraint functions. Although the value of the objective function is much different for the two equilibrium-assignment runs, performance of the algorithm is not significantly altered. By the fourth iteration, λ values of less than 0.10 are attained in both examples. Therefore, four iterations would appear to be sufficient for large networks over a reasonable range of capacity-restraint functions.

COMPARISON OF EQUILIBRIUM AND FHWA ITERATIVE ASSIGNMENTS

The equilibrium-assignment objective function was computed for a conventional FHWA iterative assignment to determine how well this heuristic approximates equilibrium link loadings. The results for the iterative assignment are shown in the right-hand column of Table 4. These calculations were made by using the standard FHWA capacity-restraint functions and are directly comparable with the adjacent column. The objective

Table 4. λs and objective functions for two sample runs.

	Equilibrium Assignment				FHWA Iterative Assignment
	CATS Capacity Restraint		FHWA Capacity Restraint		FHWA Capacity Restraint
Iteration	λ	Objective Value	λ	Objective Value	Objective Value
1	0.34	151×10^9	0.34	227×10^9	481×10^9
2	0.23	120×10^9	0.21	177×10^9	917×10^9
3	0.22	112×10^9	0.20	156×10^9	728×10^9
4	0.25	112×10^9	0.07	154×10^8	790×10^9
					255×10^9[a]

[a]Objective value for assignment formed by averaging iterations 3 and 4.

Table 5. FHWA iterative and equilibrium-assignment results for DuPage study (CATS capacity-restraint function).

		FHWA Iterative		Equilibrium	
Counted Volume Group Range	Average Count	Average Volume	RMS Error (%)	Average Volume	RMS Error (%)
0-500	243	1 702	1 012.3	1 750	960.0
500-1 000	706	1 495	226.3	1 556	226.9
1 000-2 000	1 456	2 099	93.2	2 086	89.7
2 000-3 000	2 462	2 535	54.7	2 665	55.0
3 000-5 000	3 971	4 045	47.4	4 014	45.6
5 000-10 000	7 002	6 808	37.1	6 831	41.1
10 000-15 000	12 057	11 623	32.4	11 632	27.2
15 000-20 000	16 780	16 735	27.9	16 270	26.0
20 000-25 000	21 714	19 815	21.5	18 028	23.9
30 000-40 000	36 300	35 644	7.5	27 446	26.7
Entire volume range	6 352	6 383	39.8	6 223	43.2

Table 6. Average assigned volumes by using different capacity-restraint functions after four iterations of the equilibrium-assignment algorithm.

		CATS Capacity Curve		FHWA Capacity Curve		Adjusted FHWA Capacity Curve	
Counted Volume Group Range	Average Count	Average Volume	RMS Error (%)	Average Volume	RMS Error (%)	Average Volume	RMS Error (%)
0-500	243	1 750	960.0	1 688	978.1	1 683	979.8
500-1 000	706	1 556	226.9	1 601	244.9	1 621	247.5
1 000-2 000	1 456	2 086	89.7	2 187	110.0	2 161	105.6
2 000-3 000	2 462	2 665	55.0	2 916	67.7	2 862	65.8
3 000-5 000	3 971	4 014	45.6	4 331	52.9	4 259	50.7
5 000-10 000	7 002	6 831	41.1	7 067	42.6	7 038	42.2
10 000-15 000	12 057	11 632	27.2	11 393	29.4	11 472	28.5
15 000-20 000	16 780	16 270	26.0	14 581	23.0	14 939	23.1
20 000-25 000	21 714	18 028	23.9	17 104	29.8	17 276	28.2
30 000-40 000	36 300	27 446	26.7	28 485	24.0	28 652	23.6
Entire volume range	6 352	6 223	43.2	6 293	45.1	6 291	44.3

function for the mean of the third and fourth FHWA iterations is almost 50 percent greater than the objective function for the equilibrium-algorithm loadings after four iterations. Clearly, the conventional iterative approach produces a rather poor approximation of equilibrium.

The comparison of equilibrium and FHWA iterative assignment was further investigated by comparing the results of two assignments by using CATS' capacity-restraint functions. The link flows given in Table 5 are those produced by the fourth iteration of the algorithm. Included in this table are approximately 600 links in DuPage County for which traffic counts were available. The data listed in the table come from the output of the PLANPAC program CAPRES. Each flow entry is the average flow assigned on all links in the link's class, and the root mean square (RMS) error column lists the RMS error as a percentage of the average count for the class. For a selected set of links with traffic counts within DuPage County, the two assignments showed significant differences. The equilibrium-assignment flows are generally less than the FHWA assignment flows on higher-flow links. Whether this is a general bias between the two techniques is impossible to tell at this point; the results of Table 5 may just point up the limitations in the capacity-restraint functions.

IMPACT OF DIFFERENT CAPACITY-RESTRAINT FUNCTIONS

In order to examine the above point one step further, different functions were tested to determine how they affected the results of the equilibrium-assignment algorithm. CATS capacity-restraint functions were used in the algorithm first, then the FHWA set of curves was used, and finally an adjusted set of FHWA curves was inserted in the algorithm. The adjusted curves were tested because of an apparent underassignment of high-volume links and overassignment of low-volume links by the algorithm when the FHWA curves were used. The adjusted FHWA capacity curves were set so that the capacity of a high-capacity link is effectively increased by 10 percent and the capacity of a low-capacity link is decreased by 10 percent.

Table 6 provides some results from these three equilibrium-assignment runs, which incorporate different capacity-restraint functions. There are no substantial differences between any of the assignments. The use of CATS original capacity-restraint function provides an assignment slightly closer to actual counts, but the results are not significantly better than the remaining two assignments. All three assignments tend to overpredict traffic on low-volume links, partially because the local street network over which the beginning and ending segments of trips travel is incomplete. Comparison of the second and third assignments shows that the effect of the adjustment to the FHWA curves is almost negligible.

The changes that do occur, however, are in the desired direction, which indicates that some control over the assignment can be exerted through capacity-restraint functions. Since the equilibrium-assignment algorithm produces a convergent series of assignments, it should be possible to calibrate these functions according to route type or location in an urban area.

CONCLUSIONS

Although our experience with applications of equilibrium assignment to large-scale, congested networks is still limited, we believe that the results reported in this paper provide convincing evidence that equilibrium assignment should always be preferred to FHWA iterative assignment for congested networks. We reach this conclusion for three reasons:

1. Equilibrium assignment provides a better assignment in terms of the overall objective of equal travel times over all paths used between each origin and destination pair,
2. The computational effort is similar and may be less in some cases in which the equilibrium algorithm converges quickly, and
3. Equilibrium assignment can be readily incorporated into FHWA's PLANPAC battery; moreover, it is already available in UTPS.

The preliminary results we have presented concerning the ability of equilibrium assignment to reproduce observed 24-h flows are not as convincing. There are two reasons for this result. First, the capacity-restraint functions are probably too crude. This problem has been explored slightly here, but more study and experimentation are needed. Second, the use of equilibrium assignment to produce 24-h assignments may be inappropriate in that only the peak periods have truly congested flow. All-or-nothing assignment may be sufficient for off-peak periods. Additional study of this question is needed to determine the actual cause of these apparent differences between ground counts and assigned flows.

ACKNOWLEDGMENT

The basic research on equilibrium assignment, on which this paper is based, has been conducted by many individuals during the past 10 years. Since it was not our purpose to review the development of equilibrium-assignment methods, we have not referred to this literature, except in the few cases in which it was directly pertinent. We are grateful for the advice and encouragement that we received from David Gendell of the Federal Highway Administration and Thomas Hillegass of the Urban Mass Transportation Administration.

REFERENCES

1. S. Nguyen. An Algorithm for the Traffic Assignment Problem. Transportation Research, Vol. 4, 1974, pp. 391-394.
2. L. J. LeBlanc, E. K. Morlok, and W. P. Pierskalla. An Efficient Approach to Solving the Road Network Equilibrium Traffic Assignment Problem. Transportation Research, Vol. 9, 1975, pp. 309-318.
3. M. Florian, ed. Traffic Equilibrium Methods. Springer-Verlag, New York, 1976.
4. UTPS Reference Manual. Urban Mass Transportation Administration; Federal Highway Administration, July 7, 1977.
5. J. G. Wardrop. Some Theoretical Aspects of Road Traffic Research. Proc., Institution of Civil Engineers, Part 2, Vol. 1, 1952, pp. 325-378.
6. E. K. Morlok. Introduction to Transportation Engineering and Planning. McGraw-Hill, New York, 1978.
7. Computer Programs for Urban Transportation Planning: PLANPAC/BACKPAC General Information Manual. Federal Highway Administration, April 1977. GPO Stock No. 050-001-00125-0.
8. M. Florian and S. Nguyen. An Application and Validation of Equilibrium Trip Assignment Methods. Transportation Science, Vol. 10, 1976, pp. 374-390.

Publication of this paper sponsored by Committee on Passenger Travel Demand Forecasting.

Equilibration Properties of Logit Models

Alex Anas, Department of Civil Engineering, Northwestern University, Evanston, Illinois

Despite the importance of supply-demand equilibration in travel-demand forecasting and urban planning, no attention has been paid to the equilibration properties of logit models of travel demand and residential mobility. The preponderance of logit models in travel demand and related fields suggests that these properties are worth examining if these models are to become useful forecasting tools. This paper demonstrates the basic price equilibration properties of logit models for simplified versions of six typical problems encountered in travel-demand and residential-location forecasting. Measures of the differential price of any two alternatives are derived in closed form and shown to reflect the well-known logit property of the independence from irrelevant alternatives as long as the population of travelers and households is one homogeneous group. It is shown that

this property is lost when the population consists of several segments that have distinct preferences. In such cases closed-form solutions are not possible and numerical procedures are necessary.

Many problems in transportation systems and urban planning require an equilibrium relation between demand and supply in order to measure or evaluate system performance. The crucial steps for the planner or system analyst are (a) the estimation of demand functions, (b) the estimation of supply functions, and (c) the performance of a consistent forecast for a future state by equilibrating demand and supply.

In recent years economists, transportation planners, and systems analysts have contributed to the development and empirical estimation of a class of demand functions based on the logit and related models of discrete choice. Logit models have been applied widely in travel-demand and modal-choice analysis and to a lesser extent in the related areas of housing-market and residential-location analysis. The best-known works on the subject are those of McFadden (1) and Domencich and McFadden (2). Despite the preponderance of logit models as tools of demand analysis, no attention has been paid to the equilibration properties of these models. This issue finds brief mention in the recent book by Domencich and McFadden (2). As they put it:

> If the travel-demand function is structured so that all of the decisions incorporated within it are allowed to be responsive to the performance of the transportation system, then provisions must be made to equilibrate demand and the performance of the transport system to estimate properly the effects of changes in the transportation system on trip interchanges. It is not the purpose of this study to analyze or develop equilibration procedures, but the implications of a policy-sensitive demand model on other modeling requirements should be noted.
>
> . . . failure to equilibrate demand and system performance properly could result in substantial error in estimating the expected impact of a facility change on travel volumes and service levels.

In many instances it is realistic to assume that supply or capacity will be inelastic, at least in the short run. For such cases, an equilibration problem determines price adjustments that clear the market by matching demand and supply for each alternative in the market. From the practical point of view, the importance of price adjustments in forecasting may be demonstrated by the following scenario. Suppose that a logit model of residential location has been estimated for a city by using data from 1975. It is now desired to use this model to forecast residential-location patterns for 1980 under the assumption that transportation services to subarea A of the city will be much improved between 1975 and 1980. In the meantime, let us assume that the housing stock in the same area remains approximately constant due to such factors as zoning, unavailability of vacant land, and high costs of redevelopment. The 1980 travel improvements will strengthen the demand for subarea A. If the forecasting procedure assumes that housing prices will remain unchanged between 1975 and 1980, the demand for housing in subarea A may well exceed the supply of housing units there, assuming other subareas receive comparatively minor travel improvements. In other areas demand may be found to be below the supply. To correct this mismatch, housing prices should increase in those zones where demand exceeds supply and should decrease in those zones where supply exceeds demand. The housing market is equilibrated when a new set of housing prices is found such that demand is less than or equal to supply in every zone.

TRANSPORTATION ANALYSIS AND URBAN PLANNING PROBLEMS

In this paper several fundamental equilibration properties of logit models are demonstrated within the context of six specific problems that are typical in transportation analysis and urban planning. Problems A through E are united by the assumption that there is one homogeneous population of commuters or households, and this assumption enables closed-form solutions. Several properties of the logit-demand structure are reflected in these solutions:

1. Price differentials (or the relative prices of alternatives) are unique, although the level of prices is nonunique up to the arbitrary specification of any one price;
2. The well-known logit property of the independence from irrelevant alternatives (IIA) implies that the relative prices of two alternatives (or locations) are determined independently of the information for all other alternatives (or locations); and
3. Price adjustments tend to absorb advantages that result from travel improvements so that they are reflections of these.

Later we relax the assumption of a homogeneous population and introduce several population segments that have different utility functions and choice behavior. It is shown that the IIA property no longer applies to the relative prices of two locations. Closed-form solutions are not possible, but I have developed and tested a numerical solution method (3).

Problem A: Parking Fees and Bus Fares

Suppose that a city's downtown receives commuters from a suburb through two travel modes. One is automobile, which requires parking in public lots operated by the city. The other alternative is to take the bus, which is also operated by the city. Each commuter pays a parking fee or a bus fare. The city operates a rush-hour bus capacity of S_B seats and maintains exactly S_A parking spaces. It receives N suburban commuters daily and we assume that there is no carpooling; that is, each automobile commuter drives alone. We also assume that $S_A + S_B = N$. What should be the parking fare and what should be the price of a two-way bus trip, assuming that both modes operate without congestion?

Suppose that each commuter decides whether to be a bus rider or a driver in such a way that aggregate demand is logistic and given by

$$f_i = \exp(\alpha P_i + K_i) \Big/ \sum_{j=1}^{2} \exp(\alpha P_j + K_j) \qquad i = 1,2;\ \alpha < 0 \tag{1}$$

where

f_1 and f_2 = the proportion of commuters that take automobile and bus, respectively,
P_1 and P_2 = the parking fee and two-way bus fare, and
$K_i \equiv \sum_{n=1} \beta_n Q_{in}$ = an abbreviation for the remaining utility terms.

P_1 and P_2 are the unknowns to be determined by the city, which seeks

$$Nf_1(P_1, P_2) = S_A \tag{2a}$$

and

$$Nf_2(P_1,P_2) = S_B \tag{2b}$$

which, by using Equation (1), can be rewritten as

$$N/\{1 + \exp[\alpha(P_2 - P_1) + (K_2 - K_1)]\} = S_A \tag{3a}$$

and

$$N/\{1 + \exp[\alpha(P_1 - P_2) + (K_1 - K_2)]\} = S_B \tag{3b}$$

By rearranging either Equation 3a or 3b we get,

$$P_1 - P_2 = [(1/\alpha)(K_2 - K_1)] - [(1/\alpha)\ln(S_B/S_A)] \tag{4}$$

The right-hand side of Equation 4 is the amount by which the prices should differ so that the number of drivers exactly matches the number of parking spaces and the number of riders exactly matches the number of bus seats. From Equation 4 we note several properties. First, the equilibrium prices are nonunique: any two prices that have the same difference (P_1-P_2) will do. Second, take the case where $S_B = S_A$. In this case we have $P_1 - P_2 = (1/\alpha)(K_2 - K_1)$, from which we know that if $K_2 > K_1$ then $P_2 > P_1$—the less attractive mode is priced lower. Third, suppose that more buses are added and an equal number of parking spaces is closed. From Equation 4 this would require increasing P_1 (the price of parking, which is now scarcer) or decreasing P_2 (the price of a bus trip, which is more available). Next, suppose that a third mode (train) is introduced with the number of seats (S_T) such that $S_A + S_B + S_T = N$, with P_3, the two-way train fare, and K_3, the remaining utility. Then, the above derivation can be repeated to derive Equation 4, but also

$$P_1 - P_3 = (1/\alpha)(K_3 - K_1) - (1/\alpha)\ln(S_T/S_A) \tag{4a}$$

and

$$P_2 - P_3 = (1/\alpha)(K_3 - K_2) - (1/\alpha)\ln(S_T/S_B) \tag{4b}$$

Note that Equations 4 and 4a will satisfy

$$N/\{1 + \exp[\alpha(P_2 - P_1) + (K_2 - K_1)] + \exp[\alpha(P_3 - P_1) + (K_3 - K_1)]\} = S_A \tag{5}$$

Equations 4a and 4b are of the same form as Equation 4, and any one of these is a direct reflection of the property of IIA—the price difference for any two modes is independent of any other mode. Given an arbitrary price for any one mode, Equations 4, 4a, and 4b can be used to make a unique determination of all the other prices. But how should this price level be determined? It seems reasonable to assume that the city should set these prices so as to cover the cost of operating the modes net of any subsidies from other sources (assumed to be zero here). Let the total costs be given by $C = C(S_A, S_B, S_T)$. Total daily revenues are $R = P_1S_A + P_2S_B + P_3S_T$. Setting $R = C$ we can substitute from Equations 4, 4a, and 4b for any two of the prices and solve for the third, thus determining the break-even price level.

Problem B: Supply of Buses and Parking Spaces

In problem A we assumed that the supply of parking spaces and bus seats is fixed. In this problem we allow the public authority to determine jointly both the price levels (P_1 and P_2) and also the market size of each mode (S_A and S_B) such that $S_A + S_B = N$. This problem may be posed as follows: The city contracts with a bus company that supplies buses and another firm that supplies parking space. Each of these firms operates under regular, upward sloping supply functions such that $S_A = F_1(P_1)$ and $S_B = F_2(P_2)$. The public authority must determine the regulated prices (P_1 and P_2) under which the two firms should operate. By using the similarity with problem A we know that P_1 and P_2 should satisfy

$$F_1(P_1) + F_2(P_2) - N = 0 \tag{6}$$

and

$$P_1 - P_2 = (1/\alpha)(K_2 - K_1) - (1/\alpha)\ln[F_2(P_2)/F_1(P_1)] \tag{7}$$

where Equation 7 is a restatement of Equation 4 and assures that $Nf_i(P_1, P_2) = F_i(P_i)$ for each i. If N is fixed, P_1 and P_2 can be found from Equations 6 and 7. Alternatively, if N is considered flexible another relationship is needed to replace Equation 6. This may be

$$(P_1 - c_1)F_1(P_1) + (P_2 - c_2)F_2(P_2) = 0 \tag{8}$$

where c_1 and c_2 are the costs of supplying a marginal capacity. Equation 8 states that both operations taken jointly break even. This may happen in two ways. Either $P_1 = c_1$ and $P_2 = c_2$ or $P_1 > c_1$ and $P_2 < c_2$ (or equivalently $P_1 < c_1$ and $P_2 > c_2$), but $(P_1 - c_1)F_1(P_1) = -(P_2 - c_2)F_2(P_2)$. This means that mode 1 produces a surplus of $\tau_1 = (P_1 - c_1)F_1(P_1)$ and mode 2 needs a subsidy of $\sigma_2 = (c_2 - P_2)F_2(P_2)$. Equation 8 assures that $\tau_1 = \sigma_2$ and thus both modes are kept in operation, by taxing mode 1 and by subsidizing mode 2.

Problem C: Demand for Housing and Location Rents

Logit models estimated by Quigley (4), Lerman (5), and Anas (6) are intended to capture the demand for residential location or type of housing. Typically, this problem may be stated as follows: Suppose that there are i = 1...I distinct zones, each of which contains S_i identical housing units. Then, the demand for zone (location) i can be expressed by the following logit model with grouped alternatives,

$$f_i = S_i\exp(U_i)\Big/\sum_j S_j\exp(U_j) \qquad i = 1 \ldots I; \quad \sum_i f_i = 1 \tag{9}$$

If we also assume that each household rents one housing unit and that the number of housing units in the rental market is equal to the number of households (N) then $N = \sum_j S_j$. This means that each housing unit will be occupied. In the short run the supplies (S_i) are assumed fixed for each i. Thus

$$Nf_i = S_i \text{ for each } i = 1 \ldots I \tag{10}$$

From Equations 9 and 10 we can write

$$Nf_i/Nf_j = S_i\exp(U_i)/S_j\exp(U_j) = S_i/S_j \tag{11}$$

We can now examine the implication of Equation 11 for rent adjustments if we first specify the utility function. Suppose it is given as

$$U_i = \alpha R_i + \beta T_i + K_i \qquad \alpha,\beta < 0 \tag{12}$$

where K_i is an abbreviation of terms such that $K_i \equiv \sum_{n=1} \gamma_n Q_{in}$ with Q_{in} a measure of the n^{th} characteristic of zone i and γ_n the corresponding utility parameter. R_i is the rent (price) of a housing unit in zone i and T_i is the generalized travel cost associated with zone i.

Equation 11 will hold only if

$$U_i = U_j \tag{13}$$

From this we derive

$$R_i - R_j = (\beta/\alpha)(T_j - T_i) + (1/\alpha)(K_j - K_i) \tag{14}$$

This result is analogous to our previous result in problem A. Suppose that the two zones are identical in all characteristics except transportation costs, then $K_i = K_j$ and the rent differential reflects the transport cost differential. The nonuniqueness and other considerations noted in problem A apply to Equation 14 as well.

Several variants of Equation 14 are worth noting. Suppose that the utility function was specified as follows, where Y represents household income,

$$U_i = \ell n\{R_i^\alpha T_i^\beta K_i\} \quad \alpha,\beta < 0, K_i \equiv \prod_{n=1} Q_{in}^{\gamma_n} \tag{15}$$

or

$$U_i = \alpha[Y - R_i - T_i] + K_i \quad \alpha > 0, K_i \equiv \sum_{n=1} \gamma_n Q_{in} \tag{16}$$

or

$$U_i = \ell n\{[Y - R_i - T_i]^\alpha K_i\} \quad \alpha > 0, K_i \equiv \prod_{n=1} Q_{in}^{\gamma_n} \tag{17}$$

By using Equation 13, Equations 15-17 will lead to the following,

$$R_i/R_j = (K_j T_j^\beta / K_i T_i^\beta)^{1/\alpha} \text{ for Equation 15} \tag{18}$$

$$R_i - R_j = (1/\alpha)(K_j - K_i) + (T_j - T_i) \text{ for Equation 16} \tag{19}$$

$$R_i - R_j (K_j/K_i)^{1/\alpha} = (Y - T_i) - (Y - T_j)(K_j/K_i)^{1/\alpha} \text{ for Equation 17} \tag{20}$$

The nonuniqueness argument applies to these as well. The IIA property of logit comes through in every case as the relative rents do not depend on any zone other than the two we are concerned with. Let $K_i = K_j$, then Equations 14 and 18-20 reduce to the following,

$$R_i - R_j = (\beta/\alpha)(T_j - T_i) \tag{21}$$

$$R_i/R_j = (T_j/T_i)^{\beta/\alpha} \tag{22}$$

$$R_i - R_j = T_j - T_i \text{ for both Equations 19 and 20} \tag{23}$$

The last of these is reminiscent of the early location rent model developed by Wingo (7) who assumed, rather arbitrarily, that rent plus transportation costs add up to the same constant at every location, namely $R_i + T_i$ = constant for every i.

In Equations 14 and 18-20, if the rent of any one zone is arbitrarily fixed, then the location rents of all other zones are uniquely determined.

Problem D: Impact of a New Travel Mode on Differential Location Rents

Assume that two locations i and j are identical in all respects and each is served by the same travel mode—automobile. Let R_i represent location rent for zone i, as before, and also let T_{i1} and T_{j1} represent travel costs by automobile to zone i and zone j. If we assume that demand is given by a logit model of joint location and mode choice

$$f_{im} = S_i \exp(U_{im}) / \sum_j \sum_k S_j \exp(U_{jk}) \qquad \sum_i \sum_m f_{im} = 1 \tag{24}$$

where U_{im} is the utility of choosing zone i and mode m for commuting to zone i. Now suppose that a new travel mode, transit, is introduced but serves only zone j and has travel cost $T_{j2} \neq T_{j1}$. Let the utility function be $U_{im} = \alpha R_i + \beta T_{im}$; then with condition $\sum_i S_i = N$ we can repeat our previous derivations in slightly different form, namely

$$Nf_{i1}/(Nf_{j1} + Nf_{j2}) = S_i \exp(U_{i1})/[S_j \exp(U_{j1}) + S_j \exp(U_{j2})] = S_i/S_j \tag{25}$$

By mutiplying Equation 25 by S_j/S_i we get

$$\exp(U_{i1}) = \exp(U_{j1}) + \exp(U_{j2}) \tag{26}$$

which implies

$$\exp(\alpha R_i)\exp(\beta T_{i1}) = \exp(\alpha R_j)[\exp(\beta T_{j1}) + \exp(\beta T_{j2})] \tag{27}$$

and

$$R_i - R_j = (1/\alpha)\ell n[\exp(\beta T_{j1}) + \exp(\beta T_{j2})] - (\beta/\alpha)T_{i1} \tag{28}$$

If we also assume that $T_{j1} = T_{i1}$, that is, that the automobile costs of the two zones are identical, then the differential rent $R_i - R_j$ is attributable purely to the impact of transit. Thus, if we let $T_{i1} = T_{j1} \equiv T_1$ we have

$$R_i - R_j = (1/\alpha)\ell n[\exp(\beta T_1) + \exp(\beta T_{j2})] - (\beta/\alpha)T_1 \tag{29}$$

Let us now take this one step further. Suppose that the introduction of transit does not create any real advantage. This would be the case if $T_{j2} = T_1$, which would reduce Equation 20 further to

$$R_i - R_j = (1/\alpha)\ell n 2 \tag{30}$$

Since $\alpha < 0$, Equation 30 implies that $R_j = R_i + |(1/\alpha)\ell n\, 2|$ where $|\cdot|$ measures the rent increase in zone j attributable to the presence of a new mode identical in transport cost to the existing mode. Note that, although the two zones are indistinguishable in terms of travel cost and all other characteristics, zone j still has a higher rent than does zone i. Intuitively, this seeming paradox is clarified as follows: Suppose that initially $R_i = R_j$ for these two zones. This would imply that $f_{j1} = f_{j2} = f_{i1}$ and thus $Nf_{j1} + Nf_{j2} = 2Nf_{i1}$. In other words, twice as many households choose zone j. Clearly then, to properly reallocate this demand and assure $Nf_{j1} + Nf_{j2} = Nf_{i1}$ rents in zone j must be higher.

We must also note that, if x new travel modes with equal transportation costs are introduced into zone j, then $R_j = R_i + |(1/\alpha)\, \ell n\, (x+1)|$.

Next, suppose that the utility function includes a mode-specific dummy variable so that $U_{i1} = \alpha R_i + \beta T_{i1}$ and $U_{j1} = \alpha R_j + \beta T_{j1}$ but $U_{j2} = \alpha R_j + \beta T_{j2} + \gamma_2$ where γ_2 measures the bias due to mode 2. From this we obtain the equivalent of Equation 28,

$$R_i - R_j = (1/\alpha)\ell n[\exp(\beta T_{j1}) + \exp(\beta T_{j2} + \gamma_2)] - (\beta/\alpha)T_{i1} \tag{31}$$

When $T_{j2} = T_{j1} = T_{i1} \equiv T_1$ we get

$$R_i - R_j = (1/\alpha)\ell n[1 + \exp(\gamma_2)] \tag{32}$$

Finally, if x new modes are introduced, each with equal transport costs, the equivalent of Equation 32 is

$$R_i - R_j = (1/\alpha)\ell n[1 + \sum_{n=2}^{x} \exp(\gamma_n)] \tag{33}$$

Problem E: Before-and-After Differential Rent Due to a Transportation Improvement

We now return to model Equation 9 of problem C. Let U_{ib} be the utility before a transportation improvement takes place and let U_{ia} be the utility after a transportation improvement. Assume that $U_{ib} = \alpha R_{ib} + \beta T_{ib}$ and $U_{ia} = \alpha R_{ia} + \beta T_{ia}$ with $T_{ia} < T_{ib}$, then what is the relationship between R_{ib} and R_{ia}? Note that Equation 9 can be written as

$$f_{ib} = 1\Big/\Big[1 + \sum_{j \neq i} (S_j/S_i)\exp(U_{jb} - U_{ib})\Big] \qquad i = 1 \ldots I \tag{34a}$$

and

$$f_{ia} = 1\Big/\Big[1 + \sum_{j \neq i} (S_j/S_i)\exp(U_{ja} - U_{ia})\Big] \qquad i = 1 \ldots I \tag{34b}$$

Again, assume that $\sum_j S_j = N$, what is $R_{ia} - R_{ib}$ in zone i, if this is the only zone affected by the transport improvement, that is, $T_{ia} < T_{ib}$ and $T_{ja} = T_{jb}$ for all $j \neq i$?

Note that

$$Nf_{ib}/Nf_{ia} = \Big[1 + \sum_{j \neq i} (S_j/S_i)\exp(U_j - U_{ia})\Big] \div \Big[1 + \sum_{j \neq i} (S_j/S_i)\exp(U_j - U_{ib})\Big] = 1 \tag{35}$$

where $U_j \equiv U_{ja} = U_{jb}$ for $j \neq i$. The above equality can be maintained only if $U_{ia} = U_{ib}$. By using the definition of utility this requires that

$$R_{ia} - R_{ib} = (\beta/\alpha)(T_{ib} - T_{ia}) \tag{36}$$

Thus, the rent increase must be such that the utility level before and after the investment remains the same, assuming that the utility level remains unchanged in all other zones not affected by the transportation improvement. For this to occur it is only necessary that the rent of any one zone unaffected by the investment remain unchanged before and after the investment. The above readily generalizes to the case of a transportation improvement that affects more than one zone—if utility remains unchanged before and after the improvement $U_{ia} = U_{ib}$ for each i, then the market is cleared before and after the improvement and Equation 36 holds for each zone i.

Next, consider the possibility that a new travel mode is introduced to every zone. In this case we are dealing with a model such as that of problem D (see Equation 23). The market will clear before and after the investment if

$$\exp(U_{i1b}) = \exp(U_{i1a}) + \exp(U_{i2a}) \tag{37}$$

where 1 denotes automobile and 2 the new mode, say transit. Assuming that automobile characteristics remain the same before and after the transit investment, the three utility functions are $U_{i1b} = \alpha R_{ib} + \beta T_{i1}$, $U_{i1a} = \alpha R_{ia} + \beta T_{i1}$, and $U_{i2a} = \alpha R_{ia} + \beta T_{i2}$. In this way, Equation 36 becomes

$$R_{ib} - R_{ia} = (1/\alpha)\ln\{1 + \exp[\beta(T_{i2} - T_{i1})]\} \tag{38}$$

Since $\alpha < 0$ this implies $R_{ia} > R_{ib}$. Note that as the transit improvement worsens the rent increase vanishes (recall $\beta < 0$):

$$\lim_{T_{i2} \to \infty} R_{ib} - R_{ia} = (1/\alpha)\ln\{1 + \exp[\beta(T_{i2} - T_{i1})]\} \tag{39}$$

Problem F: Traffic Congestion

A common equilibration problem of a different nature is that of capacity-constrained traffic flow, where the travel times or generalized costs on a network's links depend on the traffic-flow capacity of the link and the volume (number of passengers) that use the link. Unlike the destination- and housing-choice problems considered in this paper, traffic-flow equilibration is highly network sensitive, and problems can quickly become complicated beyond the reach of analytical solutions. Still, the basic nature of the problem can be illustrated for the simplest of all networks: two highway routes that connect an origin-destination pair used by a homogeneous population of drivers (N). In this case, let the proportion of drivers that use route i be logistic. Then

$$f_i = \exp(\alpha t_i + K_i)\Big/\sum_{j=1}^{2} \exp(\alpha t_j + K_j) \qquad i = 1,2 \tag{40}$$

where K_i is an abbreviation of the utility due to other (fixed) characteristics of the route i. Let the travel time (t_i) be given via a simple volume-delay function, namely,

$$t_i = t_{oi} + A_i(Nf_i/C_i)^g \qquad i = 1,2 \tag{41}$$

where

t_{oi} = the free-flow link travel time,
A_i = a link-specific parameter,
C_i = the link capacity,
Nf_i = the volume that uses link i, and
g = a parameter ($g > 0$).

By abbreviating $A_i N^g/C_i^g$ as b_i and substituting Equation 41 into Equation 40 we obtain

$$f_i = [\exp(\alpha t_{oi} + \alpha b_i f_i^g + K_i)]\Big/\Big[\sum_{j=1}^{2} \exp(\alpha t_{oj} + \alpha b_j f_j^g + K_j)\Big] \qquad i = 1,2 \tag{42}$$

Either one of these two equations can be written as

$$(f_1 - 1)\exp(\alpha t_{o1} + \alpha b_1 f_1^g + K_1) + f_1\exp(\alpha t_{o2} + \alpha b_2 f_2^g + K_2) = 0 \tag{43a}$$

or

$$\ln(f_2/f_1) = \alpha(t_{o2} - t_{o1}) + \alpha(b_2 f_2^g - b_1 f_1^g) + (K_2 - K_1) \tag{43b}$$

and should be solved for equilibrium-flow proportions f_1^*, f_2^* by using an iterative procedure.

EXCESS CAPACITY

Since the assumption that aggregate supply equals aggregate demand is somewhat unrealistic, we will examine the implications of relaxing it. In problems A and B this is achieved by assuming $S_A + S_B \geq N$ and in problems C, D, and E we must assume $\sum_i S_i \geq N$. Thus, some parking spaces or bus seats can remain unused or some dwellings can remain unoccupied. Since the residential location problem (C) is typical of the remaining problems, we will examine the implication of $\sum_i S_i \geq N$.

Suppose that we introduce a new set of nonnegative variables (v_1, v_2, . . . , v_I) that measure the number of vacant dwelling units in each zone. Then, we can write

$$\sum_i S_i - \sum_i v_i = N \tag{44}$$

The problem can now be restated as

$$Nf_i = S_i - v_i \quad i = 1 \ldots I \tag{45}$$

and more precisely as

$$NS_i \exp(\alpha R_i + \beta T_i + K_i) = (S_i - v_i) \sum_j S_j \exp(\alpha R_j + \beta T_j + K_j)$$
$$i = 1 \ldots I. \tag{46}$$

Equations 44 and 46 are I + 1 equations in the 2I unknowns, which are the rents and the vacancies. The sytem is underdetermined: Given the vacancy levels for all but any one of the zones, Equations 44 and 46 become I + 1 equations, with the rents and the remaining vacancy as the unknowns. If we fix vacancies as $\bar{v}_1, \ldots, \bar{v}_I$ so that these satisfy Equation 44 we can state

$$Nf_i/Nf_j = S_i \exp(U_i)/S_j \exp(U_j)$$
$$= (S_i - \bar{v}_i)/(S_j - \bar{v}_j) \tag{47}$$

From which we note that

$$\exp(U_i - U_j) = S_j(S_i - \bar{v}_i)/S_i(S_j - \bar{v}_j) \tag{48}$$

and

$$R_i - R_j = (1/\alpha)\ln[S_j(S_i - \bar{v}_i)/S_i(S_j - \bar{v}_j)] + (\beta/\alpha)(T_j - T_i)$$
$$+ (1/\alpha)(K_j - K_i) \tag{49}$$

which reduces to Equation 14 if $\bar{v}_i = \bar{v}_j = 0$.

Since a unique set of vacancies cannot be determined without specifying additional relationships, the effect of vacancies is to introduce a new source of nonuniqueness in the determination of market prices and to increase the uncertainty in the prediction of these prices. It has been shown in Anas (8) that one way that market prices can be determined is by specifying certain additional conditions of competitive-pricing behavior, such as profit maximization, and deriving an equilibrium set of market-clearing prices.

INTERACTION DUE TO SEVERAL CONSUMER TYPES

In each of the problems the entire population of consumers (travelers or households) were assumed to have the same utility function and choice behavior. This is a strong assumption and may not always be appropriate in practice. It is, therefore, fruitful to examine several consumer types, each with a different utility function and choice behavior. We do this for problem C. Suppose that the population of households is segmented into h = 1... H segments according to certain socioeconomic criteria and the work places of the household heads. Then, let the behavior of each segment be logistic according to

$$f_i^h = S_i \exp(U_i^h) \Big/ \sum_j S_j \exp(U_j^h) \qquad \sum_i f_i^h = 1, h = 1 \ldots H \tag{50}$$

with the utility function given as $U_i^h = \alpha_h R_i + \beta_h T_i^h + K_i^h$

where

R_i = the rent of location (zone) i,
T_i^h = the cost of commuting to zone i from the workplace of a type h household, and
K_i^h = the part of the utility function due to other characteristics of zone i.

Let N_h represent the number of households of type h and impose $\sum_h N_h = \sum_j S_j$. Now we must solve

$$\sum_h N_h f_i^h(R_1, R_2, \ldots, R_I) = S_i \quad i = 1 \ldots I \tag{51}$$

by finding $R_1, R_2, \ldots, R_I$. This is a system of I simultaneous, nonlinear equalities in I unknowns but cannot be solved in closed form. To see this, we may follow a procedure similar to that of problem C. Doing so for the case h = 1, 2

$$S_i/S_j = (N_1 f_i^1 + N_2 f_i^2)/(N_1 f_j^1 + N_2 f_j^2)$$
$$= \{[N_1 G_2 S_i \exp(U_i^1) + N_2 G_1 S_i \exp(U_i^2)]/G_1 G_2\}$$
$$\div \{[N_1 G_2 S_j \exp(U_j^1) + N_2 G_1 S_j \exp(U_j^2)]/G_1 G_2\} \tag{52}$$

where

$$G_h = \sum_i S_i \exp(U_i^h) \qquad h = 1, 2 \tag{53}$$

From Equation 52 we get,

$$N_1 G_2 [\exp(U_j^1) - \exp(U_i^1)] = N_2 G_1 [\exp(U_i^2) - \exp(U_j^2)] \tag{54}$$

Equation 54 shows that we cannot establish a simple relation for differential rent $(R_i - R_j)$. It is also seen that the IIA property no longer holds. The competition of the two household types for the housing supply in all zones establishes an interactive effect and the relative rents of i and j depend on characteristics of all the zones. A unique solution need not exist. It is true, in general, that many rent vectors will satisfy the simultaneous equations (Equation 51). Solutions can be obtained via special numerical techniques. One such application will be found in Anas (3), where problem E is solved for a 60-zone, five-household-segment spatial system for the case of a transit investment and excess capacity in housing.

CONCLUSIONS

The problems solved here and the more complex problems hinted at in the preceding part of the paper are a sample of a large number of supply and demand equilibration issues that form the basis of policy evaluation and planning analysis in transportation and related areas in urban planning. To date, most of the work dealing with logit models has confined itself to parameter estimation and crude forecasting. These forecasting exercises suffer from a serious weakness to the extent that the relevant equilibration issues are ignored, and thus the forecasts obtained are ultimately inconsistent. This paper has shown that these inconsistencies are readily rectifiable. Because of the complexity of problems that can be approached in this way, our objective has been to select simple, yet typical, problems of policy interest and to demonstrate the necessary manipulations and results for these problems. More complex problems can be solved by developing appropriate numerical simulation methods (3) or by specifying the nature of competitive pricing (8). My other work has shown that even for these problems, which involve several consumer groups and excess supply, the market-clearing distribution of prices is well behaved, even though it may not be possible to express it analytically.

ACKNOWLEDGMENT

This paper was supported by a grant from the National Science Foundation.

REFERENCES

1. D. McFadden. The Measurement of Urban Travel Demand. Journal of Public Economics, No. 3, 1973. pp. 303-328.
2. T.A. Domencich and D. McFadden. Urban Travel Demand: A Behavioral Analysis. North-Holland, Amsterdam, 1975.
3. A. Anas. The Impact of Transit Investment on Housing Values: A Simulation Experiment. Environment and Planning A, Vol. 11, 1979, pp. 239-255.
4. J.M. Quigley. Housing Demand in the Short Run: An Analysis of Polytomous Choice. In Explorations in Economic Research, Vol. 3, No. 1, 1976.
5. S. Lerman. A Behavioral Disaggregate Model of Household Mobility Decisions. Department of Civil Engineering, Massachusetts Institute of Technology, Cambridge, Ph.D. dissertation, 1975.
6. A. Anas. Empirical Calibration and Testing of a Simulation Model of Residential Location. Environment and Planning, Vol. 7, No. 8, 1975, pp. 899-920.
7. L. Wingo, Jr. Transportation and Urban Land. Resources for the Future, Inc., Washington, DC, 1961.
8. A. Anas. A Probabilistic Approach to the Structure of Rental Housing Markets. Journal of Urban Economics, 1978 (in press).

Publication of this paper sponsored by Committee on Traveler Behavior and Values.

Validation and Application of an Equilibrium-Based Two-Mode Urban Transportation Planning Method (EMME)

M. Florian, R. Chapleau, S. Nguyen, C. Achim, L. James-Lefebvre, S. Galarneau, J. Lefebvre, and C. Fisk, Centre de Recherche sur les Transports, University of Montreal

The purpose of this paper is to report on the validation and application of the two-mode urban transportation planning technique called EMME. This method may be characterized as an integrated two-mode traffic equilibrium method. Roughly speaking, this method combines a zonal aggregate-demand model with an equilibrium-type road assignment and a transit-assignment method. We describe the validation and application of the model by using data from the city of Winnipeg, Manitoba, Canada.

The purpose of this paper is to report on the validation and application of the two-mode urban transportation planning technique équilibre multimodal-multimodal equilibrium (EMME). This method may be characterized as an integrated two-mode traffic equilibrium method. It was suggested by Florian (1). Roughly speaking, this method combines a zonal aggregate-demand model (which may be a direct-demand model or an origin-destination table coupled with a suitable modal-split function) with an equilibrium-type road assignment and a transit-assignment method. The method has been described previously (2) and some of its theoretical properties have been studied by Fisk and Nguyen (3). The model was validated by using data from the city of Winnipeg, Manitoba, Canada. The equilibrium-type route-choice model for travel by private automobiles in congested urban areas was validated by Florian and Nguyen (4) in the Winnipeg road network. The transit-assignment model is essentially a shortest-route choice coupled with the diversion mechanism among sections served by common lines, which was devised by Chriqui and Robillard (5).

For the purpose of transportation planning, the city of Winnipeg is subdivided into 147 zones. The road network has 1040 nodes and 2836 one-way lines; observed link flows and link times were available for most of the links. The transit network has 56 lines, 1755 line segments, 500 egress-access links, and 800 nodes; 575 of the road network nodes are used in the coding of the transit network as well.

In the summer of 1976 the city of Winnipeg performed a speed-delay study, which consisted of measuring link volumes and link automobile travel times for 80-90 percent of the street system. In addition, bus travel times were measured for 446 transit line sections. These data served to recalibrate the volume-delay curves that were used in the road assignment and to calibrate the bus-automobile travel-time relationship required by EMME.

Since the city of Winnipeg had not previously used a transit-assignment model, the transit network was coded according to the EMME specifications, described by Achim and Chapleau (6), that permit the interface between the road and transit networks.

During the summer of 1976, the city of Winnipeg also performed an origin-destination survey of trips taken from home to work. A 17 percent sample of households was sampled and a separate survey of 23 percent of student trips was performed at about the same time. Since all of the analysis is done for the 7:30-8:30 a.m. peak hour, one of the first tasks considered was to define the departure codes, that is, the starting time of trips that will be using the road and transit networks during the peak hour. The departure codes were determined by the city of Winnipeg staff and were specified by origin, by using a subdivision of origins into 36 super zones. By using the departure codes, the corresponding trips are extracted from the survey data and multiplied by the appropriate expansion factors to obtain an estimate of the total person work trips taken in the peak hour by each mode. Then the total automobile work-trip matrix is scaled by appropriate automobile occupancy factors in order to obtain the total automobile work-trip matrix.

This matrix was then assigned to the road network and compared with the observed link volumes. Since only the work trips are sampled during the origin-destination survey, it was necessary to develop a set of adjustment factors that multiply the number of trips in the total automobile work-trip matrix in order to reflect automobile trips that are taken for purposes other than work and a certain amount of truck traffic. These factors are specified by origin to subdivide origins into 10 super zones. The determination of the most appropriate factor is a trial-and-error procedure. Where a factor is tried, the resulting assignments are compared to observed link flows and then a new factor is determined, which, it is hoped, is more appropriate. Five factors were tried until satisfactory results were obtained. In addition, trips to the University of Manitoba required special departure codes, which were specified for the subdivision of origins into 10 super zones, since this zone is relatively more distant from most origins. In the EMME computer system, the factors are converted into a vehicle-adjustment trip matrix, which is added to the total automobile work-trip matrix for the purpose of the assignment.

Once the departure codes, and hence the fixed origin-destination matrices, were determined, the modal-split function was calibrated. Due to the large size of the sample, it was possible to calibrate a zonal-aggregate logit modal-split function. We were then provided by the city of Winnipeg with a road-improvement scenario and a transit-improvement scenario. We first analyzed the base-year calibration by using the bimodal model and then proceeded to analyze the impact of the scenarios.

THE BUS-AUTOMOBILE TRAVEL-TIME FUNCTION

The purpose of this task was to develop a model that relates the travel time of a transit vehicle on a road link to the corresponding travel time for private automobiles. The model is used to take into account the change of transit travel times as a result of a change in the congestion level of a road link.

The data needed to develop this model are road link lengths, observed automobile travel times on those links, and the corresponding bus travel times. The road link lengths and automobile times were obtained from the road network data. The city of Winnipeg provided us with observed bus travel times for line sections (a line section is defined as the sequence of the corresponding road links). (The model was designed for U.S. customary units only; therefore, values are not given in SI units.)

We first created a data file that, for each line section, contains the following information:

1. Starting node,
2. Ending node,
3. Direction (inbound or outbound),
4. Line number,
5. Observed bus time,
6. Observed automobile time (for complete sequence of links),
7. Minutes per mile for the bus on the section,
8. Minutes per mile for automobiles on the section, and
9. Number of road links in the section.

The file contains observations for 470 line sections. On 25 segments, the observed transit time was smaller than the observed automobile time. Since this problem seems to be related to the accuracy of the data, these observations were not considered in the calibration of the model.

We first introduce some notation:

Let

TA = automobile time on the line section (min),
TB = bus time on the line section (min),
TMA = automobile time per mile on the line section (min/mile), and
TMB = bus time per mile on the line section (min/mile).

First, we plotted TB as a function of TA. Figure 1 shows the resulting scatter diagram; a linear function was fitted, resulting in an R^2 of 0.87. However, some contemplation of this relationship reveals that, over long sections, both the bus and the automobile times are relatively long, and, of course, on short sections, both times are relatively small (that is, they are both correlated to link length). Evidently, such a model would not capture any effect of congestion.

We proceeded then to analyze the inverse of speed (time per mile) (which is used in the formulation of volume-delay curves). A simple linear model of TMB versus TMA resulted in a poor fit of $R^2 = 0.2$. A linear model of TMB versus TMA and TA increased the R^2 to 0.49, which also was not satisfactory. In both of the above cases, we tried different models for the inbound and outbound direction but the fits, reflected in the R^2 values, were not improved.

A plot of (TMB/TMA) versus TMA showed that a nonlinear model could be more appropriate (Figure 2). An exponential model of the form

$$\ell n[(TMB/TMA) - 1] = a_0 + a_1 TMA \quad (1)$$

was estimated by linear regression. Again, with an $R^2 = 0.09$, the model was rejected. We then attempted to use a polynomial model of the form

$$(TMB/TMA) - 1 = a_1 (TMA)^{-1/2} + a_2 (TMA)^{-1} + a_3 (TMA)^{-3/2} + a_4 (TMA)^{-2} \quad (2)$$

which was estimated with a stepwise linear regression. The only term that entered in the regression was $a_1 (TMA)^{-1/2}$ and it resulted in an R^2 of 0.62, which, considering the accuracy of the data, was the first satisfactory result obtained. The analytical form of this model (M1) is

$$(TMB/TMA) - 1 = 1.97\sqrt{1/TMA} \quad (3a)$$

or

$$TMB = TMA + 1.97\sqrt{TMA} \quad (3b)$$

As an alternative, we considered a linear model of the form $TMB = m(t_0) + TMA$, where t_0 is the inverse of the free-flow speed of the road link. Values of t_0 were obtained from the road network data. A linear regression gave an R^2 of 0.62; the model (M2) is as follows:

$$TMB = TMA + 1.43\, t_0 \quad (4)$$

where $1.43\, t_0$ is a constant penalty in minutes per mile for transit vehicles, which is related in some way to the link type.

The next step was to make an evaluation of the predictive ability of models M1 and M2. Since we are mainly interested in predicting good transit impedances (origin to destination path times), we decided to compare for each line and each direction in the data file the sum of the predicted travel times on each section against the corresponding observed times. The results were good for most of the lines (within 10 percent) except for express services and for some high-speed regular lines.

It became evident that a natural way to improve the

Figure 1. Bus travel times versus automobile travel times.

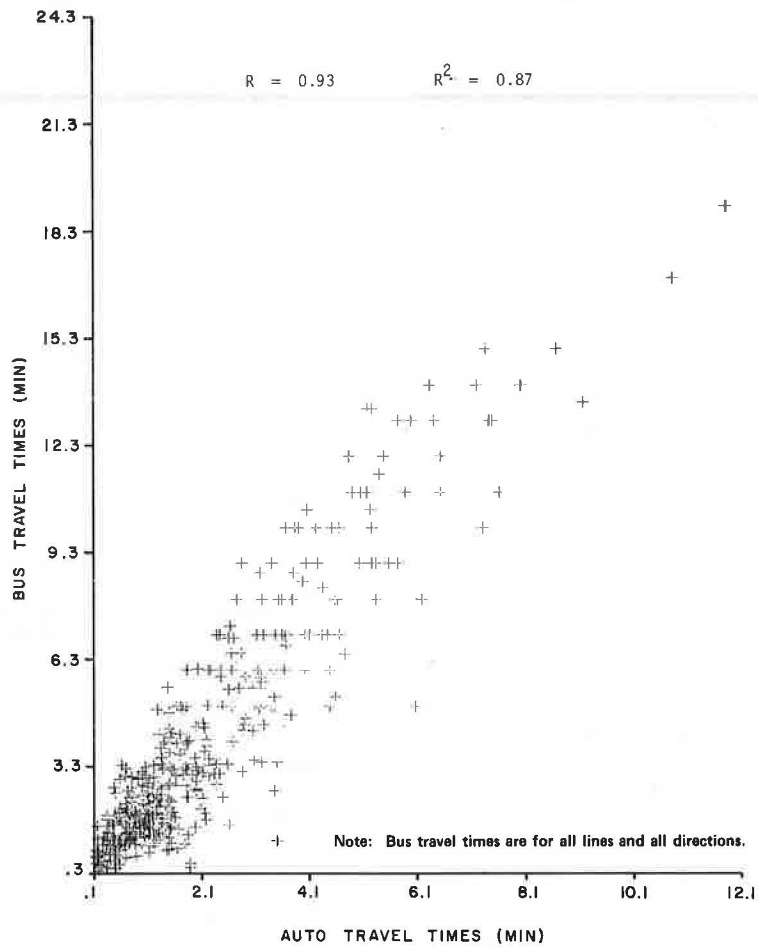
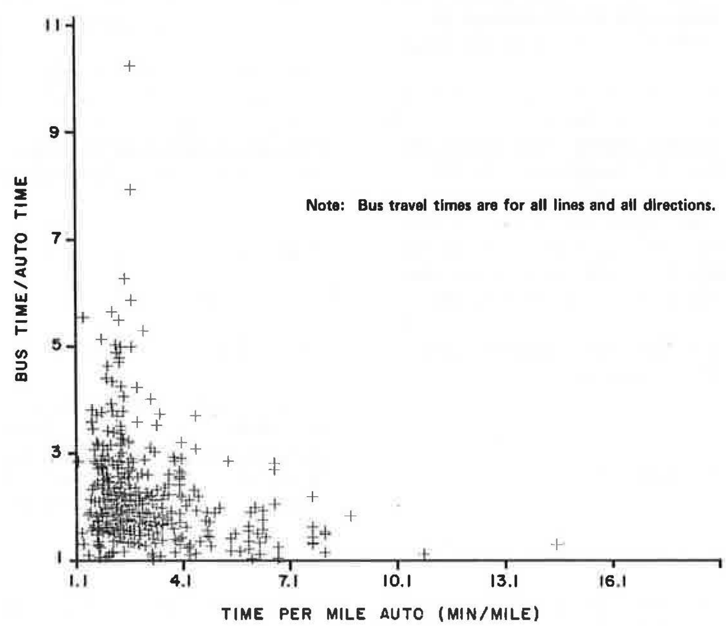


Figure 2. Bus and automobile time versus time per mile by automobile.



models was to stratify the data according to service type. The three types considered were

1. Feeder—0 observations in data file,
2. Regular—426 observations, and
3. Express—44 observations.

Models of the same form as M1 and M2 were estimated for the express service. The R^2 values were close to 0.5 and the models were not significant because of the rather small number of observations available. In the case of regular service, the recalibration of models M1 and M2 resulted in the relations:

$$TMB = TMA + 2.1\sqrt{TMA} \tag{5a}$$

$$TMB = TMA + 1.49\, t_0 \tag{5b}$$

The R^2 values improved slightly (0.64), but overall the models did not change significantly. We then subdivided the regular service into two categories by considering the average observed speed of each line. All the lines that ran at less than 10 mph were classified as regular and the others as fast regular.

For the fast-regular lines, the recalibration of model M1 results in

$$TMB = TMA + 2.15 \qquad R^2 = 0.84 \tag{6}$$

and the recalibration of model M2 results in

$$TMS = TMA + 0.9\, t_0 \qquad R^2 = 0.84 \tag{7}$$

For the regular lines, model M1 becomes

$$TMB = TMA + 9.14/\sqrt{TMA^1} \qquad R^2 = 0.78 \tag{8}$$

and model M2 becomes

$$TMB = TMA + 2.12\, t_0 \qquad R^2 = 0.73 \tag{9}$$

This time the comparison, for each line and direction, of the sum of observed and predicted times on each section showed that TMB = TMA + 2.15 is a good model for fast-regular lines. In the case of regular lines, both models had to be rejected. Our next step in the analysis of the regular lines was to go back to the previous form of the model, that is, to estimate a function of the form

$$TMB = TMA + \alpha\sqrt{TMA} \tag{10}$$

that had proved to be satisfactory for regular lines, except for the fast ones. The estimation resulted in an $\alpha = 3.21$ and an $R^2 = 0.73$. Unfortunately, the comparison of the sum of line-section times showed that the previous model ($\alpha = 2.1$, $R^2 = 0.64$), which had been estimated on all regular-lines data (fast regular and regular), gave better results than did the new one, which had been estimated by using data for regular lines (< 10 mph) only.

The above analysis suggested that it may be advantageous to define fast-regular lines by using a higher speed value. But further experiments indicated that the results could not be improved in this way.

In consideration of the above analysis and the fact that we did not have sufficient data for feeder and express services, we finally selected and implemented the following bus travel-time relationships. On transit-only links, the user-defined travel times are used. On transit links that correspond to road links, four cases are considered:

1. For a feeder service (line type 3) the user-defined line speed is used on all links, independent of automobile speed;
2. For an express service (line type 1) the bus speed is the same as automobile speed (TMB = TMA);
3. For a fast-regular service (line type 2) we have TMB = TMA + 2.15 (regular lines with an average speed of 10 mph or more were considered as fast lines); and
4. For a regular service (undefined line type) we have TMB = TMA + 2.1 $*\sqrt{TMA}$ where TMB = minutes by mile for bus and TMA = minutes by mile for automobile.

The relationships were applied to predict the transit travel time on each of the 1755 transit links of the coded network. On the basis of those predicted times, transit paths between selected origin-destination pairs were calculated. An analysis of the transit times and paths suggested that we should change the classification of some of the lines. After a few iterations of this procedure, we made final classifications for all of the lines.

An important fringe benefit of having included a bus time model in EMME is that the user does not have to define a travel time for each of the transit links; thus the coding of the network is made much easier.

RECALIBRATION OF THE VOLUME-DELAY CURVES

The volume-delay curves used by the city of Winnipeg were developed in the early 1960s by Traffic Research Corporation and had the functional form

$$S_a(v_a) = d_a\left(\delta + \alpha[(v_a/\ell_a) - \gamma] + \{\alpha^2[(v_a/\ell_a) - \gamma]^2 + \beta\}^{1/2}\right) \tag{11}$$

We modified this functional form by replacing it with the simpler BPR formula:

$$S_a(v_a) = d_a t_0 [1 + \alpha(v_a/c_a)^\beta] \tag{12}$$

where

d_a = the link length,
v_a = the link volume,
ℓ_a = the number of lanes of the link, and
c_a = the practical capacity of the link.

The other parameters are calibrated from the observed data. The initial transformation was done by Branston (7). He estimated a practical capacity for each of the volume-delay curves and then calibrated the constants α, β of the BPR formula by using the predicted times of the Traffic Research Corporation functions.

We then recalibrated the BPR curves obtained in this way by using the 1976 data and the following procedure. For each volume-delay curve, the observed data were aggregated by using a subdivision of the link volumes (v_a) into intervals, and mean values were computed for each interval. The curves and the resulting mean values of the travel times were plotted and analyzed; as a result, new free-flow speeds were determined and then the curves were replotted. This procedure was repeated three times, resulting in a new set of α, β, and t_0. Table 1 shows the values that were actually used.

It was evident from the plots used to determine the free-flow speed that certain links, which exhibited observed times below and to the right of the curves, would be better predicted by delay curves that represent higher-capacity links. In order to assist the city of Winnipeg in this reclassification of links to different curves, a report was produced for all links for which an observed flow was available. This report gives the time predicted by the currently assigned curve and also other curves that would predict the travel time better and still respect the speed limit. This report was used to reclassify links on a route basis. Links that had large differences between predicted and observed times were plotted on a map in order to determine the links of an avenue or street that had to be reclassified. In some cases the number of lanes was corrected as well. This analysis also resulted in the correction of some observed travel times and volumes. In total, 159 links were reclassified, the number of lanes was changed for 23 links, the observed time was updated for 192 links, and the observed volume was updated for 21 links. Figure 3 shows plots of the origin-to-destination travel times along shortest paths computed by using the volume-delay curves versus the observed times.

CALIBRATION AND VALIDATION OF THE ROAD-NETWORK ASSIGNMENT

The calibration of the road network was achieved by comparing the observed link volumes with the link volumes predicted by the traffic-assignment model. The comparison is performed by using specially written programs and by manually comparing screen-line totals for the observed and predicted flows. Discrepancies between observed and predicted values may be caused by the errors introduced in the total automobile origin-destination matrix or by improper coding of the road network. Since the 1976 road network differs little from the 1971 network, which was carefully calibrated, the corrections necessary to the coding of the road network were all found during the recalibration of the volume-delay curves and most of the adjustments made involved the total automobile origin-destination matrix.

This matrix is calculated from the total person work-trip matrix by using the observed modal-split and automobile-occupancy matrices and a set of adjustment factors that serve to add other-purpose trips and truck trips; that is

$$g_{pq}^{au} = (g_{pq} * r_{pq}/\gamma_{pq}) * f_{pq} \qquad (13)$$

where

(p, q) = an origin-destination pair of zones,
g_{pq} = the total person work trips between q and p,
r_{pq} = the proportion of trips by automobile,
γ_{pq} = the automobile occupancy, and
f_{pq} = the factor for other trips and truck trips.

The factors f_{pq} are given as a matrix of values for 10 groups of zones (super zones). The essence of the calibration procedure was a trial-and-error process that was aimed at finding the most appropriate factors based on the comparison of observed and predicted link volumes. While this was carried out, 19 errors in the observed link volumes were detected and corrections were made.

All the factors in the calibration procedure were determined by the staff of the city of Winnipeg by using screen-line counts. The screen lines chosen divide the city into three quadrants by using natural geographic subdivisions. A specially written program selects the links that cross each of these lines and provides the observed and predicted volumes, which are then totaled for each screen line.

First, an assignment was produced by using only the

Table 1. Volume-delay functions.

CACO	Speed Limit (mph)	α	β	Free-Flow Times (t_0)	Minutes per Mile
1	0-30	0.7312	3.6596	15.0	4.00
2	0-30	0.6218	3.5038	17.0	3.53
3	0-30	0.8774	4.4613	20.0	3.00
4	0-30	0.6846	5.1644	23.0	2.61
5	0-30	1.1465	4.4239	25.0	2.40
6	31-40	0.6190	3.6544	30.0	2.00
7	31-40	0.6662	4.9432	32.4	1.85
8	31-40	0.6222	5.1409	32.4	1.85
9	31-40	1.0300	5.5226	35.3	1.70
10	41-50	0.6609	5.0906	41.4	1.45
14	41-50	0.5423	5.7894	41.4	1.45
15	41-50	1.0091	6.5856	41.4	1.45
15	+50	0.8776	4.9287	55.0	1.09
16	+50	0.7699	5.3443	55.0	1.09
18	+50	1.1491	6.8677	55.0	1.09

Figure 3. TRC versus BPR curves.

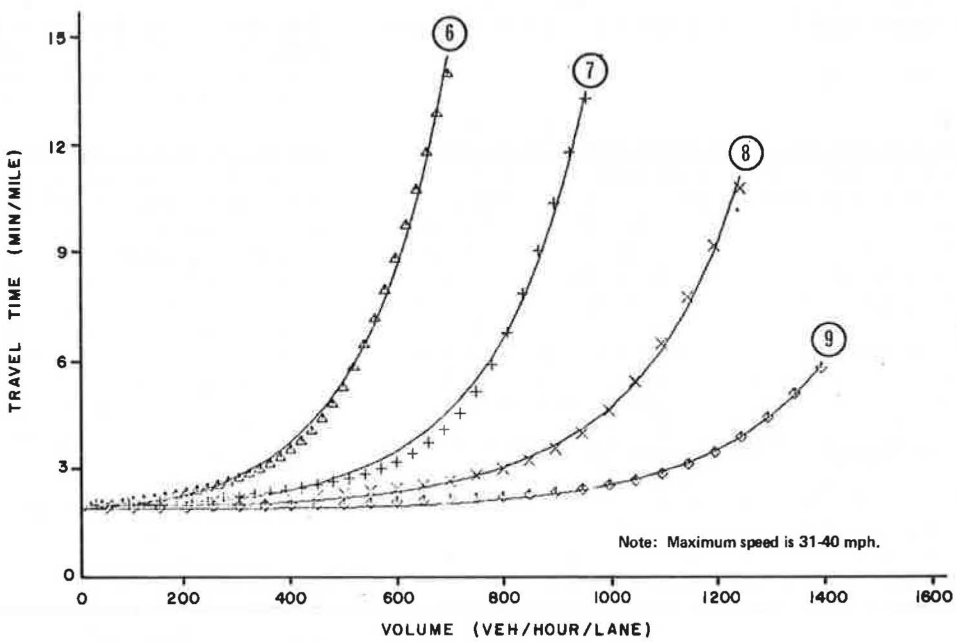


Table 2. Parameters assigned for the analysis.

Assignment	WALK	WAIT	WFAC	WPEN	WMIN	WMAX
First stage						
1	3.0	3.0	0.5	0.0	0.0	10.0
2	0.0	1.0	0.5	0.0	0.0	10.0
3	0.5	2.0	0.5	4.0	0.0	10.0
Common line section						
4	3.0	3.0	0.5	0.0	0.0	10.0
5	0.0	1.0	0.5	0.0	0.0	10.0
6	0.5	2.0	0.5	4.0	0.0	10.0
7	0.5	3.0	0.5	4.0	0.0	10.0

automobile work-trip matrix, which is obtained by setting $f_{pq} = 1$. By relating super-zone pairs with screen-line crossings it is possible to adjust the various factors to increase or decrease the interchanges across the screen lines. The correspondence is as follows:

Quadrant	Super Zones
1	1, 2, 3
2	4, 5
3	Rest + downtown (0)

Various other considerations were taken into account in determining the factors (f_{pq}), such as the low production of truck trips by residential areas and the high production of truck trips by industrial zones.

TRANSIT NETWORK VALIDATION AND CALIBRATION

This part of the project required considerable effort, since prior to this study the city of Winnipeg did not have a transit network model and the work included the definition of the network, its coding, validation, and calibration.

The purpose of the validation is to make sure that the transit system is described properly. The coded network must represent adequately all possible passenger movements and transit vehicle movements. The validation of the network consists, then, of ensuring that the coding rules have been followed correctly and that the representation of the two types of movements is satisfactory. The tools used in validation are

1. EMME data bank programs, which perform the syntactic and data consistency checks;
2. Network generation programs, which ensure that the rigorous restrictions imposed on the input data in order to realize the interface with the road network and to determine transit travel times are satisfied;
3. Graphical displays of the network;
4. Manual checks of the data; and
5. Analysis of the complete printout of the transit assignment.

This task was carried out in cooperation with the staff of the city of Winnipeg.

The calibration deals with the other aspect of the transit system, that is, the behavior of the transit passengers in the selection of paths on the network. Given the shortest-path behavior hypothesis, it is necessary to estimate the value of certain parameters of the transit path algorithm in order for it to produce satisfactory paths between the various origin-destination pairs.

The parameters to be estimated are

1. WFAC—a regularity factor relating the waiting time to the headway of the line to be boarded,
2. WMIN—the minimum waiting time,
3. WMAX—the maximum waiting time,
4. WAIT—the weight of waiting time used in the calculation of the impedance of a path in generalized time units,
5. WPEN—a constant penalty added to the impedance every time the passenger has to wait for the bus, and
6. WALK—the weight of walking time (access-egress) used in the calculation of the impedance of a path.

A given path that contains n line sections has an impedance, in generalized time units, that is given by the expression:

$$\text{IMP} = \text{WALK} * (\text{access} + \text{egress time}) + \text{WAIT} * \sum_{\ell=1}^{n} W_\ell + \sum_{\ell=1} T_\ell + n * \text{WPEN} \qquad (14)$$

where

W_ℓ = the waiting time of the ℓth line defined as W min [max (WMIN, WFAC * HDW_ℓ), WMAX],
HDW_ℓ = the headway of the ℓth line and transfer time and is considered as being included in waiting time, and
T_ℓ = the in-vehicle time spent on the ℓth line, which is assumed to have a wait of 1.0 in the impedance calculations.

For each origin-destination pair the algorithm selects the path with minimum impedance from origin O to destination D. The best way to calibrate the transit model would be to compare the predicted paths to the actual paths obtained from the origin-destination survey. Unfortunately, in the Winnipeg survey there was no question about the path used by transit riders. The method that we used consisted of analyzing the predictions of a transit assignment by comparing it with the observed volumes on the segments. Analyses were also made on level-of-service statistics (i.e., mean total trip time, mean number of transfers, and distribution of total trip time) and on predicted line volumes. Given the all-or-nothing aspect of the assignment, only large volumes may be analyzed. The following volumes were analyzed:

1. The volume at the maximum load point of each line in both directions,
2. The location of the maximum load point,
3. The volume profiles on lines, and
4. Screen-line volumes [entering and leaving the central business district (CBD), bridges, and other high-volume links].

In the first stage of the analysis, three assignments (assignments 1-3 in Table 2) were performed by use of the parameters given.

The analysis made by the staff of the city of Winnipeg showed that assignment 1 was the best one, but the split of volumes between competing lines was not satisfactory. We then introduced the "common line section" algorithm in the model. With this algorithm the passengers are diverted over common bus lines proportionally to the frequency of each line (i.e., passengers are assumed to board the first line that arrives at the bus stop). We ran four new simulations (assignments 4-7 in Table 2).

Assignment 4, which is similar to number 1, proved to be the best one and the spread of volumes had improved significantly.

CALIBRATION OF THE MODAL-SPLIT FUNCTION

The basic data that were used for the calibration of the modal-split function are the results of the origin-destination survey that was carried out by the city of Winnipeg in the spring and summer of 1976. The survey was carried out in large part by home interviews of a sample of 20 percent of households. The actual sample size obtained was roughly 17 percent, after refusals and rejections have been taken into account. In addition, a survey questionnaire, which was to be returned by mail, was distributed to the students of the three Winnipeg universities; the effective sample of student trips was approximately 23 percent. The total sample, consisting of the individual detailed data, amounted to 52 424 questionnaires and these data were transmitted to us on a magnetic tape by the city of Winnipeg. Then, the departure codes, described earlier, were applied in order to separate the trips that occurred during the 7:30-8:30 a.m. peak. There were 17 761 individual records in the peak-hour subsample.

The number of trips that occurred during the peak hour was expanded by the proportion of the sample in each zone, which was calculated by the city of Winnipeg, in order to obtain the following origin-destination matrices:

Automobile drivers and passengers—1, automobile drivers—1′
Transit passengers—2
Total trips—3 = (1 + 2)
Modal split—4 = (1/3)
Automobile occupancy—5 = (1/1′)

(The automobile drivers and passengers origin-destination matrix shall be referred to as the automobile origin-destination matrix.) The automobile origin-destination matrix was scaled by the appropriate factor to obtain the total automobile origin-destination matrix and this last was assigned to the road network by using the equilibrium traffic assignment of EMME. The resulting origin-to-destination travel times constitute the origin-destination matrix of

Road travel times—6

and by tracing a set of shortest paths on the links that carry flow we obtain the origin-destination matrix of

Distance by road—7

Next, the transit origin-destination matrix was used to calibrate the transit assignment. Other than refinements of the transit network representation, this calibration determines the coefficients of generalized time (or cost) in the expression

$$\text{Transit impedance} = \alpha(\text{Access time} + \text{egress time} + \text{wait time}) + \beta(\text{In-vehicle time}) \quad (15)$$

As described earlier, the values for α and β, determined in cooperation with the city of Winnipeg, are 3 and 1, respectively. Thus we obtained the origin-destination matrix of

Transit impedance—8

and by tracing the shortest paths used we determined the origin-destination matrix of

Number of transit transfers—9

Since our approach is to calibrate a zonal-aggregate modal-split function, we extracted from the survey data (a) the average automobile ownership per household per zone, (b) the proportion of adults who travel at the peak hour, and (c) the proportion of students who travel at the peak hour for each origin-destination pair.

The other socioeconomic variables were obtained by the city of Winnipeg from various sources and transmitted to us. The Statistics Canada 1976 Census provided the average income per household per zone and the origin-destination survey estimated the number of jobs per zone. The parking costs per month per zone and the number of parking spaces per job per zone were evaluated by using 1971 data.

Thus, in all, a file was constructed that consisted of the dependent variable, the modal split, and the independent explanatory variables outlined above. This file contained the records for all origin-destination pairs that had more than 60 trips by both modes in the expanded matrix (3) of trips by both modes. The main reason for adopting this procedure is that the modal split for origin-destination pairs with smaller demand would have far more variability due to the relatively small number of trips in the sample.

The functional form that we chose for the calibration is that of the logistic function. Although this form achieved recent fame in its use as a disaggregate probabilistic-choice function, we use it with aggregate data due to its ease of manipulation and its property of predicting choice values with a smooth ogive-type curve. The form that we used is

$$p_{au} = 1/1 + \exp(k_0 + \sum_i k_i x_i) \quad (16)$$

where

p_{au} = the proportion of trips that occur by automobile,
k_0 = a constant, and
k_i, $i = 1, \ldots, m$ = the coefficients associated with the x_i, $i = 1, \ldots, n$ explanatory variables.

A simple algebraic manipulation results in the form $\ln(1 - p_{au}/p_{au}) = k_0 + \sum_i k_i x_i$, which is used for calibrating k_0, k_i, $i = 1, \ldots, n$ by simple linear regression. This method of estimation is often referred to as Berkson-Theil estimation to acknowledge their early work (8, 9) in aggregate logistic-function calibration.

Another functional form that we tried is the so-called "dogit" proposed recently by Gaudry (10), which adds to the logit form modal constant θ_{au}, θ_{tr} as follows:

$$p_{au} = (1/1 + \theta_{au} + \theta_{tr})[1/1 + \exp(k_0 + \sum_i k_i x_i)] + \theta_{au} \quad (17)$$

However, in all of the trials that we performed, the best values for θ_{au}, θ_{tr} were always zero; that is, the logistic function was satisfactory and neither of the two modes considered had a fixed proportion (θ_{au} or θ_{tr}) of the modal split as an advantage.

The actual calibration test spanned a period of eight months, during which several hundred regressions were run by also using transformations of the explanatory variables. The best modal-split model for all the considered origin-destination pairs is given in Table 3.

We were not entirely satisfied with this model because the best fit obtained with a transformation of variables was not much better, as can be seen in Table 4.

We then subdivided the origins into subgroups by using a criterion related to the error introduced by the modal-split function. We reasoned that errors on individual origin-destination pairs were unavoidable; however, the model should not distort the origin-destination matrix.

That is, there should not be too much bias introduced on demand totals by origins and destinations. Thus, we subdivided the origins into subgroups according to the error introduced by the model on origin totals; that is, origins that had negative deviations were grouped together and origins that had positive errors and origins that had acceptable error formed a second and third subgrouping. Finally, we obtained four modal-split models as shown in Table 5.

BASE-YEAR CALIBRATION—BIMODAL MODEL

The execution of a bimodal assignment in EMME requires the simultaneous use of the vehicle assignment, the transit assignment, and the modal-split function. Each is calibrated independently and then used jointly in the computations. Since each introduces a certain error by its calibration, there will be some differences between the observed values and the output of the bimodal computations for the base year. Fortunately, these differences are not large and are well within the variances that are acceptable in calibration of transportation models.

The staff of the city of Winnipeg asked that we apply the modal-split function to all of the origin-destination pairs, even though we had calibrated the model by using only origin-destination pairs that had more than 60 trips in the expanded total trip matrix. This became necessary because only about 28 percent of the total trips were represented by that sample. Thus we applied, to all origins that were not represented in the calibration data, the initial modal-split model (model 1). For all other origins, we applied the corresponding modal-split models to all of the relevant destinations. The results were surprisingly good. Only 307 trips (or 0.5 percent of the total number of trips) are the difference between the observed total number of trips by automobile and the differences that result on trip ends (that is, origin and destination totals) are mostly of the order of up to 8 percent. The predicted origin-destination matrix is plotted versus the observed origin-destination matrix in Figure 4.

We judged these demand differences acceptable in view of the general consideration that the true demand varies daily and differences of the order of 10 percent between various days of the week are accepted to be commonplace. Further, these differences were not sufficiently high to materially change the orders of magnitude of the link flows on the important arteries.

The computation times on the CDC-Cyber-176 of the University of Montreal for the base year bimodal run are as follows:

Computation Step	Time (s)
Generate transit network	3.66
Calculate bus frequency	1.23
Calculate transit impedance	1110.55
Initialize road traffic demand	3.44
Perform road traffic assignment	2509.44
Modify transit link times	67.61
Calculate fixed transit demand	1.77
Calculate demand function (transit)	0.00
Perform transit assignment	423.91
Modify transit capacity	0.00

The costs are given below.

Function	Cost ($)
Central processor	189.20
Input-output	22.80
Fast memory	560.40
Total	772.40

CONCLUSION

There are several ways in which EMME may be used to simulate the impact of contemplated improvement scenarios. One may use the single-mode assignment mod-

Table 3. Model 1 parameter values.

Variable	Parameter Value	95 Percent Confidence Interval
Constant (k_0)	2.563	1.758 to 3.369
Transit impedance	-0.0122	-0.220 to -0.002 42
Automobile time	0.0220	0.001 92 to 0.0422
Proportion men	-3.279	-4.117 to -2.441
Parking cost	0.0745	0.0532 to 0.0957
Automobile availability	-1.904	-2.726 to -1.082
R^2	0.60	
R	0.77	

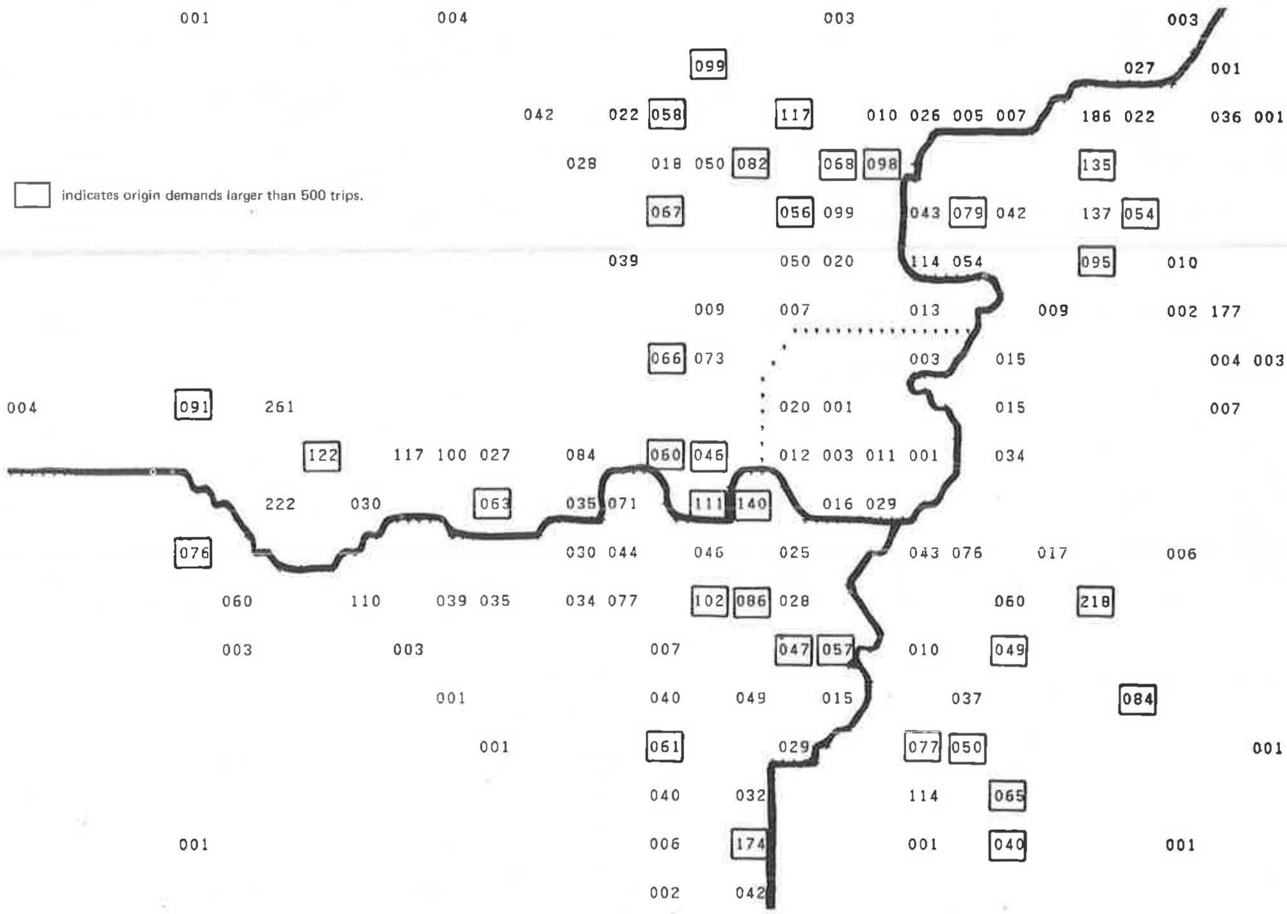
Table 4. Model 2 parameter values.

Variable	Parameter Value	95 Percent Confidence Interval
Constant (k_0)	16.566	10.000 to -23.131
(Transit impedance)^2	-0.000 758	-0.000 127 to -0.000 242
(Automobile time)^2	0.000 242	-0.000 180 to 0.000 665
(Proportion men)^2	-2.256 654	-3.478 to -2.234
Parking cost	0.363 40	0.276 to 0.444
ℓn (income)	-1.752	-2.446 to -1.058
R^2	0.64	
R	0.80	

Table 5. Model 3 parameter values.

	Model 3a		Model 3b		Model 3c		Model 3d	
Variable	Value	95 Percent Confidence Interval	Value	95 Percent Confidence Interval	Value	95 Percent Confidence Interval	Value	95 Percent Confidence Interval
Constant	2.352	1.004 to 5.708	1.516	0.297 to 2.735	3.071	-0.014 0 to 6.156	2.935	1.860 to 4.010
Transit impedance	-0.0133	-0.0532 to 0.0265	-0.0101	-0.022 1 to 0.001 78	-0.0315	-0.051 3 to -0.0118	-0.0139	-0.0408 to -0.131
Automobile time	0.0334	-0.0503 to 0.117	0.0253	0.000 186 to 0.050 4	0.0733	0.034 3 to 0.112	0.0323	-0.0302 to 0.0949
Proportion men	-2.799	-5.646 to 0.0491	-3.719	-4.827 to -2.611	-3.101	-5.043 to -1.159	-2.499	-4.020 to -0.978
Parking cost	0.0959	0.0225 to 0.169	0.0968	0.069 4 to 0.124	0.471	0.004 71 to 0.0895	0.0332	-0.0124 to 0.0787
Automobile availability	-3.023	-6.359 to 0.312	-1.232	-2.394 to -0.069 8	-1.502	-4.686 to 1.682	-1.844	-3.139 to -0.548
R^2	0.72		0.67		0.78		0.64	
R	0.85		0.82		0.88		0.80	
Number of origin-destination pairs	26		135		36		45	

Figure 4. Predicted automobile demand by origin for bimodal run.



ules and thus simulate the impact of the scenario without changing the modal shares of the demand. This may be appropriate for some situations where only marginal improvements are made and the only interest is to anticipate the changes in route choice that result due to the modifications. However, most current transportation planning methods have this capability. The other way to use EMME is to simulate the impact of each scenario with a full bimodal run, which would predict the anticipated changes in modal share of demand as well. This capability is so far unique to EMME.

The main conclusion that we draw from this project is that the use of sophisticated models, such as EMME, is feasible and the simulation of scenarios results in refined and fully detailed evaluations, which would not be possible otherwise. The main obstacles are the quality of the available data and the calibration of the demand model. Fortunately, we had access to very good data and we succeeded to calibrate a satisfactory modal-split model.

The costs of building up the necessary data base and calibrating the model are relatively high; however, the use of the model is not expensive. The figure of $800 for each bimodal simulation is reasonable, when one considers that the analyst's time to set up a scenario and analyze the EMME output is one to two days.

ACKNOWLEDGMENT

We would like to thank Ed Guertin and Jarvis Kohut of the city of Winnipeg for their precious collaboration and the Transportation Research and Development Center of Transport Canada for their financial support of this work.

REFERENCES

1. M. Florian. A Traffic Equilibrium Model of Travel by Car and Public Transit Modes. Transportation Sciences, Vol. 11, No. 2, 1977, pp. 166-179.
2. M. Florian, R. Chapleau, S. Nguyen, C. Achim, L. James, and J. Lefebvre. EMME: A Planning Method for Multi-Modal Urban Transportation Systems. Proc., World Transportation Research Conference, Rotterdam, the Netherlands, 1977.
3. C. Fisk and S. Nguyen. Properties of the Two-Mode Equilibrium Assignment Model with Flow Dependent Transit Costs. Centre de Recherche sur les Transports, Univ. de Montréal, Publication 121, 1979.
4. M. Florian and S. Nguyen. An Application and Validation of Equilibrium Trip Assignment Methods. Transportation Science, Vol. 10, 1976, pp. 374-390.
5. C. Chriqui and P. Robillard. Common Bus Lines. Transportation Science, Vol. 9, No. 2, 1975, pp. 115-121.
6. C. Achim and R. Chapleau. EMME-Coding the Transit Network. Centre de Recherche sur les Transports, Univ. de Montréal, Publication 49, 1976.
7. D. Branston. Link Capacity Functions: A Review. Transportation Research, Vol. 1, 1976, pp. 223-236.
8. J. Berkson. Maximum Likelihood and Minimum χ^2 Estimate of the Logistic Function. Journal of

the American Statistical Association, No. 50, 1955, pp. 130-161.
9. H. Theil. Principles of Econometrics. Wiley, New York, 1971, pp. 65-76.
10. M. Gaudry and M. G. Dagenais. The Dogit Model. Centre de Recherche sur les Transports, Univ. de Montréal, Publication 82, 1978, 14 pp.

Publication of this paper sponsored by Committee on Traveler Behavior and Values.

Confidence Intervals for Choice Probabilities of the Multinomial Logit Model

Joel Horowitz, U.S. Environmental Protection Agency

This paper describes three methods for developing confidence intervals for the choice probabilities in multinomial logit models. The confidence intervals reflect the effects of sampling errors in the parameters of the models. The first method is based on the asymptotic sampling distribution of the choice probabilities and leads to a joint confidence region for these probabilities. This confidence region is not rectangular and is useful mainly for testing hypotheses about the values of the choice probabilities. The second method is based on an asymptotic linear approximation of the relation between errors in models' parameters and errors in choice probabilities. The method yields confidence intervals for individual choice probabilities as well as rectangular joint confidence regions for all of the choice probabilities. However, the linear approximation on which the method is based can yield erroneous results, thus limiting the applicability of the method. A procedure for setting an upper bound on the error caused by the linear approximation is described. The third method is based on nonlinear programming. This method also leads to rectangular joint confidence regions for the choice probabilities. The nonlinear programming method is exact and, therefore, more generally applicable than the linear approximation method. However, when the linear approximation is accurate, it tends to produce narrower confidence intervals than does the nonlinear programming method, except in cases where the number of alternatives in the choice set is either two or very large. Several numerical examples are given in which the nonlinear programming method is illustrated and compared with the linear approximation method.

The multinomial logit formulation of urban travel-demand models has a variety of theoretical and computational advantages over other demand-model formulations and is receiving widespread use both for research purposes and as a practical demand-forecasting tool (1-3). However, travel-demand forecasts derived from logit models, like forecasts derived from other types of econometric models, are subject to errors that arise from several sources, including sampling errors in the estimated values of parameters of the models, errors in the values of explanatory variables, and errors in the functional specifications of the models. Knowledge of the magnitudes of forecasting errors can be important in practice, particularly if either the errors themselves or the costs of making erroneous decisions are large. This paper deals with the problem of estimating the magnitudes of forecasting errors that result from sampling errors in the estimated values of the parameters of logit models. Specifically, the paper describes techniques for developing confidence intervals for choice probabilities and functions of choice probabilities (e.g., aggregate market shares, changes in choice probabilities caused by changes in independent variables) derived from logit models, conditional on correct functional specification of the models and use of correct values of the explanatory variables.

A model's forecasting error can be characterized in a variety of ways, including average forecasting error and root-mean-square forecasting error, in addition to confidence intervals for the forecast. Among the various error characterizations, only the confidence interval provides a range in which the true value of the forecast quantity is likely to lie. Methods for developing confidence intervals for the forecasts of linear econometric models are well known (4). However, these methods are not applicable to logit models, which are nonlinear in parameters. Koppelman (5, 6) has analyzed the forecasting errors of logit models and has described the ways in which various sources of error contribute to total error in forecasts in choice probabilities. Koppelman's error measures do not include confidence intervals for the choice probabilities although, as will be shown later in this paper, one of his error measures can be used to derive approximate confidence intervals.

Three methods for estimating confidence intervals for the choice probabilities of logit models are described in this paper. All of the methods lead to asymptotic confidence intervals in that they are based on the large-sample properties of the estimated parameters of the models. The first method is based on the exact asymptotic sampling distribution of the choice probabilities and leads to a joint confidence region for these probabilities. This region is useful mainly for testing hypotheses about the values of the choice probabilities. The region is not rectangular and, therefore, is difficult to use in practical forecasting. Moreover, the methods used to derive the confidence region cannot be readily extended to functions of the choice probabilities.

The second method is based on an asymptotic linear approximation of the relation between sampling errors in models' parameters and sampling errors in choice probabilities. The linear approximation method yields confidence intervals for individual choice probabilities as well as rectangular joint confidence regions for all of the choice probabilities. The method can easily be extended to functions of the choice probabilities. However, the linear approximation on which the method is based can yield erroneous results, thus limiting the method's applicability. A procedure for placing an upper bound on the error caused by the linear approximation is described.

The third method is based on nonlinear programming. This method yields rectangular joint confidence regions for the choice probabilities and can be extended to functions of the choice probabilities. The method does not require approximation of the relations between sampling errors in models' parameters and sampling errors in choice probabilities and, therefore, is more generally applicable than is the linear approximation method. Several numerical examples are given in which the nonlinear programming method is illustrated and compared with the linear approximation method.

PROPERTIES OF THE LOGIT MODEL

In the multinomial logit model, the probability that individual n selects alternative i from a set of J_n available alternatives is given by

$$P_{in} = \exp(V_{in}) \Big/ \sum_{j=1}^{J_n} \exp(V_{jn}) \tag{1}$$

where P_{in} is the probability that alternative i is chosen by individual n, and V_{jn} ($j=1, \ldots, J_n$) is the systematic component of the utility of alternative j to individual n.

For each alternative i, V_{in} is assumed to be a linear function of appropriate explanatory variables. Thus

$$V_{in} = \sum_{m=1}^{M} X_{imn} \alpha_m \tag{2}$$

where

- M = the number of explanatory variables,
- X_{imn} = the value of the m th explanatory variable for alternative i and individual n, and
- α_m = the coefficient of explanatory variable m.

The values of the coefficients (or parameters) α_m ordinarily are not known a priori and are estimated from observations of individuals' choices by using the method of maximum likelihood. Details of the estimation procedure and the statistical properties of the estimated coefficients are described by McFadden (7).

Denote the estimated coefficients by $\{a_m; m=1, \ldots, M\}$. For each alternative i and individual n define $\hat{V}_{in}$ by

$$\hat{V}_{in} = \sum_{m=1}^{M} X_{imn} a_m \tag{3}$$

$\hat{V}_{in}$ is the estimated systematic utility function for alternative i and individual n. $\hat{V}_{in}$ is a random variable by virtue of its dependence on the random variables $\{a_m\}$. Define

$$\hat{P}_{in} = \exp(\hat{V}_{in}) \Big/ \sum_{j=1}^{J_n} \exp(\hat{V}_{jn}) \quad (i = 1, \ldots, J_n; n = 1, \ldots, N) \tag{4}$$

$\hat{P}_{in}$ estimates the probability that individual n makes choice i and is the forecast of the choice probability that is used in applications of the logit model. Accordingly, the subsequent sections of this paper are concerned with the development of ranges about the $\hat{P}_{in}$ that are likely to contain the true choice probabilities P_{in}.

Assume that the coefficients $\{a_m\}$ have been estimated by the method of maximum likelihood by using a data set that consists of observations of N individuals' choices. Then for large N, the estimated coefficients $\{a_m\}$ are asymptotically jointly normally distributed with mean values $\{\alpha_m\}$ and covariance matrix A^{-1}, where

$$A_{rs} = -\sum_{n=1}^{N} \sum_{j=1}^{J_n} (X_{jrn} - X_{\cdot rn})(X_{jsn} - X_{\cdot sn}) P_{jn} \quad (r,s = 1, \ldots, M) \tag{5}$$

and

$$X_{\cdot rn} = \sum_{j=1}^{J_n} X_{jrn} P_{jn} \tag{6}$$

In addition, the quadratic form

$$Q(\underline{a}, \underline{\alpha}) = \sum_{i=1}^{M} \sum_{j=1}^{M} (a_i - \alpha_i) A_{ij} (a_j - \alpha_j) \tag{7}$$

tends asymptotically to the chi-square distribution with M degrees of freedom.

Let one of the J_n alternatives available to individual n be considered a numeraire, and denote this alternative by t. Then the random variables $\{\hat{V}_{in} - \hat{V}_{tn}; i=1, \ldots, J_n; i \neq t\}$ are linear combinations of the asymptotically normally distributed random variables $\{a_m\}$ and are themselves asymptotically jointly normally distributed with mean values $\{V_{in} - V_{tn}; i=1, \ldots, J_n; i \neq t\}$ and covariance matrix C_n^{-1}, where

$$(C_n^{-1})_{ij} = \sum_{r=1}^{M} \sum_{s=1}^{M} (A^{-1})_{rs} (X_{irn} - X_{trn})(X_{jsn} - X_{tsn}) \tag{8}$$

and ($i, j=1, \ldots, J_n; i, j \neq t$). In addition the quadratic form

$$R(\hat{\underline{V}}_n, \underline{V}_n) = \sum_{\substack{i=1 \\ i,j \neq t}}^{J_n} \sum_{j=1}^{J_n} (C_n)_{ij} [(\hat{V}_{in} - \hat{V}_{tn}) - (V_{in} - V_{tn})] \times [(\hat{V}_{jn} - \hat{V}_{tn}) - (V_{jn} - V_{tn})] \tag{9}$$

is asymptotically distributed as chi-square with $J_n - 1$ degrees of freedom.

In practical applications of logit models, the probabilities P_{in} and, therefore, the matrix A in Equation 5 are not known due to their dependence on the unknown coefficients $\{\alpha_m\}$. Therefore, P_{in} is approximated by $\hat{P}_{in}$ in Equation 5. This approximation is used without further comment in the rest of this paper.

In the following discussion the subscript n, which denotes the individual, will not be used unless needed to prevent confusion. The choice probabilities will be understood to apply to an individual. The explanatory variables X_{imn} will be assumed to have known, fixed values. All uncertainty in the choice probabilities will be due to their dependence on the unknown coefficients $\{\alpha_m\}$.

Confidence Intervals for Choice Probabilities in Binary Logit Models

If there are only two alternatives in the choice set (J=2), then C^{-1} is a scalar. Therefore, if $Z_{\epsilon/2}$ is the 100 $(1 - \epsilon/2)$ percentile of the standard normal distribution, a $100(1-\epsilon)$ percent confidence interval for $V_1 - V_2$ is

$$(\hat{V}_1 - \hat{V}_2) - Z_{\epsilon/2} C^{-1/2} \leq V_1 - V_2 \leq (\hat{V}_1 - \hat{V}_2) + Z_{\epsilon/2} C^{-1/2} \tag{10}$$

Denote the left- and right-hand expressions of inequalities by b and B, respectively. Then, the expressions for the 100 $(1-\epsilon)$ confidence intervals for P_1 and P_2 in the binary logit model are

$$1/[1 + \exp(-b)] \leq P_1 \leq 1/[1 + \exp(-B)] \tag{11}$$

and

$$1/[1+\exp(B)] \leq P_2 \leq 1/[1+\exp(b)] \qquad (12)$$

These simple expressions for confidence intervals exist only for binary choice models.

Joint Confidence Regions for the Choice Probabilities Based on Asymptotic Sampling Distribution

Equation 4 for the estimated choice probabilities can be rewritten as

$$\hat{P}_i = \exp(\hat{V}_i - \hat{V}_t) \Big/ \Big[1 + \sum_{j \neq t} \exp(\hat{V}_j - \hat{V}_t)\Big] \qquad (i \neq t) \qquad (13)$$

$$\hat{P}_t = 1 - \sum_{j \neq t} \hat{P}_j \qquad (14)$$

where t denotes the numeraire alternative. Equation 13 defines a transformation from the random variables $\{\hat{V}_j - \hat{V}_t;\ j \neq t\}$ to the random variables $\{\hat{P}_i;\ i \neq t\}$. This transformation has a nonsingular Jacobian matrix. Accordingly, the joint probability-density function of the random variables $\{\hat{P}_i;\ i \neq t\}$, conditional on $\hat{P}_t$, can be derived by using standard procedures (8). The result is

$$\begin{aligned} f(\{\hat{P}_i \quad i \neq t\} \mid \hat{P}_t) &= (2\pi)^{(1-n)/2} |C|^{1/2} \Big(\prod_{j=1}^{J} \hat{P}_j\Big)^{-1} \\ &\times \exp\Big\{-(1/2) \sum_i \sum_j C_{ij} [\log(\hat{P}_i/\hat{P}_t) - \log(P_i/P_t)] \\ &\times [\log(\hat{P}_j/\hat{P}_t) - \log(P_j/P_t)]\Big\} \end{aligned} \qquad (15)$$

where $|C|$ denotes the determinant of the matrix C and the quantity on the left-hand side denotes the joint probability-density function of $\{\hat{P}_i;\ i \neq t\}$, conditional on $\hat{P}_t$.

The conditioning of density function 15 on $\hat{P}_t$ can be removed by noting from Equation 14 that $\hat{P}_t$ is completely determined by $\{\hat{P}_i;\ i \neq t\}$. Thus, the joint probability-density function of all of the $\hat{P}_i$ $(i=1, \ldots, J)$ is

$$f(\{\hat{P}_i \quad i = 1, \ldots, J\}) = \delta\Big(1 - \sum_{j=1}^{J} \hat{P}_j\Big) f(\{\hat{P}_i, i \neq t\} \mid \hat{P}_t) \qquad (16)$$

where δ is the Dirac delta function. Equation 16 constitutes a multivariate generalization of the univariate S_B distribution (9). The univariate distribution has been applied in a transportation context by Westin (10), who used the distribution to develop aggregate forecasts from a binary logit model.

The distribution in Equation 16 is highly intractable. To develop a confidence region for P_i it is more convenient to work with the distribution of the logarithms of the choice probabilities than with the distribution of the probabilities themselves. Specifically, Equation 4 implies that

$$\log(\hat{P}_i/\hat{P}_t) - \log(P_i/P_t) = (\hat{V}_i - \hat{V}_t) - (V_i - V_t) \qquad (17)$$

Equations 9 and 17 together imply that the random variable R^* defined by

$$\begin{aligned} R^*(\hat{\underline{P}}, \underline{P}) = \sum_{\substack{i=1 \\ i, j \neq t}}^{J} \sum_{j=1}^{J} & C_{ij} [\log(\hat{P}_i/\hat{P}_t) - \log(P_i/P_t)] \\ & \times [\log(\hat{P}_j/\hat{P}_t) - \log(P_j/P_t)] \end{aligned} \qquad (18)$$

has the chi-square distribution with J-1 degrees of freedom. Let $\chi^2(\epsilon, K)$ denote the 100 $(1-\epsilon)$ percentile of the chi-square distribution with K degrees of freedom. Then, the inequality

$$R^*(\hat{\underline{P}}, \underline{P}) \leq \chi^2(\epsilon, J-1) \qquad (19)$$

together with Equation 14 define a joint 100 $(1-\epsilon)$ percent confidence region for $\{P_i;\ i=1, \ldots, J\}$. Specifically, given estimated values of $\{\hat{P}_i;\ i=1, \ldots, J\}$, the confidence region consists of the set of all P_i $(i=1, \ldots, M)$ such that Equation 14 and inequality 19 are satisfied.

The confidence region defined by Equation 14 and inequality 19 is not rectangular and, therefore, is difficult to use in practical forecasting. In particular, the confidence region does not directly yield constants b_i and B_i $(i=1, \ldots, J)$ such that $b_i \leq P_i \leq B_i$ with a specified level of confidence. However, the confidence region can be used to test hypotheses about the values of the P_i. Let the null hypothesis be $P_1 = P_1^*$, $P_2 = P_2^*$, $\ldots$, $P_J = P_J^*$, and assume that $\Sigma P_i^* = 1$. Substitute P_i^* for P_i in Equation 18 and compute R^*. Then, the null hypothesis is rejected at the ϵ significance level if R^* fails to satisfy inequality 19.

The method used to develop inequality 19 for individual choice probabilities cannot be extended to functions of the choice probabilities, such as aggregate market shares and changes in choice probabilities caused by changes in explanatory variables. The number of utility components $\hat{V}_i - \hat{V}_t$ in such functions exceeds the number of dependent variables (e.g., aggregate shares, changes in choice probabilities) defined by the functions. Therefore, equations such as Equation 17, which define one-to-one mappings between the utility components and the dependent variables, do not exist, and chi-square distributed quadratic forms analogous to R^* cannot be developed. Moreover, the sampling distributions of aggregate shares and changes in choice probabilities contain intractable integrals that prevent these distributions from being used to form confidence regions.

Confidence Regions Based on a Linear Approximation

Equation 4 for the estimated choice probabilities can be expanded in a Taylor series about $\hat{V}_j = V_j$ $(j=1, \ldots, J)$ to obtain

$$\hat{P}_i = P_i + \sum_{j=1}^{J} (\partial P_i/\partial V_j)(V_j - V_j) + \Delta \qquad (i = 1, \ldots, J) \qquad (20)$$

where Δ is a remainder term. As the size of the sample used in estimating the $\hat{V}_j$ approaches infinity, Δ converges in probability to zero and $\hat{P}_i$ converges in probability to (11):

$$\hat{P}_i = P_i + \sum_{j=1}^{J} (\partial P_i/\partial V_j)(\hat{V}_j - V_j) \qquad (21)$$

The random variables $\{\hat{V}_j - V_j\}$ are asymptotically jointly normally distributed with mean values of zero and covariance matrix D^{-1}, where

$$(D^{-1})_{jk} = \sum_{r=1}^{M} \sum_{s=1}^{M} X_{jr} X_{ks} (A^{-1})_{rs} \qquad (22)$$

and A is the matrix defined in Equation 5. Therefore, $\hat{P}_i$ is asymptotically normally distributed with mean value P_i and variance

$$\widehat{\text{var}}(\hat{P}_i) = \sum_{j=1}^{J} \sum_{k=1}^{J} (\partial P_i/\partial V_j)(\partial P_i/\partial V_k)(D^{-1})_{jk} \qquad (i = 1, \ldots, J) \tag{23}$$

It follows that an asymptotic 100(1-ϵ) percent confidence interval for P_i is

$$\hat{P}_i - Z_{\epsilon/2}[\text{var}(\hat{P}_i)]^{1/2} \leq P_i \leq \hat{P}_i + Z_{\epsilon/2}[\text{var}(\hat{P}_i)]^{1/2} \tag{24}$$

where $Z_{\epsilon/2}$ is the 1-ϵ/2 percentile of the standard normal distribution. The numerical value of $\text{var}(\hat{P}_i)$ can be approximated by substituting $\hat{V}$ for V and $\hat{P}$ for P in Equation 23. Equation 21, which is a well-known approximation in mathematical statistics, formed the basis of Koppelman's analysis of errors in disaggregate models (5, 6).

Equation 24 can also be used to develop rectangular joint confidence regions for the P_i. Let I_i be a 100(1-ϵ/J) confidence region for P_i as given by Equation 24. Then

$$\Pr(P_1 \epsilon I_1, P_2 \epsilon I_2, \ldots, P_J \epsilon I_J) \geq 1 - \epsilon \tag{25}$$

Thus $\{P_i; i=1, \ldots, J\}$ is contained in the J-dimensional rectangular region $P_1 \epsilon I_1, \ldots, P_J \epsilon I_J$ and has a confidence level that equals or exceeds 100(1 - ϵ) percent.

The confidence interval defined by inequalities 24 and the joint confidence region defined by inequality 25 can easily be generalized to apply to functions of choice probabilities, including aggregate market shares and changes in choice probabilities caused by changes in explanatory variables. The generalization consists of substituting the functions of interest in place of the choice probabilities in Equations 21-24. The generalization of Equation 23 to aggregate market shares is given by Koppelman (5, 6).

The advantages of the confidence regions defined by inequalities 24 and 25 are substantial: The regions are rectangular, generalizable to functions of the choice probabilities, and computationally tractable. However, because of the regions' reliance on the asymptotic approximation of Equation 21, the accuracy of the confidence levels associated with the regions can vary greatly and may be highly erroneous. This variation in accuracy is illustrated in the following examples.

Consider the univariate, binomial logit model

$$P_i = \exp(\alpha X_i)/[\exp(\alpha X_1) + \exp(\alpha X_2)] \qquad (i = 1, 2) \tag{26}$$

where X_i is the explanatory variable of the model evaluated for alternative i and α is a constant. Let a be the maximum likelihood estimator of α, and let the sampling variance of a be σ^2. Assume that $X_1 = 0$, $X_2 = 0.1$, a = 3, and $\sigma = 1$. Then from inequalities 24, a 95 percent confidence interval for P_1 is $0.378 \leq P_1 \leq 0.474$. The confidence level associated with this interval also can be computed without using approximation 21 by noting that $0.378 \leq P_1 \leq 0.474$ is equivalent to $1.041 \leq \alpha \leq 4.980$. Using the asymptotic normality of the estimated coefficient a, the confidence level associated with $1.041 \leq \alpha \leq 4.980$ and, therefore, with $0.378 \leq P_1 \leq 0.474$ can be computed to be 95.12 percent. Thus, in this example, inequalities 24 yield an accurate estimate of the confidence level.

Now let $X_2 = 1.0$ while X_1, a, and σ remain unchanged. Then inequalities 24 yield $-0.041 \leq P_1 \leq 0.136$ as a 95 percent confidence interval for P_1. If the confidence level associated with this interval is computed directly from the asymptotic distribution of a without using the approximation 21, a confidence level of 87.5 percent is obtained. A true 95 percent confidence interval for P_1 is $0 \leq P_1 \leq 0.205$. Thus, in this case inequalities 24 yield erroneous results.

Nonlinear Programming Approach to Developing Confidence Regions

A method for deriving joint rectangular confidence regions for multinomial logit-choice probabilities without using approximation 21 is described in this section. Denote the vectors of true coefficients ($\alpha_1, \ldots, \alpha_M$) and estimated coefficients ($a_1, \ldots, a_M$) by $\underline{\alpha}$ and $\underline{a}$, respectively. Let $Q(\underline{a}, \underline{\alpha})$ be the quadratic form defined in Equation 7, and let $\chi^2(\epsilon, M)$ be the 100(1 - ϵ) percentile of the chi-square distribution with M degrees of freedom. Recall that P_i (i = 1, ..., J) is a function of $\underline{\alpha}$. Given $\underline{a}$ and ϵ, define $b_i(\epsilon)$ and $B_i(\epsilon)$ for each i by the following nonlinear programming problems:

$$b_i(\epsilon) = \min P_i(\alpha) \qquad (i = 1, \ldots, J) \tag{27}$$

subject to $Q(\underline{a}, \underline{\alpha}) \leq \chi^2(\epsilon, M)$

$$B_i(\epsilon) = \max P_i(\underline{\alpha}) \qquad (i = 1, \ldots, J) \tag{28}$$

subject to $Q(\underline{a}, \alpha) \leq \chi^2(\epsilon, M)$. The maximizations and minimizations are carried out over variations in $\underline{\alpha}$. Then the inequalities

$$b_i(\epsilon) \leq P_i \leq B_i(\epsilon) \qquad (i = 1, \ldots, J) \tag{29}$$

define a rectangular joint confidence region for the P_i with confidence level equal to or greater than 100(1 - ϵ) percent (12).

Another rectangular joint confidence region for the P_i with the same confidence level can be computed by considering P_i to be a function of the utilities ($V_1, \ldots, V_J$). Let $R(\hat{\underline{V}}, \underline{V})$ be the quadratic form defined in Equation 9. Then the solutions to the nonlinear programming problems

$$b_i(\epsilon) = \min P_i(\underline{V}) \qquad (i = 1, \ldots, J) \tag{30}$$

subject to $R(\hat{\underline{V}}, \underline{V}) \leq \chi^2(\epsilon, J-1)$

$$B_i(\epsilon) = \max P_i(\underline{V}) \qquad (i = 1, \ldots, J) \tag{31}$$

subject to $R(\hat{\underline{V}}, \underline{V}) \leq \chi^2(\epsilon, J-1)$ define joint lower and upper confidence limits for P_i with confidence level equal to or greater than 100(1-ϵ) percent. The maximizations and minimizations are performed over variations in $\underline{V}$. The confidence limits thus defined are closer together than the 100(1-ϵ) confidence limits defined by problems 27 and 28 when J-1 < M.

The confidence limits defined by problems 27 and 28 can easily be extended to functions of the choice probabilities. The extension consists of using the relevant functions of the choice probabilities as the objective functions of problems 27 and 28. For example, if P_{in} is the probability that individual n chooses alternative i, the aggregate market share of alternative i in a population of N individuals is

$$\Pi_i = (1/N) \sum_{n=1}^{N} P_{in} \qquad (i = 1, \ldots, J) \tag{32}$$

Π_i is a function of α through the P_{in}. Joint confidence limits b_i and B_i for the Π_i with confidence level equal to at least 100(1-ϵ) percent are given by

$$b_i(\epsilon) = \min \Pi_i(\underline{\alpha}) \qquad (i = 1, \ldots, J) \tag{33}$$

subject to $Q(\underline{a}, \underline{\alpha}) \leq \chi^2(\epsilon, M)$

$$B_i(\epsilon) = \max \Pi_i(\underline{\alpha}) \qquad (i = 1, \ldots, J) \tag{34}$$

subject to $Q(a, \alpha) \leq \chi^2(\epsilon, M)$.

The joint rectangular confidence region that results from the asymptotic linear approximation (inequalities 24 and 25) can be obtained by solving the nonlinear programming problems

$$b_i^*(\epsilon) = \min P_i^*(\underline{V}) \qquad (i = 1, \ldots, J) \tag{35}$$

subject to $R(\hat{\underline{V}}, \underline{V}) \leq \chi^2(\epsilon/J, 1)$

$$B_i^*(\epsilon) = \max P_i^*(\underline{V}) \qquad (i = 1, \ldots, J) \tag{36}$$

subject to $R(\hat{\underline{V}}, \underline{V}) \leq \chi^2(\epsilon/J, 1)$, where $P_i^*(\underline{V})$ is the expression obtained by exchanging P_i with $\hat{P}_i$ and $\hat{V}_j$ with V_j in Equation 21, and b_i^* and B_i^*, respectively, are the lower and upper confidence limits for P_i obtained by the linear approximation method. As the accuracy of the asymptotic linear approximation increases, problems 35 and 36 approach equivalence with the problems

$$b_i^*(\epsilon) = \min P_i(\underline{V}) \qquad (i = 1, \ldots, J) \tag{37}$$

subject to $R(\hat{\underline{V}}, \underline{V}) \leq \chi^2(\epsilon/J, 1)$

$$B_i^*(\epsilon) = \max P_i(\underline{V}) \qquad (i = 1, \ldots, J) \tag{38}$$

subject to $R(\hat{\underline{V}}, \underline{V}) \leq \chi^2(\epsilon/J, 1)$. Problems 37 and 38 differ from problems 30 and 31 only in the right-hand sides of their constraints. Comparison of problems 37 and 38 with problems 30 and 31 and problems 27 and 28 provides a means of determining whether the asymptotic linear approximation or the nonlinear programming method yields a smaller joint confidence region for the choice probabilities when the linear approximation is accurate. If $J-1<M$, the linear approximation yields narrower confidence limits for each of the P_i whenever

$$\chi^2(\epsilon/J, 1) < \chi^2(\epsilon, J-1) \tag{39}$$

If $M \leq J-1$, the linear approximation yields narrower limits whenever

$$\chi^2(\epsilon/J, 1) < \chi^2(\epsilon, M) \tag{40}$$

Conditions 39 and 40 will be satisfied at normal confidence levels unless the number of coefficients M is very small or the number of alternatives J is either two or very large. For example, if $M = 4$ and $\epsilon = 0.05$, conditions 39 and 40 will be satisfied if $3 \leq J \leq 24$. If $M = 5$ and $\epsilon = 0.05$, the conditions will be satisfied if $3 \leq J \leq 61$. Thus, when the asymptotic linear approximation is accurate it will tend to produce smaller joint confidence regions than will the nonlinear programming method unless the choice set either is large or contains only two alternatives. Numerical illustrations of the differences in the sizes of the linear approximation and nonlinear programming confidence regions are given in a later section.

A BOUND ON THE ERROR IN THE CONFIDENCE LEVEL

The errors in the linear approximation confidence levels of a binary choice model were previously computed exactly. This exact computation is not possible for models that have more than two alternatives in their choice sets. In multinomial models, nonlinear programming can be used to establish upper bounds on the errors in the confidence levels obtained from inequalities 24.

Let P_i^* be defined as in problems 35 and 36, and let σ^* be the linear approximation estimate of the standard deviation of $\hat{P}_i$ obtained from Equation 23. Note that P_i^* depends on the true coefficients α through $\underline{V}$. For arbitrary positive K and k define the following sets:

$$S_1(K) = \{\underline{\alpha} \mid |P_i^* - \hat{P}_i| \leq K\} \tag{41}$$

$$S_2(K) = \{\underline{\alpha} \mid |\hat{P}_i - P_i| \leq K\} \tag{42}$$

$$S_3(K) = \{\underline{\alpha} \mid |P_i^* - P_i| \leq K\} \tag{43}$$

$$S_4(K, k) = \{\underline{\alpha} \mid |P_i^* - \hat{P}_i| \leq K - k\} \tag{44}$$

$$S_5(K, k) = \{\underline{\alpha} \mid |P_i^* - \hat{P}_i| \leq K + k\} \tag{45}$$

The sets S_1 through S_5 all depend on the estimated coefficients $\underline{a}$ and, therefore, are random events. Let $Pr(S_j)$ be the probability of the event S_j ($j = 1, \ldots, 5$). Note that

$$S_2 \cap S_3 \subset S_5 \tag{46}$$

and

$$Pr(S_2 \cap S_3) \geq Pr(S_2) + Pr(S_3) - 1 \tag{47}$$

Therefore,

$$Pr(S_2) \leq Pr(S_5) + [1 - Pr(S_3)] \tag{48}$$

Also,

$$S_4 \cap S_3 \subset S_2 \tag{49}$$

and

$$Pr(S_4 \cap S_3) \geq Pr(S_4) + Pr(S_3) - 1 \tag{50}$$

Therefore,

$$Pr(S_2) \geq Pr(S_4) - [1 - Pr(S_3)] \tag{51}$$

when probabilities 48 and 51 are combined,

$$Pr(S_4) - [1 - Pr(S_3)] \leq Pr(S_2) \leq Pr(S_5) + [1 - Pr(S_3)] \tag{52}$$

$P_i^* - \hat{P}_i$ is asymptotically normally distributed with mean zero and standard deviation σ^*, by virtue of Equation 21. Let Φ denote the cumulative standard normal distribution function. Then asymptotically

$$Pr(S_4) = 2\,\Phi\,[(K - k)/\sigma^*] - 1 \tag{53}$$

$$Pr(S_5) = 2\,\Phi\,[(K + k)/\sigma^*] - 1 \tag{54}$$

$$Pr(S_1) = 2\,\Phi\,(K/\sigma^*) - 1 \tag{55}$$

Inequality 52 and Equations 53-55 imply

$$\begin{aligned} 2\{\Phi[(K-k)/\sigma^*] - \Phi(K/\sigma^*)\} - [1 - Pr(S_3)] &\leq Pr(S_2) - Pr(S_1) \\ &\leq 2\{\Phi[(K+k)/\sigma^*] - \Phi(K/\sigma^*)\} + [1 - Pr(S_3)] \end{aligned} \tag{56}$$

Given a confidence level $100(1-\epsilon)$ percent, let K be given by the solution to

$$Pr[S_1(K)] = 1 - \epsilon \tag{57}$$

Note that in the linear approximation method for developing confidence intervals P_i^* and P_i are considered to be equal. Therefore, $100(1-\epsilon)$ is the confidence level that the linear approximation assigns to the interval $|P_i - \hat{P}_i| \leq K$, whereas $100\,Pr[S_2(K)]$ is the confidence level that is obtained if the linear approximation is not used. Thus, $100[Pr(S_2) - Pr(S_1)]$ is the error in the confidence level that is made by using the linear approximation, and inequalities 56 bound this error. Specifically, for any k

$$| Pr(S_2) - Pr(S_1) | \leq \max[F^+(K, k), F^-(K, k)] \quad (58)$$

where

$$F^+(K, k) = 2\{\Phi[(K + k)/\sigma^*] - \Phi(K/\sigma^*)\} + [1 - Pr(S_3)] \quad (59)$$

and

$$F^-(K, k) = -2\{\Phi[(K - k)/\sigma^*] - \Phi(K/\sigma^*)\} + [1 - Pr(S_3)] \quad (60)$$

In practice it is usually difficult or impossible to evaluate $Pr(S_3)$. Thus, inequality 58 is not directly useful. However, it is possible to establish a computationally tractable lower bound on $Pr(S_3)$. Given a number δ that satisfies $0 < \delta < 1$, define $k(\delta)$ by the following nonlinear programming problem:

$$k(\delta) = \max | P_i^*(\underline{\alpha}) - P_i(\underline{\alpha}) | \quad (61)$$

subject to $Q(\underline{a}, \underline{\alpha}) \leq \chi^2(\delta, M)$, if $M \leq J-1$ and $R(\hat{\underline{V}}, \underline{V}) \leq \chi^2(\delta, J-1)$ otherwise

$$Pr[| P_i^* - P_i | \leq k(\delta)] \geq 1 - \delta \quad (62)$$

and

$$Pr\{S_3[k(\delta)]\} \geq 1 - \delta \quad (63)$$

Given K and δ, define $G^+(K, \delta)$ and $G^-(K, \delta)$ by

$$G^+(K, \delta) = 2(\Phi\{[K + k(\delta)]/\sigma^*\} - \Phi(K/\sigma^*)) + \delta \quad (64)$$

Table 1. Joint 95 percent confidence intervals for the choice probabilities in a three-alternative mode choice model.

		Linear Approximation Method		Nonlinear Programming Method		
Alternative	P	b	B	b	B	R
1	0.402	0.338	0.467	0.338	0.470	1.02
2	0.312	0.262	0.362	0.263	0.366	1.02
3	0.286	0.234	0.337	0.236	0.341	1.02

Note: P = estimated choice probability; b = lower confidence limit; B = upper confidence limit; and R = width of nonlinear programming confidence interval divided by width of linear approximation interval.

Table 2. Joint 95 percent confidence intervals for the choice probabilities in a 20-alternatives destination choice model.

		Linear Approximation Method		Nonlinear Programming Method		
Alternative	P	b	B	b	B	R
1	0.022	0.017	0.027	0.017	0.028	1.10
2	0.029	0.023	0.035	0.023	0.037	1.10
3	0.017	0.011	0.022	0.012	0.023	1.11
4	0.035	0.027	0.043	0.026	0.044	1.10
5	0.024	0.013	0.035	0.015	0.039	1.13
6	0.034	0.029	0.039	0.029	0.040	1.09
7	0.056	0.039	0.073	0.040	0.078	1.11
8	0.036	0.030	0.042	0.030	0.043	1.10
9	0.025	0.020	0.031	0.020	0.032	1.10
10	0.049	0.041	0.057	0.040	0.058	1.09
11	0.111	0.075	0.147	0.077	0.157	1.11
12	0.083	0.075	0.091	0.074	0.092	1.10
13	0.089	0.066	0.112	0.066	0.117	1.10
14	0.066	0.056	0.076	0.056	0.078	1.10
15	0.080	0.069	0.090	0.069	0.091	1.10
16	0.077	0.064	0.089	0.064	0.091	1.10
17	0.018	0.010	0.025	0.011	0.028	1.13
18	0.063	0.052	0.074	0.052	0.075	1.10
19	0.044	0.033	0.056	0.033	0.058	1.10
20	0.043	0.037	0.049	0.037	0.050	1.10

Note: P = estimated choice probability; b = lower confidence limit; B = upper confidence limit; and R = width of nonlinear programming confidence interval divided by width of linear approximation interval.

$$G^-(K, \delta) = -2(\Phi\{[K - k(\delta)]/\sigma^*\} - \Phi(K/\sigma^*)) + \delta \quad (65)$$

Then

$$G^+(K, \delta) \geq F^+[K, k(\delta)] \quad (66)$$

$$G^-(K, \delta) \geq F^-[K, k(\delta)] \quad (67)$$

and

$$| Pr(S_2) - Pr(S_1) | \leq \min_{\delta}\{\max[G^+(K, \delta), G^-(K, \delta)]\} \quad (68)$$

Inequality 68 defines a computationally tractable upper bound on the error in the confidence level obtained from inequalities 24.

The degree to which the right-hand side of inequality 68 overestimates the error made by linear approximation 21 can be illustrated with the model of Equation 26. It was shown that when $X_1 = 0$, $X_2 = 0.1$, $a = 3$, and $\sigma = 1$ in Equation 26, the linear approximation assigns a confidence limit of 95 percent to a particular confidence interval for the coefficient α, whereas a confidence level of 95.12 percent is obtained for the same interval when the linear approximation is not used. In this case the linear approximation makes an error of 0.12 percent in the confidence level. When $X_2 = 1.0$ and the other parameters remain unchanged, the linear approximation assigns a confidence level of 95 percent to an interval whose confidence level is found to be 87.5 percent when the linear approximation is not used. In this case, the linear approximation makes an error of 7.5 percent in the confidence level. Inequality 68 gives an upper bound on the error in the confidence level of 1.2 percent when $X_2 = 0.1$ and 31 percent when $X_2 = 1.0$. Although inequality 68 considerably overestimates the error made by the linear approximation in both cases, the error estimates obtained from inequality 68 do distinguish between a case in which the linear approximation is useful (e.g., $X_2 = 0.1$), and a case in which the linear approximation is not useful (e.g., $X_2 = 1.0$).

Inequality 68 can be extended to functions of the choice probabilities, such as aggregate market shares. The extension is accomplished by substituting the desired functions in place of $\hat{P}_i$, P_i, and P_i^* in equations and inequalities 41-68 and by using the Q form of the constraint in problem 61.

NUMERICAL EXAMPLES

To illustrate and compare the linear approximation and nonlinear programming methods for developing confidence regions, both methods were applied to two multinomial logit models: a 3-alternative model of work-trip mode choice (5) and a 20-alternative model of destination choice for nonwork trips (13). Typical values of the explanatory variables were used in each case. The nonlinear programming problems 27, 28, 31, 32, and 61 were solved by using the sequential unconstrained minimization technique (14).

Joint 95 percent confidence limits for the choice probabilities of the mode choice model are shown in Table 1. The upper and lower confidence limits of the choice probabilities are, respectively, approximately 17 percent above and below these probabilities. The nonlinear programming confidence intervals were obtained from problems 30 and 31 and are approximately 2 percent wider than the linear approximation intervals. Inequality 68 indicates that the errors in the confidence levels of the linear approximation confidence intervals considered individually are less than 1.14 percent. Considering the looseness of the bound provided by inequality 68, this suggests that the linear approximation achieves acceptable accuracy in this example.

Joint 95 percent confidence limits for the choice probabilities of the destination choice model are shown in Table 2. The upper and lower confidence limits of the choice probabilities are, respectively, roughly 10 to 40 percent above and below these probabilities, depending on the alternative. The nonlinear programming confidence intervals are approximately 10 percent wider than the linear approximation intervals. Inequality 68 indicates that the errors in the confidence levels of the linear approximation confidence intervals considered individually are less than 0.8 percent, again suggesting that the linear approximation is acceptably accurate.

CONCLUSIONS

This paper has described three methods for developing confidence regions for the choice probabilities of the multinomial logit model. One method involves a direct application of the asymptotic sampling distribution of the choice probabilities and yields joint confidence regions for these probabilities. The confidence regions are not rectangular and, therefore, are useful mainly for testing hypotheses about the choice probabilities.

The other two methods are based, respectively, on a linear approximation of the relation between errors in the coefficients of a model and errors in the choice probabilities, and on a nonlinear programming approach to developing confidence intervals. Both of these methods produce joint rectangular confidence regions for the choice probabilities, and both can be applied to functions of the choice probabilities, such as aggregate market shares and changes in choice probabilities caused by changes in explanatory variables. The linear approximation method also can be used to develop confidence intervals for individual choice probabilities.

The linear approximation method is computationally simpler than the nonlinear programming method. Moreover, when the linear approximation on which the method is based is accurate, the linear approximation method produces a smaller confidence region for a given confidence level than does the nonlinear programming method, unless the choice set either is very large or contains only two alternatives. However, the linear approximation method has the disadvantage that it can yield erroneous results.

A procedure for bounding the error made by the linear approximation method has been described in this paper. However, this procedure is based on nonlinear programming, and the computational effort involved in implementing it can equal or exceed the computational effort involved in developing confidence regions by the nonlinear programming method. If there are a priori reasons for believing that the linear approximation method will yield accurate results in a particular application, then the computational simplicity of this method makes it preferable to the nonlinear programming method. However, if the accuracy of the linear approximation method is questionable and resources for implementing the bounding procedure are not available, then the nonlinear programming method will yield more reliable results than will the linear approximation method.

The linear approximation and nonlinear programming methods for developing confidence regions can be applied to other utility maximizing models with linear-in-parameters utility functions (e.g., multinomial probit) by substituting the choice probabilities of the desired model in place of the logit probabilities used in this paper.

ACKNOWLEDGMENT

Thanks are due to Garth McCormick for providing a computer program that implements the sequential unconstrained minimization technique. The views expressed in this paper are mine and are not necessarily endorsed by the U.S. Environmental Protection Agency.

REFERENCES

1. T. A. Domencich and D. McFadden. Urban Travel Demand. North-Holland-Elsevier, New York, 1975.
2. B. D. Spear. Applications of New Travel Demand Forecasting Techniques to Transportation Planning. U.S. Department of Transportation, 1977.
3. M. Ben-Akiva and T. J. Atherton. Methodology for Short-Range Travel Demand Predictions. Journal of Transport Economics and Policy, Vol. 11, 1977, pp. 224-261.
4. H. Theil. Principles of Econometrics. Wiley, New York, 1971.
5. F. S. Koppelman. Travel Prediction with Models of Individual Choice Behavior. Center for Transportation Studies, Massachusetts Institute of Technology, Cambridge, MA, Rept. CTS 75-7, 1975.
6. F. S. Koppelman. Methodology for Analyzing Errors in Prediction with Disaggregate Choice Models. TRB, Transportation Research Record 592, 1976, pp. 17-23.
7. D. McFadden. Analysis of Qualitative Choice Behavior. In Frontiers in Econometrics (P. Zarembka, ed.), Academic Press, New York, 1974, pp. 105-142.
8. M. Fisz. Probability Theory and Mathematical Statistics. Wiley, New York, 1963.
9. N. Johnson. Systems of Frequency Curves Generated by Methods of Translation. Biometrika, Vol. 36, 1949, pp. 149-176.
10. R. B. Westin. Predictions from Binary Choice Models. Journal of Econometrics, Vol. 2, 1974, pp. 1-16.
11. H. Chernoff. Large Sample Theory: Parametric Case. Annals of Mathematical Statistics, Vol. 27, 1956, pp. 1-22.
12. C. R. Rao. Linear Statistical Inference and Its Applications. 2d ed. Wiley, New York, 1973.
13. J. Horowitz. Disaggregate Model for Nonwork Travel. TRB, Transportation Research Record 673, 1978, pp. 65-71.
14. A. V. Fiacco and G. P. McCormick. Nonlinear Programming: Sequential Unconstrained Minimization Techniques. Wiley, New York, 1968.

Publication of this paper sponsored by Committee on Passenger Travel Demand Forecasting.

Discrete Multivariate Model of Work-Trip Mode Choice

David Segal, Department of Economics, Oberlin College, Oberlin, Ohio, and Department of City and Regional Planning, Harvard University, Cambridge, Massachusetts

This paper applies discrete multivariate analysis to the specification and estimation of factors that govern work-trip mode choice. Where large data sets are available, this technique is found to have two important advantages over conditional logit analysis: Better model specification is facilitated and parameters can typically be estimated at sharply lower cost. The study focuses on the mode-choice behavior of 9880 Washington, D.C., area households that made work trips in 1968. Perhaps the most striking result is that in-vehicle travel time seems to have a nonlinear impact on the mode-choice logit (log-odds of drive alone versus bus), which has potentially important consequences for policy. For the range in which bus is faster than automobile, changes in bus (or automobile) in-vehicle travel time have the well-known results reported by other studies. But for the interval within which driving is faster than bus, decreases in bus in-vehicle travel time that fall short of making the bus mode absolutely faster than driving will, if our estimates are correct, fail to increase ridership significantly.

This paper examines urban travel mode-choice behavior by using a discrete multivariate technique. The purpose of the exercise is twofold. First, the technique itself is shown to have some advantages over conditional logit analysis when large data sets are available (1-3). It allows for more careful analysis of model specification and for this reason it often provides better goodness of fit. And the estimation algorithms on which it draws are simpler and typically cheaper to use than those of conditional logit analysis.

A second aim of the paper is substantive. Empirical models of travel mode choice within cities have generally assumed all of the explanatory variables to be linear in their impact on the mode-choice logit (log-odds). Some confirmation of this assumption is provided here; however, there are important exceptions. One is in-vehicle travel time, which is found to be nonlinear in a major respect: Although the log-odds of transit travel increase as transit becomes faster than automobile, the converse is found not to be true. The log-odds of driving do not increase significantly over the range in which driving in-vehicle travel time is less than travel time by bus. This result may be significant for policy, suggesting that efforts to improve transit speed that fall short of making transit absolutely faster than driving may have a negligible impact on ridership.

The paper considers the choice among three modes (driving, bus, and automobile passenger) of a sample of 9880 households that reported work trips in the Washington, D.C., area in 1968. The data were collected in a home-interview survey of some 25 000 households in that area conducted by the Washington Metropolitan Area Council of Governments. They have been used extensively in published studies of the determinants of mode-choice behavior, as well as in developmental studies of the estimation of disaggregate travel demand by using the conditional logit technique (4). The data set used here was the home-interview survey, augmented by engineering level-of-service data provided by R. H. Pratt Associates. A file that merges the two data sets was prepared by Cambridge Systematics, Inc. It is known as the Second Auto Ownership Project Master File and is used in the analysis of this paper. Although the data for these studies and the one here have a common origin, our approach and estimation technique are discrete multivariate.

There is a basic equivalence between discrete multivariate analysis (DMA) and the logit model. DMA may be likened to a logit model in which all of the variables are categorical. The value taken by a given observation on any particular variable is then represented by a dummy that is equal to one if the value falls within a prescribed level of the variable and zero otherwise. An observation in the logit formulation is thus a vector of ones and zeros; the frequencies with which that observation occurs in the data set may be viewed as a weight. In the DMA formulation, a data set is aggregated into a multidimensional array of counts. Each dimension of the array is a variable, and each variable in turn has a specified number of categories or levels.

Different algorithms are used to estimate parameters by the two forms of analysis. The logit model has relied on the Newton-Raphson estimating procedure, which can be costly when large numbers of parameters must be estimated from large data sets (5, p. 48; 6, p. 122). Because the algorithm requires that a matrix of second-order partial differentials be inverted at every iteration in the estimation procedure, and because the number of computations needed to invert grows as the square of the number of parameters to be estimated, analysts who use conditional logit analysis have a cost incentive to economize on parameters. Therefore, analysts have tended to limit the number of parameters to be estimated and to draw samples from large data sets. Many of the studies that use the Washington data have been based on 10 percent samples.

An economy of parameters, however, usually makes sense only if there is an economy of data. For the Washington data set this is clearly not the case. DMA uses all of the data, as noted earlier, in the form of cross-classified tabulations of counts. Some information may be lost in the categorization procedure; however, Aigner, Goldberger, and Kalton (7) have shown that, (a) for certain underlying specifications in which an explanatory variable is uniformly, normally, or exponentially distributed and (b) where four or five categories are developed for such a variable, only 10 percent of the information in the data set is lost (contrasted with as much as 90 percent or more in some of the studies sampling from the Washington data).

Not only can the underlying structure of a model be more readily revealed by DMA, it can be done at relatively low cost. Models that have as many as 100 or more parameters and tens of thousands of observations can currently be estimated at a marginal cost of as low as $0.75 at any large computing center by using one of the cyclic ascent algorithms. As suggested earlier, the Newton-Raphson technique can also be applied to such models at a somewhat greater expense if the data are first cross-classified and if the cell frequencies are then employed as weights by using a choice-based sampling approach that views the cells themselves as dummy variables [see Manski and Lerman (8)]. To be sure, the cost advantages of DMA are obscured in situations in which data collection is expensive. It is ideally suited for analysis of some forms of census data.

The balance of this paper considers the model employed here to estimate mode-choice logit coefficients, the data set, estimation procedure, and results.

THE MODEL

Let us begin with the question of mode-choice probabilities between automobile and bus. Later the choice set is broadened to include traveling as an automobile passenger.

Let P_D = probability (mode = drive) be proportional to $F(X_1\beta)$ where $F(X_1\beta)$ is a function that describes how the probability of driving is related to a set of explanatory variables (X_1) that includes both automobile level-of-service and socioeconomic variables. With binary mode choice, the probability of going to work by bus is P_B = probability (mode = bus) = $F(X_2\beta) = 1-F(X_1\beta)$. Any one of several functional forms can be used to relate the probability measures to $X_1\beta$ or $X_2\beta$; however, we follow the work cited earlier and employ the logistic distribution. We thus define

$$P_D = \exp(X_1\beta)/[\exp(X_1\beta) + \exp(X_2\beta)] \tag{1}$$

Multiplying both the numerator and denominator by exp $(-X_1\beta)$ gives

$$P_D = 1/\{1 + \exp[-(X_1 - X_2)\beta]\} \tag{1a}$$

Let $X = X_1 - X_2$ be the differences in the mode characteristics for driving and taking the bus that affect the choice probabilities. Clearly the socioeconomic variable levels, which for a given household are invariant with mode choice, will cancel out if included in both X_1 and X_2. One of the ways that conditional logit estimating procedure gets around this potential difficulty is by including such variables in only one of the level-of-service attribute vectors, say X_1. This parallels the procedure followed here in which the socioeconomic variables enter the analysis only after the differences for the level-of-service attributes have been computed. That is, X in Equation 3 has as its component elements (a) level-of-service differences between a pair of modes and (b) socioeconomic levels.

The probability measure ranges from zero to one as $X\beta$ goes from $-\infty$ to $+\infty$. If, as we assume, Equation 1 represents the probability of driving, then the probability of taking the bus to work becomes

$$P_B = (1 - P_D) = \exp(-X\beta)/[1 + \exp(-X\beta)] = 1/[1 + \exp(X\beta)] \tag{2}$$

Rearranging Equations 1 and 2 gives

$$\begin{aligned} L &= \log(P_D/P_B) = \log P_D - \log(1 - P_D) \\ &= -\log[1 + \exp(-X\beta)] - \{\log[\exp(-X\beta)] - \log[1 + \exp(-X\beta)]\} \\ &= X\beta \end{aligned} \tag{3}$$

where L is the logit or log of the ratio P_D/P_B, which reports the odds of driving relative to taking the bus and where, as noted earlier, $X = X_1 - X_2$. The properties of the logistic function and its advantages in studying the determinants of disaggregate travel demand have been widely reported (9, 10). In the literature the logit is shown to depend on the mean utility of a given alternative. It is assumed that individual utility deviations from mean utility in a homogeneous market segment are statistically independent for different alternatives. The logit is thus governed by a stochastic utility function, which in Equation 3 is represented without the error term. A function whose arguments are linearly additive is usually specified, as in Equation 3; X represents the utility of a set of differences between the mode characteristics of automobile and those of bus as it is evaluated by a household of a given socioeconomic stratum, and β is a vector of weights that must be estimated.

The conditional logit model assumes the independence of irrelevant alternatives (IIA). This means that, if a third mode is introduced into the analysis (for example, automobile passenger with utility function arguments X_3) the log-odds of driving versus being a passenger in an automobile are reported by $(X_1 - X_3)\beta$. The parameters in the vector β are unaffected by the introduction of a new model.

The conditional logit model has been widely used to estimate the coefficients of Equation 3; however, the discrete multivariate model has not. It should not be surprising, therefore, that relatively little attention has been given to nonlinearities in, and interaction among, the explanatory variables in empirical mode-choice analysis.

DATA AND ESTIMATION TECHNIQUE

Often the data that are available to the analyst are already categorized, as in the case of census data or, in this study, the mode-choice and income data. The data that are reported as continuous, such as travel cost and time, must be grouped. The data are fashioned as a multiway table of counts or frequencies. Such a table may often have as many as five or six dimensions, one for each variable. The category levels of each of the variables become the rows (or columns) of faces of the table. The multiway table that is thus formed is the basic input in discrete multivariate analysis.

The estimation procedure is a straightforward generalization of analysis of variance (ANOVA) techniques and involves finding a set of cell frequencies that fits a specified set of marginal configurations (2, chapter 3). Specification of a particular set of configurations to be fit is equivalent to specification of a model—which interactions, if any, matter. The set of possible models is hierarchical, meaning that to specify an interaction effect of a given order among a group of variables is to specify, simultaneously, all lower-order interaction effects. For a given model, maximum likelihood estimates (of the original cell frequencies) are obtained by taking an initial set of frequencies (often ones) and iterating them through a sequence of cycles that brings configurations of them successfully closer to the configuration totals specified by the model. The Deming-Stephan iterative proportional fitting algorithm is used here to obtain maximum likelihood estimates (MLEs) in this fashion (2, p. 84).

Once MLEs have been estimated, along with the appropriate goodness-of-fit statistic, it is possible to estimate the logit coefficients (β). Their interpretation is as follows: Each of the elements of x is a dummy variable that equals one if the observation falls within the level of the variable in question and zero otherwise. For any given variable, the β coefficients, which hereafter are termed w coefficients following the notation of the discrete multivariate literature, sum to zero:

$$\sum_i w_{1(i)} = \sum_j w_{2(j)} = \sum_k w_{3(k)} = 0 \tag{4a}$$

$$\begin{aligned} \sum_i w_{12(ij)} &= \sum_j w_{12(ij)} = \sum_i w_{13(ik)} = \sum_k w_{13(ik)} \\ &= \sum_j w_{23(jk)} = \sum_k w_{23(jk)} = 0 \end{aligned} \tag{4b}$$

$$\sum_i w_{123(ijk)} = \sum_j w_{123(ijk)} = \sum_k w_{123(ijk)} = 0 \tag{4c}$$

Table 1. Logit parameter estimates for model number 16, driving versus bus.

Variable		Level	
Number	Name	Number	Description
1	Mode choice	1	Drive
		2	Bus passenger
2	In-vehicle travel time	1	Driving time exceeds bus time by 0-60 min (26.3 min average)
		2	Bus time exceeds driving time by 0-15 min (7.16 min average)
		3	Bus time exceeds driving time by more than 15 min (29.6 min average)
3	Out-of-vehicle travel time	1	Driving relatively slower—range from 20 min (or more) slower to 5 min faster (average of 2.05 min slower)
		2	Driving faster by 5-20 min (11.9 min average)
		3	Driving faster by more than 20 min (32.9 min average)
4	Travel cost	1	Bus cheaper by \$0.50-\$5.00 (\$1.20 average)
		2	Bus cheaper by \$0.0-\$0.50 (\$0.23 average)
		3	Automobile cheaper (\$0.24 average)
5	Household income	1	\$0-\$6000
		2	\$6000-\$15 000
		3	\$15 000 or more

The w-terms (or their u-term counterparts in log-linear models) are often omitted in analyses outside economics, where the greater interest lies mainly in issues of model specification per se and goodness of fit. Their virtue is that they are analogous to, and can be compared directly with, β coefficients of the conditional logit model. They report the partial effects on the logit of a particular level of a given variable, all other things being equal.

The models tested in this paper are based on five-way tables whose dimensions report one response variable (mode choice) and four explanatory variables (in-vehicle travel time, out-of-vehicle travel time, out-of-pocket travel cost, and household income). The first three explanatory variables are level-of-service variables; the fourth is a socioeconomic variable. In addition, the data were stratified by the number of automobiles available per worker in the household: zero, one, or two or more. The results presented in the next section are for the middle category of automobile availability, which is by far the best represented in the sample. (In subsequent analyses, the automobile-availability variable will be included both as an explanatory variable and as a response variable, determined simultaneously with mode choice. Here the effort was directed at keeping the analyses relatively simple. The logit coefficients reported in Table 1 are thus conditioned on the availability of one automobile per worker in mode-choosing households.)

In the analysis we set conditions on the values of the explanatory variables. This stems from the fact that we are not interested in the factors that govern the margin totals for these variables (or interactions among them) or in the totals. Rather, we are concerned with the factors that govern mode choice. (As we shall see in Table 1 and the accompanying text, this means that we must fit all models under consideration to the margins that reflect full interaction among the explanatory variables. Doing this effectively adjusts for the effects of variation in frequencies across the categories of margin totals for the explanatory variables.)

For the binary-choice model of Equation 3 each of the four explanatory variables is assigned three levels. A complete description of the levels for each of the variables is provided in Table 2. The five-way table that emerges is dimensioned 2x3x3x3x3 and thus has 162 cells. The table described here is complete. Had the automobile-availability variable been included as an explanatory, level-of-service variable, the resulting table would have been incomplete because of the logical impossibility of a household owning no automobiles and opting for the drive mode. (A table of similar dimensions is fit for the drive-or-passenger choice. The two tables have a combination of 243 cell frequencies.)

As a rule, the category limits for the different explanatory variables were set in such a way as to distribute frequencies more or less evenly across categories. The data that were used to establish category limits were, for the level-of-service variables, the differences between the automobile (drive alone) and bus characteristics, as reported in the Washington survey (4). Only work trips were included in the analysis. With the exception of the automobile-passenger data, all level-of-service data were taken exactly as they appeared in the data file. In-vehicle travel time was adjusted for car-pool passengers by adding 10 min to the drive-alone time. Travel costs for passengers were adjusted downward on a pro rata basis under the assumption that car-pools carry an average of 2.5 occupants.

THE RESULTS

Model Selection

A model is selected by evaluating alternative combinations of interactions between the response variable (designated as 1 in Table 3) and combinations of the explanatory variables, denoted as 2, 3, 4, and 5. Theory suggests which variables belong in the model, and a set of goodness-of-fit statistics for alternative models is used to designate an appropriate specification for the variables.

The number of possible models that fit the 2345 margins and that include combinations of these variables interacted with the mode-choice variable is well over 100. A subset of 20 such models is shown in Table 3. (The analyses reported in Tables 2 and 3 involve just such a dichotomous mode choice—between driving to work and taking the bus. There are 8513 observations on these two modes, of which 948 are bus riders and 7565 drive to work alone.)

In the selection of a model there are three, roughly equivalent, ways to proceed. A first method of approach involves beginning with just the independent effects of the explanatory variables, interacted with the response variable. The decision rule that is followed involves the inclusion of higher-order interactions if and only if the expenditure of degrees of freedom can be justified through improved goodness of fit. When such a gain is no longer possible or when undue complexity would be introduced into a model by specifying further interaction, the procedure is brought to a stop. The model that cannot be improved on without violating the decision rule is chosen as best.

Table 2. Partial effects on logit (log-odds of driving when bus is the alternative).

Variable	Household Income ($000s)	Bus Faster	Automobile Slightly Faster	Much Faster
In-vehicle travel time	0-6	-0.348	0.114	0.234
	6-15	-0.554	0.704	-0.150
	15+	-0.490	0.153	0.337
Out-of-vehicle travel time	0-6	0.056	-0.384	0.328
	6-15	-1.014	-0.308	1.320
	15+	-0.833	-0.298	1.131
Cost	0-6	-0.599	-0.190	0.789
	6-15	-0.780	0.202	0.578
	15+	-0.780	0.214	0.566

Table 3. Choice of a model.

Model Number	Model	df	G^{2a}	p^b
1	2345 12 13 14	74	129.32	-
2	2345 12 13 14 15	72	129.23	-
3	2345 134 12	70	121.92	-
4	2345 124 13	70	115.04	0.001
5	2345 123 14	70	107.72	0.003
6	2345 123 14 15	68	107.56	0.002
7	2345 125 145	66	364.68	-
8	2345 125 135	66	276.84	-
9	2345 135 145	66	232.16	-
10	2345 123 124 134	62	86.06	0.023
11	2345 125 135 145	60	89.63	0.008
12	2345 1235	54	252.16	-
13	2345 1245	54	340.87	-
14	2345 1345 125	48	74.88	0.008
15	2345 1245 135	48	63.96	0.064
16	2345 1235 145	48	62.46	0.078
17	2345 1234 125 135 145	40	37.91	0.500+
18	2345 1345 1245	36	49.23	0.070
19	2345 1234 1245	36	46.29	0.117
20	2345 1235 1245	36	42.21	0.22

Note: 1 = Mode choice, 2 = In-vehicle travel time, 3 = Out-of-vehicle travel time, 4 = Cost, and 5 = Income.
[a] G^2 is a goodness-of-fit statistic, defined as $G^2 = 2\Sigma$ (Observed) log(Observed/Expected).
[b] The probability measure (p) has the following interpretation. If a given model is correct as specified, then there is a probability that the χ^2 statistic associated with it (G^2) is as large as it is simply by chance.

A second approach involves fitting the fully saturated model (12345) and ranking the standardized values of the estimates for its w-term parameters. The procedure then calls for specification of a model that excludes all interactions whose corresponding standardized w-terms fail to meet some specified level of significance (3, p. 73).

A third manner of model selection involves ranking the models in descending order of their degrees of freedom from the independence model to the fully saturated. The selection rule is as follows—choose the least complicated model (fewest number of interactions) that satisfies some preestablished significance criterion, for example, that there is no less than 1 chance in 20 that the chi-square statistic associated with the model is as large as it is simply by chance.

When this last technique and its selection criteria are applied to the models of Table 3 the chosen model is number 16—2345 1235 145. This model specifies that all of the level-of-service variables should be interacted with income and that, in addition, the two travel-time variables should be interacted with one another (and jointly with income).

With the selected model in hand, the inclusion of the various terms and interactions can easily be defended. For example, that the two kinds of travel time should be included in the model interactively can be inferred by comparing model 11 with the chosen one, model 16. Inclusion of variables 2 and 3 as part of the larger interaction comes at a cost of 12 degrees of freedom (60 less 48). G^2, however, decreases by 27.17. From a table of chi-square statistics we learn that, under the assumption that variables 2 and 3 should enter the model as in number 11 (the null hypothesis), there is less than a 1 percent probability that a difference as great as 27.17 would arise simply by chance. The null hypothesis is therefore rejected. Similar reasoning and computations may be used to test other aspects of the specification of model number 16.

That cost belongs in the model, interacted with income, can be defended by comparing model numbers 12 and 16. The difference between them is 6 degrees of freedom and a G^2 value of 189.70. A chi-square value this high would occur, if model 12 were correct, with a probability of only 0.000 001. It is interesting to note that the statistical grounds for the inclusion of cost as an explanatory variable are even stronger in the discrete multivariate model here than in some comparable conditional logit specifications where its t-value is 4.5 (6, pp. 158-163) or often a lot lower (11).

Empirical Results

Table 2 sums the parameters for each of the level-of-service variables and its interactions with income: $[w_{2(j)} + w_{25(jm)}]$, $[w_{3(k)} + w_{35(km)}]$, and $[w_{4(l)} + w_{45(lm)}]$.

Figures 1 and 2 plot the information of Table 2 graphically with the y-axis coordinates representing partial effects on the logit and the x-axis coordinates showing mean values for the level-of-service frequencies for each income class. The slopes of the two segments of each function can thus be interpreted as arc elasticities. The graphs of the functions also reveal the extent of any nonlinearities that might exist in the variables as well as the impact on the logit of income when it is interacted with the explanatory variables.

Specific results are described below.

Nonlinearities in Variables

In contrast with the literature that uses the conditional logit model and reports coefficients that are invariant with levels of the explanatory variables, the model here found some significant nonlinearities. The coefficient for the cost variable increases (as one might expect) over the range in which cost advantages favor automobile use. Although the out-of-vehicle travel time variable is approximately linear for low- and high-income households, there is a possible nonlinearity in this variable for the middle-income group. (This is unclear, however, as the coefficients for $\hat{w}_{35(2m)}$ have low standardized values.)

The most striking nonlinearity appears in the in-vehicle travel time variable. For each of the income categories, as well as for all income groups taken together, the effect of this variable flattens out over the range in which the automobile is faster than the bus. From a policy standpoint this would suggest that increases in bus in-vehicle travel time that fall short of rendering the bus absolutely faster than the drive mode may have a negligible impact on transit ridership. Further analysis, opening up four or five categories for the in-vehicle travel time variable, appears to be in order.

Interactions

The interaction between the cost and income variables has the expected result (as depicted in Figure 2), at least in the upper ranges (automobile cheaper) of the cost variable. The cost savings of automobile use cause lower-income households to favor the automobile more readily than do higher-income households.

The interaction between the in-vehicle travel time and out-of-vehicle travel time variables is also reported as significant, according to Table 3.

Onerousness of Out-of-Vehicle Travel Time Versus In-Vehicle Travel Time

Results reported in the literature regarding the relative onerousness of out-of-vehicle travel time and in-vehicle travel time are generally confirmed here. The slopes of the out-of-vehicle travel time functions of Figure 1 tend to be greater than the average slopes of the in-vehicle travel time functions, within and across income classes, by a multiplicative factor of two or three. What is particularly interesting is that, for the range in which bus is faster than automobile, the onerousness of in-vehicle travel time and out-of-vehicle travel time tends to be equal for two of the three income groups, which suggests that for this range commuters are as sensitive to

Figure 1. Differential effects of travel time on logit by income class.

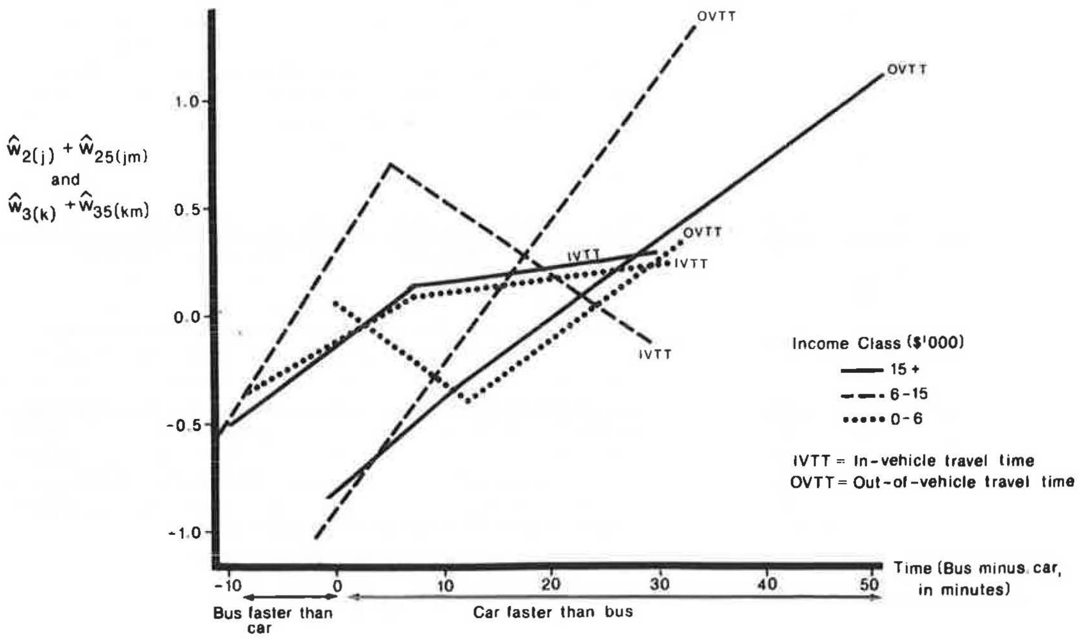
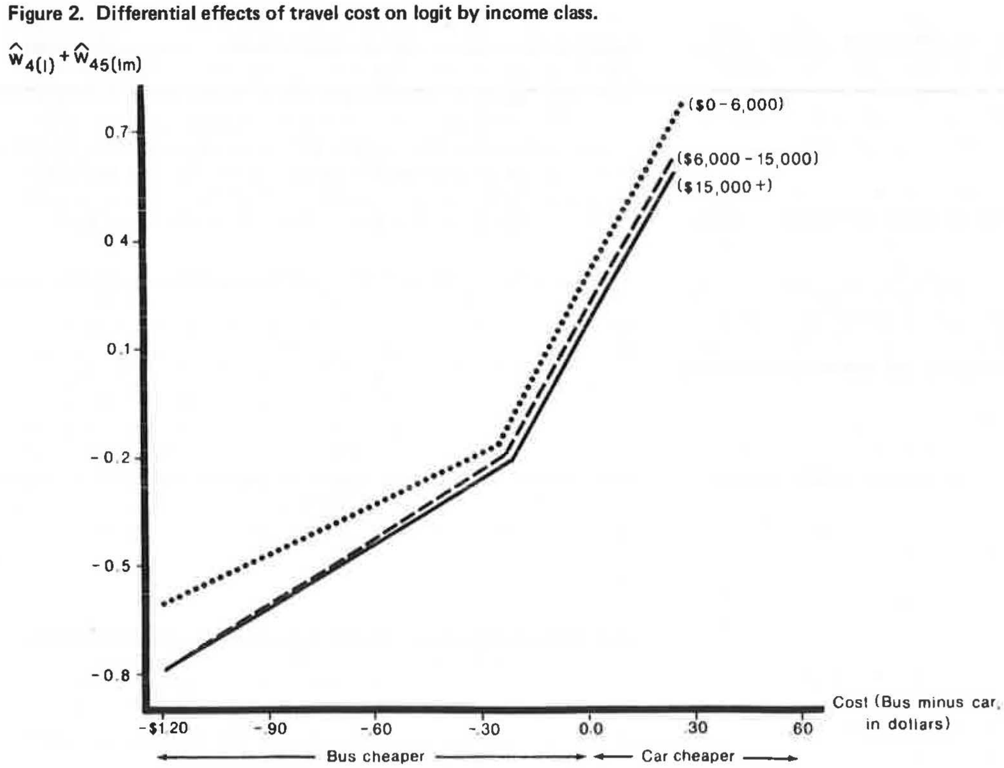


Figure 2. Differential effects of travel cost on logit by income class.



changes in bus-travel time as they are to changes in access time.

Effects of Income

Increases in household income have the predicted effect on the log-odds of driving. High-income households favor the drive mode relative to low-income households, when one automobile is available per worker. As regards higher-order effects, when income is interacted with the level-of-service variables, the major distinction that emerges is between the bottom- and top-income categories: The middle-income category usually proves not to add significantly to goodness of fit.

Automobile-Passenger Mode and IIA Assumption

The theory of the conditional logit model as well as the model employed here assumes that the introduction of a third mode (for example, the shared-ride option for a commuter) will not affect the parameters of the model. If the attributes of such a mode are designated X_3, then according to theory the substitution of X_3 for X_2 in Equations 1 and 1a should leave the β- or w-terms unaffected.

The best way to test for the validity of the IIA assumption here would be to perform a test analogous to a Chow test. This would involve the entire set of 9880 observations, collapsed across all alternatives. A dummy variable would be introduced (call it variable 6) that would take a value of zero if the drive-bus option is described by a particular observation and one if the drive-passenger option is relevant. At issue in the test would be whether this new variable has any two-way interaction with variable 1 and one of the remaining variables. If such interactions emerge as significant then we would reject the hypothesis of IIA, that is, that the w coefficients are the same for the two pairs of alternatives.

An alternative to such a test, which is vastly cruder, involves simply comparing the parameters of the drive or bus alternative with those of the drive or shared-ride alternative. We performed such a test. Because the G^2 statistic for the latter data set was excessively high when the model of Table 2 was used, we employed for both data sets the least complicated higher-order model (a) that passed the standard significance test and (b) of which the model of Table 2 is a proper subset. This was the model: 2345 1235 1245 1345. The probability statistic exceeded 0.4 for each of the two data sets to which this model was fitted.

Of concern was the issue of how the difference between the coefficients for the two data sets compared with the sum of the standard deviations. As the ratio of the coefficient differences to the sum of the standard deviations exceeded 2.0 for only 9 percent of the 148 parameters that were estimated, one might be tempted, by using such a test, to reject the hypothesis that the coefficients for the two alternatives are different—that IIA is not a valid assumption.

ACKNOWLEDGMENT

I wish to acknowledge the assistance and helpful comments of John R. Meyer and Gary Fauth.

REFERENCES

1. D. McFadden. Conditional Logit Analysis of Qualitative Choice Behavior. In Frontiers in Econometrics (Paul Zarembka, ed.), Academic Press, New York, 1974.
2. Y. M. M. Bishop, S. E. Fienberg, and P. W. Holland. Discrete Multivariate Analysis: Theory and Practice. MIT Press, Cambridge, MA, 1975.
3. S. E. Fienberg. The Analysis of Cross-Classified Categorical Data. MIT Press, Cambridge, MA, 1977.
4. M. Ben-Akiva and T. J. Atherton. Methodology for Short-Range Travel Demand Predictions: Analysis for Carpooling Incentives. Journal of Transport Economics and Policy, Vol. 11, Sept. 1977.
5. S. J. Haberman. The Analysis of Frequency Data. Univ. of Chicago Press, Chicago, 1974.
6. T. A. Domencich and D. McFadden. Urban Travel Demand: A Behavioral Analysis. North-Holland, Amsterdam, 1975.
7. D. J. Aigner, A. S. Goldberger, and G. Kalton. On the Explanatory Power of Dummy Variable Regressions. International Economic Review, Vol. 16, June 1975, pp. 503-510.
8. C. F. Manski and S. R. Lerman. The Estimation of Choice Probabilities from Choice Based Samples. Econometrica, Vol. 45, No. 8, Nov. 1977, pp. 1977-1988.
9. D. McFadden, A. Talvitie, and others. Demand Model Estimation and Validation. Urban Travel Demand Forecasting Project, Institute of Transportation Studies, Univ. of California, Berkeley, Phase I Final Rept. Series, Vol. V, June 1977.
10. E. Hanushek and J. Jackson. Statistical Methods for Social Scientists. Academic Press, New York, 1978.
11. M. G. Richards and M. E. Ben-Akiva. A Disaggregate Travel Demand Model. Lexington Books, Lexington, MA, 1975.

Publication of this paper sponsored by Committee on Passenger Travel Demand Forecasting.

Small-Area Trip-Distribution Model

David L. Kurth*, North Central Texas Council of Governments, Arlington, Texas
Morton Schneider, John Hamburg and Associates, Inc., Rockville, Maryland
Yehuda Gur*, Urban Systems, Inc., Chicago

A model for predicting trip tables for small areas based on the access and land development travel function is described along with the results of an initial test of the model. The model provides trip tables required for subregional analyses without the need for windowing into a regional data set. The model requires minimum-path friction skim trees and trip-end data for the small area as input. Trip-end data can be derived from ground

counts on links that enter the small area or from the results of an assignment of a regional trip table. Test results from a small area in Hudson County, New Jersey, suggest the validity of the model. The need for further refinements to the model is discussed in the paper.

In recent years, emphasis has been placed on short-range, small-scale, non-capital-intensive solutions to transportation problems and needs. This emphasis has made development of new tools for quick, inexpensive analysis of subregional alternative actions necessary. Regional models are not always cost effective nor sufficiently detailed for analyzing small-scale or subregional alternatives. Some examples of new tools are windowing and network-aggregation programs and small-area and microassignment programs (1-3).

One need for many subregional or small-area analyses is a trip table. This paper describes a model for developing a small-area trip table. The model provides a powerful tool for the analysis of subregional plans since it does not require data from outside the area of interest.

The small-area trip-distribution model was developed to provide an initial starting point for a process to estimate a trip table based on observed link volumes. The model can be used independently.

PURPOSE OF THE MODEL

The small-area trip-distribution model (SMALD) is based on the access and land development (ALD) travel function (4). It produces a small-area trip table based on link volumes at boundary points of entry and exit (boundary load nodes), productions and attractions at points internal to the small area (internal load nodes), and minimum-path friction skim trees within the area of interest.

A number of definitions are necessary for a clear understanding of the model:

1. Small area—an area where a major portion of the trips have one or both trip ends outside of the area under consideration;
2. Boundary load node—a point where a link crosses the cordon line that defines the area (also referred to as a point of entry or exit);
3. Internal load node—a point internal to the small area where trips originate or terminate; represents an analysis zone;
4. Skim tree—a matrix that gives the generalized cost (friction) of the shortest paths between all load nodes (customarily, friction is a linear combination of travel time and travel cost); and
5. Domain—the part of the region served by a point on the network (usually a load node).

It may be possible to use a standard trip-distribution model to produce a small-area trip table, but the process is conceptually inferior to SMALD. Regional trip-distribution models work on the premise that a major portion of the trips have both trip ends within the area of interest. However, by definition, this assumption is violated in small areas. In addition, standard trip distribution does not consider a source of valuable information on the characteristics of trips that enter at boundary load nodes; that is, the type of service provided by the link as characterized by its functional class at the boundary load node. A trip that enters the small area on an expressway will have different characteristics from a trip that enters the area on an arterial or local street or a trip that is internally generated. This information is essential to SMALD. SMALD explicitly considers that the small area is surrounded by more region that attracts trips (see Figure 1).

It is possible to extract a small-area trip table from a regional trip table by using the Urban Transportation Planning System (UTPS) program NAG (5). NAG performs an all-or-nothing assignment of a regional trip table on the regional network, traces the trips, and records them as they pass through the area of interest. If the regional trip table is unavailable or unreliable, NAG cannot be used for deriving a small-area trip table.

Thus, SMALD has been developed to fill a void. It finds reasonable trip tables for small areas based on data from only the small area. However, it accounts for and uses the fact that the area is surrounded by more region that attracts trips.

SMALD THEORY

SMALD is based on the ALD travel function and a gravity-type distribution process:

$$V_{ij} = P_i F_{ij} R_j / I_i \tag{1}$$

where

V_{ij} = the interchange between point i and point j,
P_i = the productions at point i,
R_j = the attractions at point j,
F_{ij} = decay function, and
I_i = the sum of F_{ij} and R_j.

Decay functions are based on the single-mode ALD travel function:

$$F_{ij} = K_2(2\sqrt{At})/At \tag{2}$$

where

A = a system constant,
t = any measure of separation (e.g., distance, time, or friction), and
K_2 = the modified Bessel function of the second type and second order for the argument $(2\sqrt{At})$.

In SMALD the travel function is different for each type of interchange based on the domains at different types of facilities. The various travel functions are derived by integrating the basic travel function (Equation 2) over the respective domains of the facilities at the points of entry and exit. The following assumptions are made about the domains of points of entry and exit on different facility types:

1. For expressway boundary load nodes, domains expand two-dimensionally into the region external to the area of interest;
2. For arterial boundary load nodes, domains expand in only one dimension into the region external to the area of interest; and
3. For local street boundary load nodes, domains are bounded and small enough to be treated as ordinary point zones (internal load nodes).

Figure 2 shows domains for the different types of load nodes. It is also assumed that the region external to the area of interest is uniform in its accessibility and trip-end density.

The form of the effective travel functions is given in the table below. Note that K_i (where i = 0, 1, or 2) is the modified Bessel function of the second type, ith

Figure 1. Relationship of small area to region.

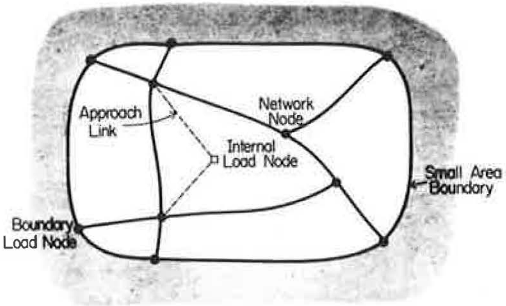


Figure 2. Domains for various load node types.

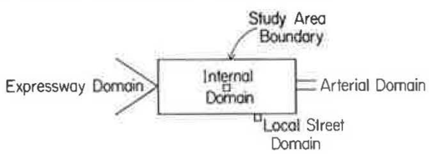
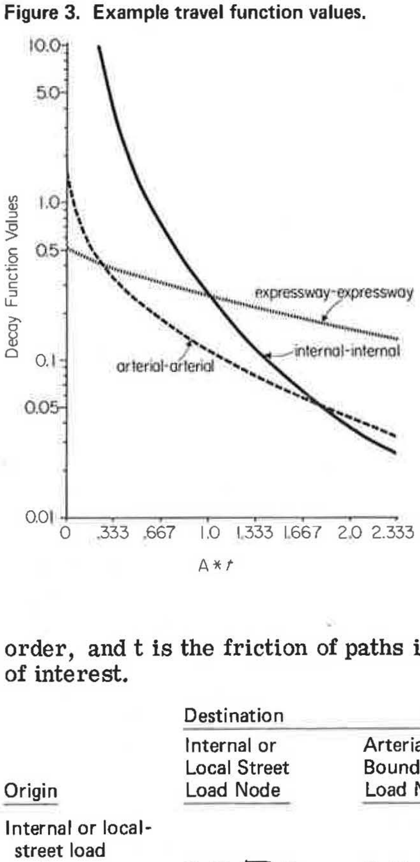


Figure 3. Example travel function values.



order, and t is the friction of paths internal to the area of interest.

	Destination		
Origin	Internal or Local Street Load Node	Arterial Boundary Load Node	Expressway Boundary Load Node
Internal or local-street load node	$K_2(2\sqrt{At})/At$	$K_1(2\sqrt{At})/\sqrt{At}$	$K_0(2\sqrt{At})$
Arterial boundary load node	$K_1(2\sqrt{At})/\sqrt{At}$	$K_0(2\sqrt{At})$	$\sqrt{At}K_1(2\sqrt{At})$
Expressway boundary load node	$K_0(2\sqrt{At})$	$\sqrt{At}K_1(2\sqrt{At})$	$AtK_2(2\sqrt{At})$

Figure 3 shows an example of three of the functions plotted on semi-logarithmic paper. Note that interchanges on higher-level facilities become more and more attractive in relationship to interchanges on lower-level facilities as the friction increases.

Derivation of the Decay Functions

In the derivation that follows, the region is assumed to be uniform—that is, it has about the same trip-end density and accessibility everywhere. Domains shown in Figure 2 assume that local streets always compete with arterial roads, arterial roads always compete with other arterial roads (but not with expressways), and expressways compete only with an arterial infrastructure of some sort. Domains have not been, and cannot be, drawn carefully. Their main and only generally stable feature is their dimensionality (i.e., point-like, linear, or geometric).

The following notations are used in the derivation that follows:

- Q = a quantity proportional to trip interchange volume (i.e., the trip decay function);
- P, p = productions, production density;
- R, r = attractions, attraction density;
- F = the ALD travel function, $K_2(2\sqrt{At})/At$;
- dS = an element of area (see Figure 2);
- C_a = a factor related to arterial domain width and travel friction such that $C_a dt = dS$;
- C_x = a factor related to expressway domain shape and travel friction such that $C_x t dt = dS$;
- $\tilde{P}$ = a quantity similar to, but not necessarily identical with, entry volume at a point of entry ($P = C_a p/A$ for arterial roads and $P = C_x p/A^2$ for expressways);
- R = a quantity similar to, but not necessarily identical with, exit volume at a point of exit ($R = C_a r/A$ for arterial roads and $R = C_x r/A^2$ for expressways);
- t = travel friction internal to small area [with a subscript it signifies friction on external segments (t_i for external before entrance to area, t_j for external after exit from area)]; and
- A = sensitivity to friction.

For internal-internal trips, no extended domains are involved and the function follows the base ALD travel function [see UTPS manual (5) for extended derivation]:

$$Q = PFR$$
$$= PRK_2(2\sqrt{At})/At \qquad (3)$$

For internal-arterial trips, the interchange between an internal zone and an element of area within an arterial domain is measured by

$$dQ = PF(rdS) \qquad (4)$$

so that the interchange to the entire domain is

$$Q = Pr\int_0^\infty FdS$$

$$= PC_a r\int_0^\infty (t + t_j)dt_j$$

$$= PC_a rK_1(2\sqrt{At})/A\sqrt{At}$$

$$= P\tilde{R}K_1(2\sqrt{At})/\sqrt{At} \qquad (5)$$

For internal-expressway trips, the element of area is proportional to $t_x dt_x$ rather than just dt_x, so that

$$Q = PC_x r \int_0^\infty F(t + t_j)t_j dt_j$$

$$= PC_x r K_0(2\sqrt{At})/A^2$$

$$= \tilde{P}RK_0(2\sqrt{At}) \tag{6}$$

For arterial-internal trips, production density and the elemental size of the production domain measure the interchange:

$$dQ = pdSFR \tag{7}$$

so that

$$Q = RC_a p \int_0^\infty F(t + t_j)dt_j$$

$$= RC_a pK_1(2\sqrt{At})/A\sqrt{At}$$

$$= \tilde{P}RK_1(2\sqrt{At})/\sqrt{At} \tag{8}$$

For arterial-arterial trips, both production and attraction domains measure the interchange:

$$dQ = (pdS)F(rdS) \tag{9}$$

so that

$$Q = C_a^2 pr \int_0^\infty dt_i \int_0^\infty F(t + t_i + t_j)t_j dt_j$$

$$= \tilde{P}\tilde{R}K_0(2\sqrt{At}) \tag{10}$$

For arterial-expressway trips, the attraction domain element of area is, again, proportional to $t_x dt_x$, so that

$$Q = C_a C_x pr \int_0^\infty dt_i \int_0^\infty F(t + t_i + t_j)dt_j$$

$$= \tilde{P}\tilde{R}\sqrt{At}\,K_1(2\sqrt{At}) \tag{11}$$

For expressway-expressway trips, both the production and attraction domains are proportional to $t_x dt_x$, so that

$$Q = C_x^2 pr \int_0^\infty t_i dt_i \int_0^\infty F(t + t_i + t_j)t_j dt_j$$

$$= \tilde{P}\tilde{R}AtK_2(2\sqrt{At}) \tag{12}$$

For expressway-arterial trips

$$Q = C_c C_a pr \int_0^\infty t_i dt_i \int_0^\infty F(t + t_i + t_j)dt_j$$

$$= \tilde{P}\tilde{R}AtK_1(2\sqrt{At}) \tag{13}$$

For expressway-internal trips

$$Q = C_x pR \int_0^\infty F(t + t_i)t_i dt_i$$

$$= \tilde{P}RK_0(2\sqrt{At}) \tag{14}$$

Note, again, that local street domains are assumed to be bounded and small enough to be treated as ordinary point zones (internals); wherever the word internal appears, local street boundary load node can be substituted without much damage to the mathematics.

Limitations to the Theory

Perhaps the most severe simplification in the functions listed in the preceding table is that nothing is said about competition among domains, which occurs, for example, when

1. An expressway or arterial emerges from the internal area at a different angle than it enters, so that its entry and exit domains overlap;
2. Two parallel expressways through the area or in its neighborhood cut the domain of each;
3. More than two expressways are involved, so that at least one of the domains tends to grow only linearly with distance from its boundary load node rather than geometrically;
4. An expressway domain pinches off an arterial domain only a short distance from the boundary load node; or
5. Any peculiarity of network geometry or performance causes a facility's domain to have an anomalous shape.

Problems typified by examples 1 and 2 above can be dealt with in an approximate manner by factoring the attractiveness between two boundary load nodes based on the type and severity of the anomaly. If there is an anomaly, the attractiveness between two load nodes can be expected to decrease from what would be expected in a perfect world. Attractiveness factors can be applied in a logically consistent manner. In effect, they modify the size of the domains of the facilities in question.

Problems typified by examples 3 and 4 above may be dealt with on an individual basis. In small areas, links that cut the boundary must be identified by their actual function at that point. If there are two expressways in an area, one may easily serve as an arterial for most trips around that area even though in the region it serves as an expressway. SMALD will perform satisfactorily without taking into account the actual traffic-carrying function of links at boundary load nodes, but results will be improved if this information is known.

The model's performance deteriorates as the size of the area of interest decreases, since the problem becomes increasingly dominated by the specific structure of the network external to the area. This is not considered explicitly in the model.

The present formulation of the model is unimodal (automobile trips only) and does not consider explicitly competition with walking trips. As a result, the model's performance is questionable when zones represented by internal load nodes are extremely small. In practice, this should limit zone sizes to a minimum of about 0.65 km^2 (0.25 $mile^2$).

Figure 4. Study area map.

Table 1. Equilibrium assignment summary results.

Measure	Observed	Run 1	Run 2	Run 3	Run 4
Volume (vehicle-h of travel)	15 350.61	20 866.05	18 959.47	17 659.33	17 325.30
Net error (h)		5 515.39	3 608.97	2 308.87	1 974.77
Percentage error		35.9	23.5	15.0	12.9
Absolute error (h)		6 323.76	4 711.11	3 881.09	3 409.27
Percentage error		41.2	30.7	25.3	22.2
Sensitivity to friction[a]		0.003 8	0.007 6	0.011 4	0.007 6
Average trip length (min)	2.05	2.73	2.49	2.32	2.28

[a]All nonexpressway boundary load nodes considered as local streets.

TEST APPLICATION OF THE MODEL

Data for a 15.5-km^2 (6-mile^2) test area in Hudson County, New Jersey (see Figure 4), were extracted from a regional data set that describes the New York metropolitan area. The area was rather unique since it is parallel to the Hudson River and includes the approaches to both the Lincoln and Holland Tunnels. The data include 24 internal load nodes and 34 boundary load nodes connected by 369 unidirectional links. Two-way ground counts were available for all actual network links in the small area, including all boundary crossings. Productions and attractions for internal load nodes were obtained from the trip file and modified marginally; intrazonal trips, as estimated by the regional trip distribution model, were removed.

Criteria for Calibration

At the outset of the testing, we assumed that a small-area trip table to be used as a calibration standard could be obtained by extracting it from the regional trip table by using a program similar to the UTPS program NAG. Specifically, the regional trip table was assigned by an all-or-nothing process, and trips were traced and recorded as they crossed the area's boundary. It was discovered, however, that the trip table obtained by this process was subject to the pathological quirks of all-or-nothing assignment and, as a result, was not very good.

Since no trip table was available as a standard for comparison, abstract criteria were used to test the quality of the model and calibration. The main criteria used were average trip length in the small area and total absolute link-volume error (observed versus assigned volumes) after five iterations of an equilibrium assignment of the trip table. The average trip length was determined to be 2.05 min from the total vehicle hours of travel on the links (15 350.61 vehicle-h) and the total trips (448 864 trips) in the small area. Note that the trip length was based on friction (a linear combination of time and distance expressed in minutes).

Results

Four runs of SMALD were made. All runs used the same friction skim tree. In the first three runs, the value of the system constant (parameter A in Equation 2) was varied. System constants were chosen to be equal to, two times, and three times the A value used in the trip-distribution model (ALDGRAV) for the region. In the fourth run, the functional class of all arterial boundary load nodes was assumed to be local to more accurately describe their operation due to the unique location of the data set. The A value for the fourth run was twice the regional A value.

Table 1 summarizes the results of the equilibrium assignment of the trip tables from the four tests. Net error is the sum of differences between observed and assigned volumes on all links in the network. Absolute error is the sum of absolute differences in observed and assigned volumes on all links in the network.

Based on the summaries shown in Table 1, the model tends to overpredict average trip lengths even at very high A values (theoretically, the sensitivity to friction used in SMALD should be the same as that used in the regional distribution model).

The size of absolute volume errors is encouraging. Although the errors seem high, they are about one-half of the errors that resulted when the trip table extracted from the regional trip table was assigned. The absolute volume error from the extracted trip table was 64.9 percent of the observed volume.

Although they are not shown here, the resulting trip tables appear intuitively reasonable. Specific interchanges are reasonable in their relative magnitudes. The number of right-angle and U-turn movements through the area appears to be reasonable—about 3 percent of the boundary-to-boundary trips are U-turn movements, and about 30 percent are right-angle movements.

CONCLUSIONS

The results of SMALD are encouraging. SMALD is capable of building a reasonable trip table for a small area, based on data from only that area. There are several areas for possible enhancements to the model, the most promising being methods for specifying routes through the small area. Additional improvements to and testing of the model are currently in progress. The model, in its present form, is applicable in planning studies.

ACKNOWLEDGMENT

The work described in this paper was performed under the sponsorship of the Office of Research, Federal Highway Administration. We wish to thank Steven Cohen, the contract manager, for his encouragement and helpful suggestions in all phases of the project. We are solely responsible for the findings and opinions expressed.

REFERENCES

1. C. Brown and Y. Gur. Adaptable-Zone Transportation-Assignment Package. TRB, Transportation Research Record 637, 1978, pp. 57-61.
2. John Hamburg and Associates, Inc. Micro/UROAD Users Guide. Urban Mass Transportation Administration, Aug. 1978.
3. M. Turnquist and Y. Gur. Estimation of Trip Tables from Observed Link Volumes. TRB, Transportation Research Record 730, 1979.
4. M. Schneider. Access and Land Development. In

Urban Development Models, HRB, Special Rept. 97, 1968, pp. 164-177.
5. UTPS Reference Manual. Urban Mass Transportation Administration, July 1977.

Publication of this paper sponsored by Committee on Passenger Travel Demand Forecasting.

**D. L. Kurth and Y. Gur were employed by John Hamburg and Associates, Inc., when this paper was written.*

Disaggregate Travel Models: How Strong Are the Foundations?

A. G. R. Bullen and Russell H. Boekenkroeger, University of Pittsburgh

This paper presents a review and analysis of disaggregate travel-demand modeling founded on an examination of the published literature. This analysis is directed to the conceptual foundations of the modeling process, which appear to be somewhat obscurely covered by the literature. The analysis is at two levels: (a) a review of where the modeling structure fits into the overall travel process and (b) an analysis of the foundations of the specific models and how they relate to the target processes. The particular disciplinary backgrounds that lead to the model formulations are reviewed since a qualitative interpretation appears to be lacking in the travel literature. From these analyses it is concluded that the basic random-utility travel model does not have a sufficient behavioral foundation that allows its generalized usage for all components of the currently perceived travel structure. As a consequence it would seem to have somewhat limited application for many transportation policy questions. The paper suggests that a more diversified modeling approach is required, that the traditional modeling structure should be reviewed to exclude unimportant functions and introduce more policy-relevant ones, and that, in the development of models, considerably greater attention needs to be given to the establishment of criteria for their evaluation and verification.

Disaggregate travel-demand models (DDMs) have been at the forefront of transportation systems-analysis research and academic activity for the past 10 years. The reasons for this are several. They respond to the practical need to develop more effective models for travel prediction and transportation evaluation. They provide considerable intellectual challenge in their use of sophisticated techniques, and their attractiveness has been heightened by the theoretical derivation of an apparent behavioral basis for earlier empirical developments.

Despite the quantity and sophistication of the work that has been done on DDMs, several areas of concern appear:

1. They still have not been accepted by a substantial segment of transportation practitioners;
2. They have not, so far, provided any spectacular breakthroughs in modeling or understanding; and
3. The literature reveals lingering uncertainties and continuing problems with models and data.

The main response to these difficulties has been greater technical activity in search of a more complex and sophisticated methodology. Nevertheless, the literature has not become much clearer. To many, the methodology remains unclear and the problems remain to be clarified.

This paper postulates that the immediate need is a reexamination of the foundations of the models to provide at least a clear, concise, and simplified explanation of them, if not a redirection of the modeling process. The literature provides much confusion in definition and terminology at the conceptual level. Many of the foundations of the modeling approach are subjectively derived without testing of the underlying assumptions. Some of the most important concepts are left to the references, which also remain obscure.

A clear response to these questions, including clarification of the concepts, would broaden the understanding and acceptance of DDMs. The conceptual base of the current econometric thrust is so narrow, however, as to preclude the confirmation of the strong empirical results claimed. The usefulness of the methodology is thus restricted to fairly limited applications.

BACKGROUND

DDMs are widely reported in the literature. A series of conference proceedings provide the most exhaustive reviews (1-3). Specific modeling developments are provided by Ben Akiva (4), Charles River Associates (5), Domencich and McFadden (6), and Manski (7).

DDMs were originally developed to gain greater insight into travel behavior, particularly at the individual level. This fundamental understanding was found lacking in the aggregate forecasting models generally used in the Urban Transportation Planning (UTP) process. Critiques of traditional aggregate models are abundant (1, pp. 13-19).

The initial modeling work was of an empirical nature, developing logit models of mode choice. Later theoretical work of Charles River Associates (5), McFadden (8), and Domencich and McFadden (6) provided a behavioral interpretation and foundation for the preceding empirical work. Despite this formulation, however, the nature of the DDM methodology remains overwhelmingly empirical. Conceptual difficulties and behavioral inconsistencies have arisen from time to time, and the underlying theory has often been adjusted in an ad hoc manner to account for discrepancies (7). Empirical and technical work has dominated; less attention has been given to theoretical understanding, and, unfortunately, this had led to what seems to be lack of concern for the modeling foundations.

DDMs have been looked on as accurate, inexpensive replacements for traditional forecasting models; they are capable of dealing with policy questions that the earlier methodology could not handle. Yet, with the few exceptions, DDMs have not become a standard tool for analysis in practical settings. This is despite their virtues over the UTP models (4, 6).

From time to time questions of a conceptual and theoretical nature have been raised about DDMs. These questions include the difference between aggregate and disaggregate models (2, pp. 116-126), the lack of an appropriate treatment of nonchoosers (2, pp. 173-179), the applicability of these models to policy issues (9), and the wholesale restructuring of the modeling process (10). Recently, a number of interrelated research problems have been proposed (11-13). The questions and problems, however, remain. Questions arising from the basic assumptions of DDM are being addressed at the technical level. If the underlying concepts are being considered, then the literature does not make this clear. Rather, it provides a sometimes confusing terminology and unclear references.

An examination of the assumptions of the theory may be unwarranted if the principal concern is with testing a model's predictions rather than its assumptions (14). The purpose of this paper, however, is to review the modeling processes through an examination of the concepts and assumptions.

FOUNDATIONS AND CONCEPTS

As the literature recognizes, personal travel is an extremely complex process. The actual mechanism by which this complexity is reduced to a manageable methodology is the central theme of this paper. Although some of the assumptions used in a DDM are criticized here, this is not in conflict with the process of idealization and simplification that is essential to develop a workable model of a complex phenomenon. Of principal concern are those instances where basic structures affected by the circumstances being modeled are not reflected in the modeling methodology. Alternatively, there is also concern for those instances where the foundations of the methodology are so imprecise or unclear as to make the user insensitive or unaware of the actual processes being dealt with.

Two distinct levels of concept are dealt with in the analysis of these foundations. The first and more general level that is examined is concerned with the general character of personal travel and how it relates to the overall travel methodology. This provides the background for the second and more extensive level of concept, namely those assumptions and techniques that lead to the specific DDM.

Travel is a realization of human activity structured over a spatial framework. The analysis of these spatial connections is the travel modeling problem and, as such, it has been frequently and clearly described throughout the literature (6). This initial characterization, however, is frequently followed by a precipitous leap to the description of rational economic man as a utility maximizer. At most, strictly qualitative attention has been given to the concepts and subsequent assumptions that transform the former into the latter.

This human activity is an assembly of individual activities integrated into the larger structure of some behavioral unit, generally agreed to be the household. It is here that DDMs are initially tenuous. Although they have been related to household decision questions (such as residential location and automobile ownership), their basic travel structure is concerned with the individual. The models, therefore, will have limited value for policy analyses, where changes in the structure of household interactions are likely. Changes in energy availability, vehicle size, life-style, and the role of women have significance for the internal activity structure of the household—its subtle interactions and substitutions. In these kinds of instances the assumptions of the separability of the individual utility functions, so essential to DDMs, are unrealistic.

The closest approach to the household-identification problem is the market-segmentation process currently in vogue. This, however, cannot analyze changes in household-activity structures unless these changes coincide with transitions between market segments. Market segmentation appears to be an ad hoc response to deficiencies in the abilities of the models to handle demographic or socioeconomic characteristics.

At the level of the individual, DDMs make further idealizations of the basic activity structure. Two key assumptions are made:

1. Activities spatially removed from the home have a suitable surrogate in trip purpose and
2. The separability of utility applies to all components of the activity-travel structure.

These simplifications are sometimes necessary to reduce a complex process to a reasonable model but, once again because of the separability criterion, important interactions are not explicitly considered. For assumptions of this kind, more effort should be given to identification of their range of application.

Some of the conceptual problems raised here are related to the relationship between the traditional aggregate modeling process of UTP and the disaggregate approaches. Although disaggregate theories are intended to overcome basic difficulties of the traditional methods, they are highly derivative of these methods. The traditional aggregate simulation models still dominate travel analysis thought, and some of the conceptual problems of aggregate models transfer directly to DDMs. The difficulties start (2, pp. 116-125) with the mere description aggregate versus disaggregate, which gives the impression that the individual is being analyzed. This, however, is an economic interpretation, and the study of the individual consumer is actually the study of a homogeneous aggregate of consumers and, similarly, DDMs are the study of homogeneous aggregates of travelers. Both models are aggregate. The traditional models aggregate space whereas the newer ones aggregate class of individual (or occasionally household).

The major transfer of traditional techniques revolves around the definition of the trip and the maintenance of purpose as a substitute for activity. There appears to have been little, if any, questioning of the basic trip structure of frequency, time, mode, destination, route, and purpose. Perhaps alternative structures are feasible. To approximate activity with purpose requires separability notions that are difficult to justify. The analysis of household-activity patterns is being looked at (15) and the travel implications have been conceptualized (16), but their potential impact on DDMs is limited. More understanding of activities, time consumption, spatial structure, and household interactions are needed. DDM developers have realized this, but they have tended to pass over these subjects through qualitative reasoning and strings of assumptions.

FUNDAMENTAL CHOICE CONCEPTS

The travel process just mentioned is treated in DDM in a traditional economic framework that has some formal mathematical propositions from psychology integrated into it. Theories and models of travel that originate from this framework are well documented (6, 7), but little attention has been devoted to relating the framework to the travel process. The failure to explain precisely how the underlying concepts of the models are related to these economic and psychological foundations is a source of many conceptual difficulties. To develop a cohesive basis for further discussion, these foundations are now highlighted. The material is taken from the standard references cited in the DDM literature.

Theory of Choice in Economics

Current approaches to the theory of choice establish those axioms that must be fulfilled for the existence of any choice problem. The axioms that constitute the general theory of choice ensure that (a) a universal set of choices may be partitioned into the mutually exclusive, attainable choice set; (b) all elements of the universal choice set may be compared and an induced strong ordering of the elements established; and (c) an element will be chosen and it will be the one most preferred (17, 18).

These principles must be made more specific in the consideration of a particular choice problem by asserting clear restrictions on the choices made by a choosing agent, identifying an attainable choice set, and positing the criterion that will rank the choices. In consumer theory, the criterion used is utility, and the mechanism that provides this is the utility function.

Consumer Theory in Economics

In consumer theory the choosing agent is identified as an individual consumer and the commodities that comprise his or her choice set are those that he or she has at hand. Most consumer theory considers that the commodities themselves give rise to utility. DDMs incorporate the approach of Lancaster (19), wherein the intrinsic characteristics of the commodities give rise to utility. Lancaster postulated that the characteristics possessed by a good are the same for all consumers. In DDM a somewhat modified approach is taken wherein different homogeneous segments of a population have different consumption characteristics.

To delineate the attainable choice set for individual consumers, additional assumptions are required of these assumptions, and they ensure that the preference of utility function possesses certain properties that are to be exploited. Once a consumer's utility function is known and if he or she continues to behave rationally, the demand function may be derived.

Theory of Revealed Preferences

For DDMs, McFadden (8) has identified modeling travel choices as the population analogue of the theory of revealed preferences for individual consumers, which originated with Samuelson (20, pp. 90-123), who proposed that, by observing a consumer's actions, preferences would be established. The advantage of this theory is that, being based solely on observed behavior, it is presumed to be testable. In its most general statement, the theory entails two axioms:

1. Given a choice set, the consumer must make a choice and
2. If the consumer reveals a preference, it can never be violated at the same set of prices.

In this theory an outside observer constructs the preference or utility function to conform to the rankings that a consumer makes. If the function successfully ranks the choices of consumers, then it is interpreted as explaining the behavior. However, the theory only allows us to glean information about a consumer after choices have been made. Unless some independent information exists on the way in which a consumer's preference calculus changes over time, the observer is unable to conclude anything before the fact about the process that gives rise to the observed behavior. By assuming that tastes and preferences are fixed in the short run this problem is avoided and the theory is complete.

Utility

The criterion that a consumer employs in making choices is utility, and the mechanism is the utility function. When this concept is employed in consumer theory, some meaning is invariably associated with the term. Utility is assumed to summarize a consumer's sense of well being and it is generally interpreted as a reduced form of a number of complex psychological and sociological processes. Without dealing directly with these processes, utility may be interpreted to take account of them, albeit in an unspecified manner.

The characteristics of the choice are selected for inclusion in the utility function by the observer based on his or her substantive knowledge of the choice problem. He or she may not know for sure what the characteristics are and, in the empirical analysis of consumer-choice problems, different characteristics and transformations are tried to obtain that combination that is both theoretically plausible and empirically valid. Of the two classes of variables that enter the utility function in DDM (characteristics of the chooser and the choice), utility is encapsulated in characteristics of the choice. The characteristics of the chooser are used primarily to establish homogeneous market segments of consumers.

The concept of utility is a controversial one, even within the economics discipline, and considerable argument exists about its measurement and validity (21, 22). As a basis for travel modeling, Fried and others (10) tend to dismiss it entirely. Nevertheless, it is a flexible concept, wide ranging over many disciplines, and it provides a driving mechanism for the models.

CHOICE THEORY IN PSYCHOLOGY

The study of choice behavior in psychology is a search for the laws between stimulus and response relations, which can be generalized in many cases to the gamut of human decision-making situations. Empirical analysis guides the determinations of which theories are applicable to particular choice situations (23-25). Those developing DDMs have referred to and used formal propositions of mathematical psychologists, particularly Luce (26) and Thurstone (27).

Luce's Theory of Individual Choice Behavior

Luce presupposes that choice behavior is best described as a probabilistic phenomenon. This philosophy is adopted because of observed intransitivities in individual decision making and the plausibility of a probabilistic interpretation for the majority of choice problems addressed by psychologists. Luce's theory has an axiomatic foundation, with the standard probability axioms as its starting point. He assumes only mathematically well-defined sets of choice alternatives.

The core of the model is the choice axiom, which consists of two parts. The first part states that, if all pairs of discriminations among the elements of a universal set are imperfect, then the choice probabilities for any subset are identical to those for the universal choice set, conditional on the subset having been chosen. The second part states that if one particular element is never chosen over another, then the former element may be deleted from the universal set without affecting any of the choice probabilities.

Two consequences of the choice axiom that have been used by DDMs are the constant ratio rule, leading to independence from irrelevant alternatives, and the numerical ratio scale for characterizing alternatives in the choice set. The constant ratio rule states that the probabilities of choosing one alternative versus another do not depend on the total set of alternatives. It is the ratio of probabilities, not the probabilities themselves, that is invariant. The constant ratio rule maintains the assumption of pairs of discrimination among alternatives as well as transitiveness of choices. These are also two of the more important basic axioms of choice theory in economics. The choice axiom also implies that a numerical ratio scale exists over the choice set. In DDM, utility is represented in terms of a numerical ratio scale.

Thurstone's Law of Comparative Judgment

Thurstone's law of comparative judgment (27) is based on the notion that choice alternatives (as a stimulus) are subjectively experienced by an individual as intrinsically variable, and this accounts for the variability in individual judgments. Alternatives are treated as normal random variables and are called discriminal processes that represent the indirectly observable psychological values involved in choice. A case V Thurstone model is formally comparable to Luce's choice axiom, and it is the one of importance for DDM. The discriminal processes are assumed to have identical variances and common covariances, such that the marginal distributions differ only in their locations along the axis. The different stimuli or characteristics of the alternatives, described by real valued scale functions, are identically and independently distributed normally about their mean values. Thurstone's case V model is more familiarly known to economists and transportation analysts as the random-utility model.

Thurstone, Luce, and the Double Exponential Distribution

For pairs of discrimination problems, Luce's choice axiom, which results in the logistic distribution and the normal distribution of Thurstone's case V model, produces similar results except for the tails of the distributions (28, p. 216). Conceptual differences between Luce and Thurstone notwithstanding, McFadden (8) and Yellott (29), independently and under different assumptions, have demonstrated for multiple-choice comparisons that, if the random variables for Thurstone's model are restricted to differ only in their means, then Luce's choice axiom and Thurstone's case V random-utility model are formally equivalent. The double exponential distribution provides the linkage between the two. This distribution is referred to as the Weibull in travel literature and the Gumbel in some other disciplines, where Weibull is reserved for an alternative extreme value form. The principal result of this finding is that the multinomial-logit model has a random-utility interpretation along the lines of Thurstone's case V model. By assuming the double exponential as the underlying probability distribution, an explicit model for determining individual-choice probabilities results.

CONCEPTUAL ISSUES IN THE BASIC MODELS

The purpose of this section is to identify and discuss conceptual and theoretical issues of DDM, particularly as they relate to the concepts just highlighted. Some of the issues mentioned here have been presented elsewhere in the literature (1-3, 11). The specific organization given to this discussion focuses on the issues of how travel is characterized and modeled. It is this particular aspect in which the literature is obtuse and usually concentrates on the technical aspects of the models.

Basic-Choice Model: The Probability of What?

The heart of DDM is a basic-choice model, which in its elementary form is written

P (i:A) = probability of choosing i from the travel choice set A.

In dealing with this simple-choice concept as a starting point, however, the transportation literature presents a confusing and often inadequate notion of precisely what concept of choice is being developed and, more importantly, precisely what behavioral ideas are involved. There are three possible interpretations of the probability-of-choice model presented above. They involve to varying degrees the analyst, or observer, and the subject, or consumer.

Model A—The probability involved refers to a sampling probability that the subject, who has completed a fixed choice, will be selected by the observer.

Model B—The probability involved refers to the probability of choice by the subject where his or her choices vary randomly over repeated trials.

Model C—The statistical methodology implied in model A is being used on a group of model B subjects to estimate their probability distributions.

Invariably, DDMs are of the type described in model A. This is often clearly stated (6, 30), but on balance this distinction is left unclear by much of the literature. The question at this juncture then is why the psychologists are references for the basic choice. Clearly, for DDM to be behavioral in any more than a strict statistical sense (where independent variables explain the behavior of the dependent variable), something else is being implied. Are disaggregate models trying to get at model B through model C or what? McFadden (25) uses the mathematical methodologies of the psychologists by restating the choice axioms in the context of model A. The generalized framework of Manski (7) combines observer and subject in the context of model A, but this requires a narrowly defined individual-choice mechanism. Formal similarities aside, the underlying choice concepts of DDM are not those of psychology. Model C presents serious theoretical and conceptual problems.

Conceptual and behavioral confusion first arises from the different probability definitions implied. Model A represents the relative frequency view of probability, and model B implies the degree of confirmation concept of probability, as defined by Carnap (31). These are two of the major definitions of the several put forward by various authors. By adopting this view, probability may be taken to have a substantive meaning in particular applications. Thus, model A and de facto disaggregate models are incapable of logically supporting testing of behavioral hypotheses. By its very structure model A must be an aggregate model.

Model B is a true individual model and is thus disaggregate, wherein a probabilistic mechanism is used to reflect the degree of uncertainty of a decision maker regarding his or her alternatives. Model A, on the

Table 1. Comparison between choice theories of Luce-Thurstone and disaggregate travel models.

Dimension	Disaggregate Models	Luce-Thurstone
Type of probability	Sampling probability, relative frequency	Subjective probability degree of confirmation
Nature of choice experiment	Complex, traditional submodels provide choice sets	Simple
Choice subjects	Aggregates of persons (market segments)	A single individual
Number of trials	Single observation for each individual	Many observations for each individual
Individual decisions structure	Fixed	Random
Underlying individual preferences	True individual preferences unknown	Preferences applied repeatedly to similar choices
Attributes of choice alternatives	Function of attributes of dissimilar choices determined by the observer	A single attribute is varied in a predetermined manner
Intervening processes	Many, not all of which are known, understood, or examined	Controlled by the observer

other hand, results in a sampling probability of choices arrived at by decision makers from systems where the alternatives are fixed. The probability mechanism arises from the variation of that set of characteristics of alternatives for the subject unknown to the observer. The varying preferences are accounted for by the joint consideration of fixed statistical distributions of these unknown characteristics. The justification for these distributions, which are the behavioral core of the DDM is, at best, fuzzy. Unlike the Luce model, behavior is not directly modeled but is inferred from the apparent differences that individuals as consumers of travel indicate in their preference structure. The probabilistic core of DDM, therefore, appears to be predicated on the error term in the model structure and the data base it is calibrated from. Some of the comparative differences between the model concepts are summarized in Table 1.

Deriving the Basic Travel Model

The individual probabilistic choice models of psychology described earlier (model B) are a means of exploring intransitivities of behavior in simple-choice experiments. In the context of simple-choice experiments with repeated trials, the characteristics of the alternatives (their utilities) are treated as random variables that reflect the subjective preferences of an individual choosing agent. The associated response is uniquely determined on each presentation by the choosing agent. The alternatives are all known to the choosing agent and to the person conducting the experiment.

A travel model begins with the random-utility model (model A), which has been interpreted by economists as an econometric interpretation of maximizing behavior. In DDM this interpretation results in the fixed utilities of travel choices being treated as random variables by an observer who samples from the personal travel data set (5, 6, 30). The particular application of the random-utility model used in these models is more in the spirit of deterministic modeling than probabilistic modeling. Consequently, the randomness results not from a lack of rationality or uncertainty on the part of the traveler as to the utility of his or her alternative choices, but from a lack of information on the part of the observer as to which individual is chosen and the true utility of the alternatives.

The characteristics that are specified by the observer comprise the mean utility in the random-utility model, and those characteristics that are not specified are assumed to be part of the intrinsic utility, which each individual considers uniquely, or that utility that the observer does not have knowledge of. The socioeconomic characteristics of the traveler included in the utility function serve the primary purpose of segmenting the sample into homogeneous groups that have similar tastes and preferences. Within each market segment it is assumed that demand has a structure determined by behavioral regularities, which remain stable over time and space. As individuals are sampled from the data set, only the choices made or their revealed preferences are known to the observer, since he or she has no knowledge of the actual alternatives at the time the observed choice was made.

A core conceptual problem is the random distribution of unknown tastes, which is the essential behavioral driving force of the DDM. It has a particular set of properties assigned to it, yet little is actually known about it. It remains unknown, and must remain unknown, for the model as such to survive. The model is data specific. If more behavioral variables emerge they cannot come out of the distribution of unknowns, so a new model is specified. The distribution of tastes then must change its dimension but maintain its distributional properties. There has been no interest in establishing any information about the details of this basic behavioral process. Perhaps this indifference to the behavioral core of the model is responsible for Luce's apparent lack of interest in travel modeling (28).

Less fundamental technical questions arise. The independence of irrelevant alternatives issue has been widely thrashed around, but it presents a conceptual singularity fatal to the imputed behavioral basis of the model. The implication of the distributional independence requirements on the model are rarely addressed. Also, why are extreme value distributions used? In most applications of these, the use of an order statistic is clearly related to the modeling purpose and the parent statistical distribution contributes to that purpose.

The Basic Model Applied to a Perceived Travel Structure

The next stage of the travel-modeling process is to apply the basic random-utility model to the perceived travel structure. As already discussed, this perception is highly derivative of the existing UTP process and the available data bases. Two general approaches have been taken: the recursive approach and the simultaneous approach. The problems discussed below apply to either. The basic choice model is applied to every phase of the travel process, although its derivation has been largely in terms of mode choice. The travel process as conceptualized in conventional UTP submodels is purely descriptive. This breakdown of travel choices (frequency, mode, destination, time of day, route, and purpose) seems to be accepted as a matter of faith. There appears to be little discussion on whether alternative structures may be desirable, whether each of these components is equally important, and whether all components are relevant to the analyses to which the models might be put. This structure will be examined from the point of view of relating a qualitative view of the elements of the travel process to the basic model. The table below summarizes this analysis.

Travel Component	Behavioral Process
Mode	The individual's perception of the modes is constant—model A
Destination	Model A or model B, depending on trip purpose
Route	Generally model A but some model B by regular commuters
Frequency	A renewal point process
Time of day	A renewal process or scheduling process

For mode choice, the choice model is applied to the fixed preferences of a variable population, with complete knowledge of their travel alternatives. Each individual arrives at a consistent choice. Since the random-utility model has been derived in this context, this is a reasonable approach to what can be visualized. Generally, mode choice appears constant and, if the individual does randomly vary choice of mode, it is probably for reasons unrelated to the variables usually calibrated. Mode choice is a model A choice mechanism that has led to the random-utility model formulations.

For the choice of destination, there is the possibility of randomly varying individual choice, as given by model B. The choice set will be extremely complex since trip purpose does not define activity very well. Depending on the activity engaged in at the destination, some forms of the mechanisms supplied by both models A and B will be in evidence. Most work trips entail the fixed preference of model A, but for others, such as shopping or social and recreational trips, some form of model B mechanism may be operational. In any case, a uniform behavioral interpretation is not possible across the various purposes of travel. The choice set will also vary from household to household, confounding the determination of homogeneous market segments.

In the application to route choice the problems inherent with destination reappear. No doubt, many decision makers are displaying a fixed preference and others present more probabilistic individual behavior on a day-to-day basis. The direct application of the random-utility model presents some conceptual difficulties. Route choice deals with one of the most clearly defined choice problems, since the decision is closely related to the usual fixed attributes of cost and time and not relevant to any unknown tastes. The real decision mechanism is probably driven more by incomplete information on the part of the decision maker, a model B process.

Problems arise for the associated choices of frequency and times of day, since they fit neither model B nor model A. Clearly the traveler does not choose frequency in the preference scale of the basic model. What are its attributes? How does the taste variation fit around the do-nothing alternative, which may be a do-it-tomorrow alternative? This particular choice and that of time of day involves some kind of renewal process, a stochastic point process, which hardly fits either conceptual approach.

The simultaneous model structure lumps all of the described choices together into one model, which corresponds with model A. This can result only in models of extreme behavioral obscurity and great complexity.

CONCLUSION

The conceptual foundations of the DDM as they relate to choice theories from economics and psychology have been highlighted and explained. Behavioral interpretations attributed to DDM by the use of these choice concepts appear to be mainly formal in nature and rather obscure when related to the travel processes being modeled. Much of the DDM literature misinterprets what can be achieved from these concepts, and the application of them to the perceived travel process compounds the conceptual difficulties of the DDM. The fundamental discrepancies between the stated and actual interpretations of the DDM indicate a tenuous behavioral base and render its use for most purposes highly suspect.

The random-utility model is perceived as a significant advance, in fact and in potential, on the conventional UTP models—yet the travel structure is unchanged, the calibrated variables are little different, and the aggregation remains, albeit on a different dimension. These models are driven by variation over the population rather than by the imputed variability in the individual decision-making processes. As a consequence, the underlying behavior being modeled remains largely unexplained.

DDMs meet some important modeling objectives in that they are elegant and simple. Yet, as reasoned here, they are not proven behaviorally and, as such, they should not be considered sacrosanct and the only basis for further examination and generalization of travel. The models are helpful to have and they possess properties that may be exploited, but they are not a behavioral truth. DDMs have provided no modeling breakthroughs nor have they led to an increased understanding of travel.

A greater awareness of the complex processes that cause travel is required. Attention at all levels of the modeling process would help to conceptually structure models that are behaviorally and empirically valid. The determination of criteria for evaluation should be a parallel effort to the development of the models themselves, for the lack of clearly stated and operational criteria for evaluation is one of the causes of the confusion and inconsistency in current models.

The use of probability in DDM does not appear to have proved any new insight into the travel process. It is used in a descriptive statistical sense to take account of human variability, whether or not that variability is germane to the problem at hand. Yet, the process at hand may be susceptible to stochastic analysis since the events take place over time. Rather than use probability as a substitute for what is not known, it could perhaps reinforce what is known.

The conclusion is that DDMs lack the strong foundations, the power, or the capability to provide much additional understanding of travel structure. Beyond a predictive capability in the short run within the limits of their empirical calibration, they would appear to be limited in application. The development of more explanatory models for travel analysis will require more diverse research approaches, which will entail a concentration on assumptions rather than on methodologies.

REFERENCES

1. Urban Travel Demand Forecasting. TRB, Special Rept. 143, 1973.
2. Behavioral Demand Modeling and Valuation of Travel Time. TRB, Special Rept. 149, 1974.
3. P. Stopher and A. Meyburg, eds. Behavioral Travel Demand Models. D.C. Heath and Company, Lexington, MA, 1976.
4. M. Ben-Akiva, T. Adler, J. Jacobson, and M. Manheim; Massachusetts Institute of Technology. Experiments to Clarify Priorities in Urban Travel Forecasting Research and Development. Urban Mass Transportation Administration, 1977. NTIS: PB 295 051/7SL.
5. A Disaggregated Behavioral Model of Urban Travel Demand. Charles River Associates, Inc., Cambridge, MA, Final Rept., March 1972. NTIS: PB 210 515.
6. T.A. Domencich and D. McFadden. Urban Travel Demand: A Behavioral Analysis. American Elsevier Publishing Company, Inc., New York, 1975.
7. C. Manski. The Structure of Random Utility Models. Theory and Decision, Vol. 8, 1977, pp. 229-254.
8. D. McFadden. Conditional Logit Analysis and Qualitative Choice Behavior. In Frontiers in Econometrics (P. Zarembka, ed.), Academic Press, New York, 1973, pp. 105-142.
9. A. Talvitie. Disaggregate Travel Demand Models with Disaggregate Data, Not Aggregate Data, And for What. Institute for Transportation Studies, Univ. of California, Berkeley, Working Paper 7615, pp. 37-77.
10. M. Fried, J. Havens, and M. Thall. Travel Behavior: A Synthesized Theory. NCHRP, Draft Final Rept., March 1977.
11. D. Brand. Approach to Travel Behavior Research. TRB, Transportation Research Record 569, 1976, pp. 12-33.
12. I. Heggie. Putting Behaviour into Behavioural Models of Travel Choice. Journal of Operational Research Society, Vol. 29, No. 6, 1978, pp. 541-550.
13. R. Boekenkroeger. Disaggregate Behavioral Travel Demand Models: A Questionable Basis for Transportation Policy Analysis. Department of Civil Engineering, Univ. of Pittsburgh, MS Thesis, 1978.
14. M. Friedman. Essays in Positive Economics. Univ. of Chicago Press, Chicago, 1966, pp. 3-43.
15. F. S. Chapin. Human Activity Patterns in the City. Wiley, New York, 1974.
16. I. Heggie and P. Jones. Defining Domains for Models of Travel Demand. Transportation, No. 7, 1978, pp. 119-125.
17. J. Henderson and R. Quandt. Microeconomic Theory: A Mathematical Approach. McGraw-Hill, New York, 1971.
18. R. Shone. Microeconomics: A Modern Treatment. Academic Press, New York, 1976.
19. K. Lancaster. A New Approach to Consumer Theory. Journal of Political Economy, Vol. 1, No. 3, 1967, pp. 132-157.
20. P. Samuelson. Foundations of Economic Analysis. Atheneum Press, New York, 1976, pp. 90-123.
21. G. Clarkson. The Theory of Consumer Demand: A Critical Appraisal. Prentice-Hall, Englewood Cliffs, NJ, 1963.
22. M. Hollis and E. Nell. Rational Economic Man: A Philosophical Critique of Neo-Classical Economics. Cambridge Univ. Press, London, 1975.
23. R. Bock and L. Jones. The Measurement and Prediction of Judgement and Choice. Holden Day, San Francisco, 1968.
24. R. Luce and P. Suppes. Preference, Utility and Subjective Probability. In Handbook of Mathematical Psychology (R. Luce and E. Gallanter, eds.), Wiley, New York, Vol. 3, 1965, pp. 249-410.
25. D. McFadden. Quantal Choice Analysis: A Survey. Institute of Transportation Studies, Univ. of California, Berkeley, Working Paper 7625, Jan. 1976.
26. R. Luce. Individual Choice Behavior. Wiley, New York, 1959.
27. L. L. Thurstone. A Law of Comparative Judgement. Psychological Review, Vol. 34, 1927, pp. 273-286.
28. R. Luce. The Choice Axiom After Twenty Years. Journal of Mathematical Psychology, Vol. 15, 1977, pp. 215-233.
29. J. Yellott. The Relationship Between Luce's Choice Axiom, Thurstone's Theory of Comparative Judgement and the Double Exponential Distribution. Journal of Mathematical Psychology, Vol. 15, 1977, pp. 109-144.
30. Disaggregate Travel Demand Models. Charles River Associates, Inc., Vols. 1 and 2, Cambridge, MA, Feb. 1976.
31. R. Carnap. Logical Foundations of Probability. Univ. of Chicago Press, Chicago, 1962.

Publication of this paper sponsored by Committee on Traveler Behavior and Values.

Choice of Access Mode to Intercity Terminals

J. Sobieniak, N. D. Lea and Associates, Ltd., Vancouver
R. Westin, University of Toronto and the World Bank, Washington, D.C.
T. Rosapep and T. Shin, University of Toronto

Disaggregate demand models are developed for Canada's national capital region (Ottawa-Hull and vicinity) for the choice of access mode to intercity transportation terminals. Models that consider a choice of five alternative access modes are reported for the airport, railroad station, and intercity bus terminals. The results show that considerations of convenience (walking time, schedule frequency, and baggage handling) are dominant factors in the choice of access mode. The models are applied to test passenger preference for several proposed strategies for improving access to intercity transportation terminals in the region. The evaluation indicates that, although more direct and faster public limousine and transit services will produce a modest increase in mode share, shared-ride taxi services offer a better compromise between the low cost of public transportation and the convenience of the private automobile and conventional taxi service.

Municipal transportation and planning authorities are often asked to formulate policies and provide service for access to intercity transportation terminals. Indeed, the need for access to major terminals such as airports and railroad stations is often used as a justification for major investments in roadways and rapid transit facilities. This paper reports on the development of disaggregate demand models for choice of access mode to intercity transportation terminals [see also Rassam and others (1)]. We will give separate models for air, rail, and intercity bus terminals, stratified by personal and business trip purposes. In addition, we report on the application of the models to the prediction of the impact of several strategies for improving access to intercity terminals in Canada's national region (Ottawa-Hull and vicinity).

The results of our investigation, although compatible with professional intuition, have not been previously confirmed in the literature and therefore bear emphasis. In particular, our results show that the most important factors that determine use of public transportation modes for access to intercity terminals are the schedule frequency and the ease of access to pickup points for the service. Comparatively speaking, improvements in line-haul travel times are ineffective in increasing the use of public transportation modes for access to intercity terminals. Major investments in infrastructure to improve the travel time for public transportation to intercity terminals are unlikely to be justified on a cost-effective basis as compared to policies that promote taxi, limousine, and flexible bus service to the terminals. Shared-ride taxi offers a compromise between the low cost of public transport and the convenience of the private automobile and conventional taxi and could attract a significant share of passengers if provided in all parts of the catchment area of the region's terminals.

The results reported are based on travel surveys performed in April and May of 1978 on departing passengers at the airport, railroad station, and intercity bus terminals that serve the national capital region. The purpose of the study was to develop models for access mode choice for each of the terminals and to apply the models to test passenger preference for several strategies proposed for improving access to the intercity terminals. The technique used for calibrating the demand models was the familiar multinomial logit model, which has been extensively documented elsewhere (2, 3).

DATA

The set of access modes considered in this study are automobile driver, automobile passenger, taxi, limousine, and public transit (either regular or special bus). Not all these modes are available for each of the intercity transportation terminals we considered. For example, because of the absence of long-term parking at the bus terminal in Ottawa, drive alone was not considered to be a feasible alternative. Because only limited bus service to the airport is available, only 0.1 percent of survey respondents used this mode and it was deleted from our calibration data set for the airport. Neither the intercity bus terminal nor the railroad station is serviced by limousine.

Table 1 summarizes the variables used for our calibration.

Cost

Travel cost figures for automobile were estimated by adding an assumed operating cost of \$0.09/km to any long-term parking charge (i.e., \$2.50/day at the airport terminal). Total parking charges were determined by multiplying the per diem rate by the reported trip duration from the survey. Transit cost consisted only of the fixed \$0.55 flat fare. Similarly, limousine travel was figured at the set trip cost of \$2.75. The fare structure for taxi was \$0.80 plus \$0.44/km. In addition, a 10 percent surcharge was added to the taxi fare to account for standing charges and tipping.

Line-Haul Time

Line-haul travel times for automobile, taxi, and limousine were obtained from skim-tree values provided by the regional municipality of Ottawa-Carleton. These times were based on an average speed of 19 km/h plus an extra 3 min when passing downtown. Transit travel times were derived from current bus schedules and routes.

Waiting Time

Waiting time for limousine and transit was computed as half of scheduled headways to a maximum of 10 min plus the expected waiting time for transfer when required. No waiting time was assigned to the automobile and taxi modes.

Walking Time

Total walking time includes walking at both ends of the trip. Terminal walking times from park-and-ride lots were estimated as 7 min at the airport and 4 min at the railroad station. For driver-served passengers,

Table 1. Variables used for model development.

Abbreviation	Variable	Driver	Passenger	Taxi	Transit	Limousine	Comments
COST	Cost	X	X	X	X	X	Terminal dependent
WALK	Walking time	X	0	0	X	X	Terminal dependent
WAIT	Waiting time	0	0	0	X	X	Terminal dependent
LINE	Line-haul time	X	X	X	X	X	Terminal dependent
HBRES	Resident-home-based	1	0	0	0	0	Drive-specific variable
HBVIS	Visitor-home-based	0	1	0	0	0	Passenger-specific variable
NHBVIS	Visitor-non-home-based	0	0	1	0	1	Taxi-limousine-specific variable
BAG	Baggage	0	0	0	1	0	Transit problem = 1, other = 0
SEX	Sex	0	1	0	0	0	Female = 1, other = 0
ALT1	Alternative-specific dummy	1	0	0	0	0	Drive alone = 1, other = 0
ALT2	Alternative-specific dummy	0	1	0	0	0	Passenger = 1, other = 0
ALT3	Alternative-specific dummy	0	0	1	0	0	Taxi = 1, other = 0
PUR1[a]	Taxi purpose	0	0	1	0	0	Business = 1, other = 0
PUR2[a]	Transit purpose	0	0	0	1	0	Business = 1, other = 0

Note: X = calculated value; 0 = unaffected alternative.

[a]Alternative-specific dummy variables used for bus terminal only to account for trip purpose.

terminal times were assumed to be zero.

HBRES, HBVIS, NHBVIS

HBRES, HBVIS, and NHBVIS are dummy variables designed to account for the effects of residential status and trip origin. Three principal trends were identified in the analysis of these data and were incorporated into dummy variable definitions.

1. HBRES—Ottawa area residents who initiated their trips from home were more likely to drive alone than any other group of travelers.
2. HBVIS—Visitors to the area who began their trips to the terminal from a private residence were more likely to be automobile passengers.
3. NHBVIS—Visitors who began their trips to the terminal from a hotel or business location were more inclined to travel by taxi or limousine.

Baggage

The survey elicited information on the difficulty of using public transit because of baggage-handling problems. Passengers who responded that baggage considerations made the use of public transit difficult were identified by a dummy variable.

Sex

Female travelers showed a higher likelihood of making their trips as automobile passengers. This information was included in the model as a dummy variable.

Alternatives 1-4

Alternative-specific dummy variables were included in all models to capture the average influence of unobserved attributes for each mode.

Purpose 1, 2

Separate models were developed for business and personal travelers for the air and rail terminals. This was not possible for the bus terminal because of the small number of business travelers in the sample. The samples were therefore combined for the bus terminal only, and dummy variables were included to capture the alternative-specific effects of business trip purpose.

Besides the variables discussed above, extensive experiments were made with several other variables. Our failure to find any consistent influence of these variables is as instructive as our more positive results reported below, and these conclusions are briefly summarized here.

Household Income

In the initial analysis of our data, travelers from higher-income households (more than $20 000) appeared to have a higher propensity for using single-occupant automobiles or taxis as an access mode. Once we stratified our models by terminal and trip purpose, we were unable to identify any consistent effect of household income. Numerous experiments were made that treated household income interactively with the time and cost variables, as a dummy variable classification, and as an imputed wage rate. Our inability to find any consistent results indicates that income primarily determines choice of intercity mode and trip purpose. Conditional on these decisions, access mode choice is relatively free of income effects.

Automobiles per Driver

We initially felt that family competition for the automobile would be an important variable in determining the use of the automobile as an access mode. Once our models were stratified, however, the variable "automobiles per number of drivers in the household" lost virtually all explanatory power.

Transfers

Because of the inconvenience associated with transferring between vehicles, particularly with baggage in hand, a variable was defined equal to the number of transfers required to use public transportation. This variable was insignificant when waiting times were included in the model.

THE MODELS

Table 2 summarizes the stratifications used in defining the models calibrated. Because the number of business travelers who used the intercity bus terminal was small, business and personal travelers were combined for this terminal only. For the other terminals, separate models were estimated for each trip purpose.

The strategy used in the selection of variables for inclusion in the models was whether the coefficient of the variable had the predicted sign and whether the variable enhanced the predictive capability of the model. If variables passed these tests, they were included regardless of their statistical significance. These

Table 2. Model segmentation.

Terminal	Model Number	Market Segmentation	Sample Size	Modal Choice Set
Bus	1	Business and personal	556	Automobile passenger Taxi Transit
Rail	2	Business	96	Automobile driver Automobile passenger Taxi Transit
	3	Personal	222	
Air	4	Business	670	Automobile driver Automobile passenger Taxi Limousine
	5	Personal	198	

criteria were deemed a reasonable search procedure in an exploratory study such as this one.

Because transit service to the airport is virtually nonexistent, we were not able to include any of the time variables in either of the airport models. Without transit service as a standard of comparison, the time variables either showed no difference between the remaining modes (line-haul time) or were virtually dummy variables for one or another of the modes (walking and waiting time). Because we were able to obtain reasonable time coefficients for the other two terminals, these coefficients could be added to the air terminal models for predictive purposes.

Table 3 presents the most successful model calibrations for each terminal and market stratification.

Bus Terminal Model

This model contains a complete set of level-of-service variables (walking, waiting, line-haul travel time, and travel cost). Of the time components, travelers are most sensitive to walking time, display moderate sensitivity to waiting time, and are least sensitive to line-haul time. They also display moderate sensitivity to trip cost, as can be seen by examining the implied values of time for the model.

Bus Terminal	Values of Time ($/h)
Walking time	27.31
Waiting time	6.12
Line-haul time	3.97

In addition, transit handling difficulties are a strong negative influence on the use of public transit, and there are significant differences in modal choice probabilities based on trip purpose and being a home-based visitor to the Ottawa-Hull region.

Rail Terminal Models

These models also contain a full set of level-of-service variables, except that line-haul time has been deleted from the personal travel model because of a positive but statistically insignificant coefficient. The magnitudes of the time coefficients are compatible with those of the bus terminal model and also indicate that travel choices are most sensitive to walking time, are somewhat less sensitive to waiting time, and are relatively insensitive to line-haul time. The magnitudes of the cost coefficients are smaller than for the bus terminal. The values of time for business travelers who go to the rail terminal reflect a rather high sensitivity to time considerations as compared to travel costs.

Rail Terminal	Values of Time ($/h)
Walking time	74.24
Waiting time	41.06
Line-haul time	20.82

For personal travelers, the very small coefficient for trip cost would imply unreasonably high values of time if interpreted literally. A more conservative interpretation is that personal travelers are relatively insensitive to trip cost considerations as compared to travel time. We also note no significant difference between business and personal travelers in their sensitivity to the components of travel time. This result is at odds with the common statement that business travelers worry about time and personal travelers worry about cost.

As noted for the bus terminal model, the necessity to handle luggage is a strong deterrent to transit use for personal travelers. The failure of this variable to enter for business travelers probably reflects the shorter length of business trips and the correspondingly less luggage required. We also note that trip origin, residence status, and sex are important influences on modal-choice probabilities.

Air Terminal

As noted earlier, we were not able to obtain independent coefficients for any time variables for the airport models because of the absence of transit as a feasible airport access mode. Therefore, the only level-of-service variable that enters this model is trip cost. Although statistically significant for both the business and personal trip purpose models, the magnitude of the cost coefficient is small in absolute value and would imply implausible values of time if computed by using time coefficients from either the rail or bus model. In addition, as with the rail models, the cost coefficient for personal travelers is smaller than that for business travelers. The absence of the baggage variable from the air terminal models is due to the lack of transit as a feasible alternative. Otherwise, we see that residence status, trip origin, and sex are important determinants of modal choice. The alternative-specific dummy variables are almost all significant. We note that the difference in sign for the alternative-specific dummy variables as compared to the signs for the other terminals is due to the use of limousine as a base mode for the air terminal; transit was used as the base mode for the rail and bus terminals.

Cross-Model Comparisons

Comparisons between the five calibrated models in Table 3 yield interesting implications. First, we note the uniformity of the implications of the time coefficients in the three models where they could be estimated. In each case, walking time is the most important trip time component that influences modal choice, followed by waiting time; line-haul time is relatively least important in its effect. In addition, there is no significant difference in the travel time coefficients between business and personal travelers for the rail terminal or with travelers to the bus terminal.

In contrast, the cost coefficients show significant differences by terminal and by trip purpose. Cost considerations are most important for travelers to the bus terminal, are somewhat important for business travelers to the rail terminal, and have a small effect on

Table 3. Model calibrations.

	Model 1		Model 2		Model 3		Model 4		Model 5	
Variable	Number	t-Value	Number	t-Value	Number	t-Value	Number	t-Value	Number	t-Value
COST	-0.003 9	-4.09	-0.001 7	-1.28	-0.000 3	-0.30	-0.000 9	-4.19	-0.000 5	-2.17
WALK	-0.180	-1.63	-0.207	-1.75	-0.200	-3.52				
WAIT	-0.040	-1.80	-0.115	-1.06	-0.146	-2.60				
LINE	-0.026	-2.26	-0.058	-0.72						
HBRES					0.44	0.80	0.35	1.26	1.07	1.48
HBVIS	1.25	3.19			1.77	3.18	1.27	2.19	1.26	1.57
NHBVIS			1.03	1.76			0.65	2.44	2.92	3.60
BAG	-2.03	-6.35			-1.41	-3.00				
SEX			1.35	1.95	0.45	1.09	0.84	2.68	0.96	2.71
ALT1			-1.20	-0.91	-2.17	-2.46	1.15	4.18	1.54	1.77
ALT2	-1.64	-1.53	-3.82	-2.56	-3.96	-4.13	-0.39	-1.62	1.68	2.95
ALT3	-0.24	-0.24	-2.23	-1.44	-3.25	-3.57	1.43	9.01	1.20	2.30
PUR1	1.05	2.29								
PUR2	-1.18	-3.51								
ρ^2	0.184		0.229		0.129		0.203		0.190	
Percentage correctly predicted	65		63		58		61		63	

modal choice for air travelers and personal travelers to the rail terminal. These differences are probably due to the relative magnitude of terminal access costs as compared to the total cost of the intercity trip. Since intercity bus is the least expensive intercity mode, terminal access costs are a larger proportion of total trip costs and, therefore, play a larger role in determining access modal choice. On the other hand, air travel is the most expensive intercity transport mode, particularly because of the generally longer trip lengths and, therefore, terminal access costs are a smaller proportion of total costs and are relatively less important to the traveler. Rail represents an intermediate cost case between air and bus.

With respect to the fact that personal travelers exhibit smaller cost coefficients than do business travelers going to the same terminal, this result at first glance appears to be unintuitive. There are two considerations that make it reasonable, however. The first is that personal travel is generally of longer duration than business travel, so again the cost of access to the terminal is relatively smaller as a proportion of the total trip costs. In addition, personal travel often requires the carrying of more baggage, which is a strong dissuasion from using transit, as is evidenced by the baggage variables in two models. These considerations act together to suggest that personal travelers may be less sensitive to access costs and more concerned about baggage convenience than are business travelers. Our results suggest that trip cost considerations are not unimportant to business travelers, even if many of them will be reimbursed for out-of-pocket charges.

POLICY IMPLICATIONS

Walking and waiting times are more important factors than line-haul time in determining access modal choice. Therefore, policies and programs that encourage greater service frequencies and convenient access at the trip origin should lead to an increased market share for the mode under consideration. Policies that focus on faster line-haul travel time alone (e.g., exclusive transit lanes or expressways) will be less successful in achieving modal objectives.

Baggage-carrying considerations are a strong dissuasion from using public transit for personal travelers. Therefore, public transit is unlikely to compete satisfactorily with private automobile, taxi, and limousine for this market.

The use of an access model is strongly correlated with trip origin (home versus nonhome based) and residence status of the traveler (resident versus nonresident). Therefore, the best market for public access modes to intercity terminals will continue to be in the employment center and around hotels.

These conclusions are neither surprising nor controversial. They have definite implications about public policies for providing access to intercity terminals, however. The policies most likely to be successful will stress convenience, flexibility of service, and ease in baggage handling. Capital investments to improve transit line-haul times to intercity terminals are unlikely to attract a significant number of new passengers.

ACCESS TO INTERCITY TERMINALS IN THE NATIONAL CAPITAL REGION

The models were developed to predict access mode choice to the intercity terminals in Ottawa-Hull for different arrangements of access services and terminal locations. This is a key element in a study currently being done to evaluate alternative strategies for improving access to intercity services in the national capital region.

Current Situation

The location of the existing bus, rail, and air terminals in Ottawa-Hull and the access services provided are illustrated in Figure 1. There are no terminals of significance in Hull and nearly all passengers who originate in Hull and other areas north of the Ottawa River must use the Ottawa terminals. Currently all major intercity bus and rail services to and from Ottawa-Hull, even those to points east of the region, radiate from Ottawa on the south side of the Ottawa River because of superior highways and rail corridors. Only a few local bus and rail services pass through Hull and pick up passengers at satellite stops.

All of the Ottawa terminals can be reached by transit from most of the urban area although it may require a prolonged trip, involving several connections, especially from fringe areas. Moreover, the bus terminal is located one city block west of the major north-south arterial that carries several of the busiest north-south transit routes. These transit services do not deviate to serve the bus terminal except on weekends, so bus passengers must walk from the nearest transit stop to the bus terminal. The transit route that serves the airport is not an express service and is intended for airport employees.

Limousine services are provided between the major hotels in the central business district (CBD) and the rail station and the airport. The limousine service to and from the rail station crosses the Ottawa River and terminates in the Hull CBD, but the airport limousine service terminates in the Ottawa CBD and does not serve Hull.

Long-term parking is provided at the rail station and airport but not at the bus terminal. The nearest off-street pay lot for the bus terminal is located several blocks away.

Future Situation

A transitway is to be developed to serve the Ottawa urban area south of the Ottawa River. Initially, this transitway will be a busway, but it is designed to be upgraded to light rail transit at some future date. The transitway will not pass through any of the intercity terminals. During the planning phase, the possibility of running the transitway through the rail terminal was evaluated but rejected on the basis of the low number of trips generated by the rail terminal (relative to other trip generators to be served) and the high cost of traversing the rail lines in the terminal area. The transitway alignment selected is approximately 1.3 km west of the rail terminal. Alternative transit corridors, which would have located the transitway nearer to the bus terminal, were rejected during the planning phase due to a combination of demand, cost, environmental, and other practical considerations. The southeastern end of the transitway terminates approximately 2 km from the airport, near the limits of the urbanized area, and it was not considered feasible to extend the transitway to the airport.

It is planned to operate only regular transit services on the transitway. Approval would be required before the limousine services between the CBD and the rail station and airport would be permitted.

OPTIONS FOR IMPROVING ACCESS

The options recommended for improving access to the bus, rail, and air terminals are listed below

Terminal	Option
Bus	More direct transit
	New limousine
	Shared-ride taxi
	Optimum terminal locations
Rail station	Reduced transit and limousine line-haul times
	Shared-ride taxi
Airport	Extended limousine service and reduced line-haul times
	Shared-ride taxi

Bus Terminal

Rerouting the high-frequency north-south transit services through the bus terminal is not practical because it would impose an unacceptable time penalty on the majority of transit users. However, a bus service that runs east-west along the crosstown expressway has been introduced recently and will connect with all major transit lines that radiate from the CBD at each interchange. The new bus service also serves the bus terminal that is located near one of the expressway interchanges. The service will allow a large number of bus passengers to avoid having to travel first to the CBD in order to connect to the bus route that passes nearest to the bus terminal.

A limousine service similar to the rail station limousine, which will link the bus terminal with principal points in the Ottawa and Hull CBDs, is being considered.

Shared-ride taxi is an alternative access mode that satisfies many of the requirements to which passengers are particularly sensitive when selecting the access mode. Compared to exclusive-ride taxi, passengers must trade off cost versus a certain amount of delay (if

Figure 1. Current situation.

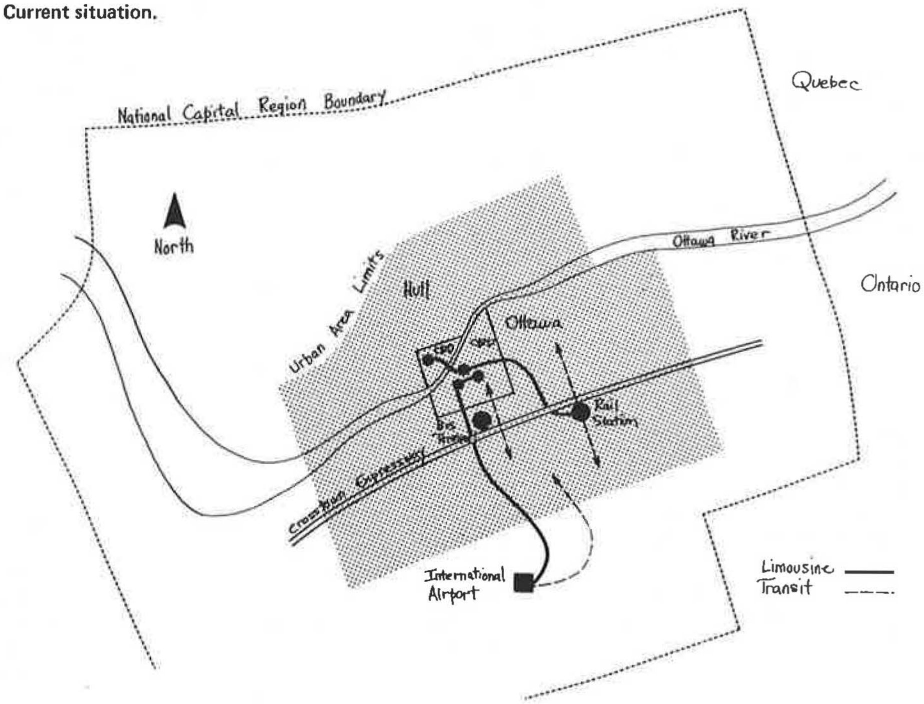


Table 4. Access mode shares for national capital region.

Terminal	Option	Automobile (%)	Taxi (%)	Limousine (%)	Transit (%)	Shared-Ride Taxi (%)
Bus	Current	34	37		29	
	More direct transit	34	31		35	
	New limousine	32	28	9	31	
	Shared-ride taxi	21	18	5	21	35
Rail	Current	44	42	5	9	
	Reduced transit and limousine line-haul time	43	39	5	13	
	Shared-ride taxi	30	28	3	9	30
Air	Current	45	47	8		
	Extended limousine and reduced line-haul time	43	42	15		
	Shared-ride taxi	41	34	8		15

not picked up or dropped off first), convenience (need to prebook a trip to the terminal), and privacy (however, Ottawa-Hull is much more socioeconomically homogeneous than most large North American cities). Most importantly, shared-ride taxi services can be provided to all areas in the region.

Another means of improving access is to relocate or add new bus terminals to bring the intercity trip end nearer to passenger origins and destinations. This can be done in a combination of ways, including moving the bus terminal to a more central location (ideally on the transitway) or establish satellite bus terminals to pick up or drop off passengers in fringe areas en route. Another possibility would be to begin and terminate some of the intercity bus routes in Hull with an en route stop at the main Ottawa bus terminal to pick up and drop off passengers.

Rail Station

The rail station is served by several transit routes that stop directly at the station and by a limousine service to hotels in the Ottawa and Hull CBDs. Access transit has been improved by the same crosstown transit service that was introduced recently at the bus terminal. When the transitway is developed, the line-haul portion of limousine and transit trip times are expected to be reduced significantly.

Airport

The limousine service that currently links the airport with the major hotels in the Ottawa CBD could be extended to serve the new hotels and commercial complexes now being developed in the Hull CBD. This service would be a simple extension to the existing service during off-peak periods and a separate direct service in the peak periods. When the transitway is developed, these services could operate along this facility to avoid road congestion and reduce line-haul time.

MODE CHOICE FOR OPTIONS

The mode choices of different categories of passengers represented in each of the models (e.g., home-based residents on a personal trip, non-home-based visitors on a business trip) were determined by dividing the national capital region into zones, determining representative values of the explanatory variables for each model and zone, and then applying these values to the models. The overall mode share for the region was calculated as the average mode share weighted by the number of passengers in each category and zone. These results are presented in Table 4.

Bus Terminal

The current share of transit to the bus terminal is fairly high at 29 percent because a large number of bus passengers originate in the center of the city or are transit dependent. The provision of more direct transit services will increase transit's share from 29 to 35 percent; most of this increase will come from taxi. New limousine services to Ottawa and Hull CBDs will attract more passengers from taxi and a few from automobile and (improved) transit.

Assuming that shared-ride taxi services would be operated as specified, they would be highly favored by passengers and could attract a large number of passengers away from the other modes (including improved transit and new limousine services) to become the predominant mode.

The results of relocating or adding new bus terminals are inconclusive and are not presented in Table 4. It was found that changes in the terminal location tend to affect all modes more or less equally and, therefore, the mode shares remain almost unchanged or change in favor of taxi, whose costs are reduced relative to transit. The mode share is also heavily dependent on whether or not long-term parking is provided at the new bus terminals.

Rail Station

Reduced transit and limousine line-haul times for using the transitway will increase transit's mode share from 9 to 13 percent but will not affect the limousine's share. The transitway affects transit line-haul times from a large number of zones within Ottawa, but it produces only a minor reduction in the limousine line-haul time. Shared-ride taxi services are expected to attract a significant number of passengers from all modes.

Airport

Extended limousine services to the Hull CBD and reduced line-haul times due to using the transitway for part of the trip will increase limousine's mode share by almost 100 percent. Shared-ride taxi is not expected to attract as large a share as at the bus and rail terminals.

CONCLUSIONS

The majority of intercity passengers in Ottawa-Hull currently use private automobile or taxi for access to the intercity terminals; however, the evaluation demonstrates that there are several ways in which public transportation services can be improved and thereby attract significant numbers of passengers. These results bear out the policy implications previously drawn from the developed models.

Improvements in transit services that provide more direct access to the terminals by reducing walking and waiting times and the number of connections can produce further modest gains for transit as an access mode. The difficulty is, of course, in providing a high level of service to the intercity terminals from all points in the terminals' catchment area.

Limousine services that provide express service between hotels and other central points in the CBD and specifically cater to passenger baggage requirements can attract a majority of passengers whose origins and destinations are in the CBD. These passengers can form a significant portion of the total trips in the catchment area.

Shared-ride taxis offer a compromise between the lower cost of public limousine and transit services and the convenience and speed of private automobile and taxi. They also provide service to nearly all parts of the catchment area of the Ottawa-Hull terminals. As a result, the evaluation estimates that shared-ride taxis can capture a substantial share of bus and rail passengers and a smaller share of air passengers. They offer a clear alternative to existing public transport services.

ACKNOWLEDGMENT

The research reported was performed under contract to the National Capital Commission. The views and conclusions are ours and not those of the National Capital Commission or of the World Bank or its affiliated organizations.

REFERENCES

1. P. Rassam, R. Ellis, and J. Bennett. The n-Dimensional Logit Model: Development and Application. HRB, Highway Research Record 369, 1971, pp. 135-147.
2. D. McFadden. Theory and Practice of Disaggregate Demand Forecasting for Various Modes of Urban Transportation. Paper presented at the Seminar on Emerging Transportation Planning Methods, U.S. Department of Transportation, Daytona Beach, FL, 1976.
3. B. Spear. Applications of New Travel Demand Forecasting Techniques to Transportation Planning. U.S. Department of Transportation, 1977.

Publication of this paper sponsored by Committee on Passenger Travel Demand Forecasting.

Use of the Gravity Model for Pedestrian Travel Distribution

G. Scott Rutherford*, G.S. Rutherford and Associates, Washington, D.C.

Knowledge of pedestrian travel behavior is very important to attempts to improve congestion problems in central business districts. This paper describes the results of the use of a traditional gravity model for predicting pedestrian trip distribution. The model is calibrated by using a data set from downtown Chicago. The results indicate that the traditional gravity model closely reproduces the characteristics of pedestrian trip distribution and might be a useful tool in the analysis of downtown travel.

A great deal of discussion is now taking place on how to improve the central business districts (CBDs) of our major cities. Many proposals that are being evaluated and implemented deal with malls, personal rapid transit, downtown people movers, and sky walks. All of these systems have implications for the mobility of people in the CBD. Since much, if not most, of the CBD mobility is provided through pedestrian journeys, these proposals will certainly affect the number and length of such journeys and compete with them for patronage. An understanding of pedestrian trip distribution is, therefore, necessary in order to evaluate the potential impact of some new suggestions for the CBD.

The purpose of this paper is to review the calibration and application of a standard gravity model for a data set collected in Chicago in 1963 (1). This data set offers more than 10 000 origin-destination interviews in the Chicago CBD and presents the opportunity to test the gravity model on pedestrian travel behavior.

THE PEDESTRIAN SURVEY

The pedestrian survey was conducted by the Chicago Area Transportation Study (CATS). The interviews were conducted by people from various city departments in Chicago's downtown, known as the Loop, due to the elevated transit line that defines it. The survey was taken during the period from 7:00 a.m. to 7:00 p.m.; each interviewer collected a predetermined number of interviews. Interviews were collected randomly along 98 stations on one side of a street about three blocks in length for each hour in the time period.

The survey collected data for each station by hour, including purpose of trip, direction of travel, and whether the respondent was coming from work. The interviewer also obtained origin and destination addresses. The total number of people interviewed was 11 632. The sample rates for each station were based on pedestrian volume counts done by regular traffic counters the previous year.

The sampling techniques employed resulted in a sample that was uniformly distributed across the Loop area (i.e., an approximately equal number of interviews at each station). This distribution has two beneficial effects from a statistical standpoint.

1. It assures that blocks with low volumes on the edge of the Loop are not ignored (if a uniform sample were taken, very few trips from low-volume areas would be sampled, thus producing a possible bias), and

2. When the sample is expanded, the tendency will be to equalize the percentage of standard error of expansion across blocks (if a certain sample percentage were taken from a low-volume location and an equal percentage from a high-volume location and both expanded, the low-volume location expansion will have a larger percentage of standard error than will the high-volume location).

By surveying larger percentages in low-volume areas and smaller percentages in high-volume areas, the tendency will be toward an expansion that has smaller variance in the percentage of standard error than if a uniform sample were taken for the entire area. The problem of even getting a uniform sample, should one want it, would be nearly insurmountable in sidewalk interviews in a location such as Chicago's Loop.

This method of sample expansion and an analysis of the pedestrian travel characteristics were presented previously (2).

THE GRAVITY MODEL

The gravity model is calibrated by using the observed trip-length distribution to adjust model parameters. Analysis performed on the Chicago data (2) indicated that these data yield distributions of trip length that not only compare with other cities fairly well but could also, if necessary, be described with a simple negative exponential relationship. In short, it was apparent that the data to support the calibration of the gravity model were complete (i.e., trip-length distributions) and showed substantial promise.

The gravity model concept derives its name from Newton's law of gravity that states that the attraction between two bodies is directly proportional to their mass (or amount of attractions) and inversely proportional to some function of the distance between them. The form of the gravity model is as follows (3):

$$T_{ijp} = P_{ip} A_{jp} F(t)_{ijp} \Big/ \sum_j A_{jp} F(t)_{ijp} \qquad i, j = 1, 2, 3, \ldots, n \tag{1}$$

where

T_{ijp} = one-way trips from block i to block j for purpose p,
P_{ip} = trips produced at block i for purpose p,
A_{jp} = trips attracted to block j for purpose p, and
$F(t)_{ijp}$ = friction factor based on the travel distance between block i and block j for purpose p (ordinarily travel time would be used but since the level of service for walking is nearly constant, it is easier computationally to substitute distance, which is then directly proportional to time).

The premise of the gravity model is that trip interchanges can be estimated based on the relative attractiveness and impedance between the blocks in question. For this application, attractiveness is measured by the ratio of the number of trips attracted to block i for purpose p versus the total trips to all blocks for purpose p:

$$\text{Attractiveness of block j for purpose p} = A_{jp} \Big/ \sum_j A_{jp} \tag{2}$$

The impedance is calculated similarly in the following fashion:

$$\text{Impedance between block i and j for purpose p} = F(t)_{ijp} \Big/ \sum_j F(T)_{ijp} \tag{3}$$

Mathematically, $F(t)_{ijp}$ is a complex function but, in general, is proportional to a function of the inverse of the distance between blocks raised to a power, as is shown below:

$$F(t)_{ijp} \propto f_1 (T/d_{ij}^{f_2(n)}) \tag{4}$$

where d_{ij} = the distance between blocks i and j and $f_2(n)$ = a factor that depends on the trip purpose and trip length. Mathematical description of F(t) is quite complex, so it is generally described as a discrete distribution of numbers and not as a mathematical expression. The F(t) values are generally referred to as friction factors or impedances; however, as is indicated by the equation, the higher F(t) is, the more trips will be assigned to the i-j interchange. The F(t) values might better be referred to as travel propensities rather than frictions; however, to avoid confusion, this paper will continue to refer to F(t) as a friction factor or impedance.

In summary, the gravity model is based on very simple intuitive assumptions that deal with spatial separation of points and rewards or benefits available at these points. It has been applied widely in transportation planning and many examples of its use are available in the literature (4, 5).

Calibration of the Gravity Model

The calibration method adjusts the F(t) values iteratively until the trip-length distribution calculated by the model on the basis of distances between blocks is essentially equivalent to the observed trip-length distribution. The equivalence point is arbitrary and depends on the judgment of the person doing the calibration; however, a criterion of ±5 percent for the difference between observed and calculated mean trip length for each purpose has been suggested (6). This calibration technique is discussed further elsewhere (3, 4).

The computer formulation of the model first reads in the necessary inputs for the calibration phase; these are

1. The observed trip-length distribution for each purpose,
2. Initial estimated for F(t) values for each purpose (these can be based on prior knowledge of simply set equal to one),
3. Observed productions and attractions by purpose for all blocks in the area being studied, and
4. A matrix containing the distances between all blocks.

The model then distributes the trips based on the previously described equation for each purpose for as many iterations as the user specifies. During each iteration, trips are distributed over all blocks, new trip-length distributions are calculated, and new F(t) values are adjusted on the basis of the length distributions.

These F(t) values then serve as input to the next iteration. Once the calculated trip-length distribution is sufficiently close to the observed distribution, the model is then considered to be calibrated. Again, this point of calibration is determined by the planner based on judgment. The final calculated values of F(t) for each purpose are then ready to be used for the trip-distribution forecasting process. The calibration process is solely to obtain the F(t) or impedance function used in forecasting with the distribution model.

Calibration Results

To demonstrate how the model is stabilized (i.e., how the calculated trip distribution approaches that observed), Figure 1 shows the change in value of calculated trip-length distribution over five iterations. As one can see, the model rapidly approaches a stable point. This is somewhat dependent on the initial F(t) values assumed; should one use an initial value of 1.0, the process might take more iterations. This application began with a set of friction factors that have a slope similar to those found appropriate in other trip-distribution modeling efforts. The final, calibrated set of friction factors, however, was substantially different from the initial set.

The result of the full calibration can be analyzed by comparing the final trip-length distributions with the observed trip-length distributions. This comparison is best demonstrated in Figure 2, which shows the total observed distribution along with the calibrated distribution; this figure shows near perfect correlation. Another comparison is made in the table below, which gives observed and calculated mean trip lengths by purpose. Again, close agreement is apparent.

Purpose	Mean Trip Length Observed	Mean Trip Length Calculated
To work	296	299
To home	335	311
To shop	274	274
Work-related business	299	300
Personal business	299	297
Social-recreation	247	244
All purposes	296	292

These curves are the result of the three initial iterations of one purpose to gain approximate values plus four additional iterations of each of the six purposes. It should be pointed out that the purposes that have

Figure 1. Percentage of trips by distance for five calibration iterations—"to work" trip purpose.

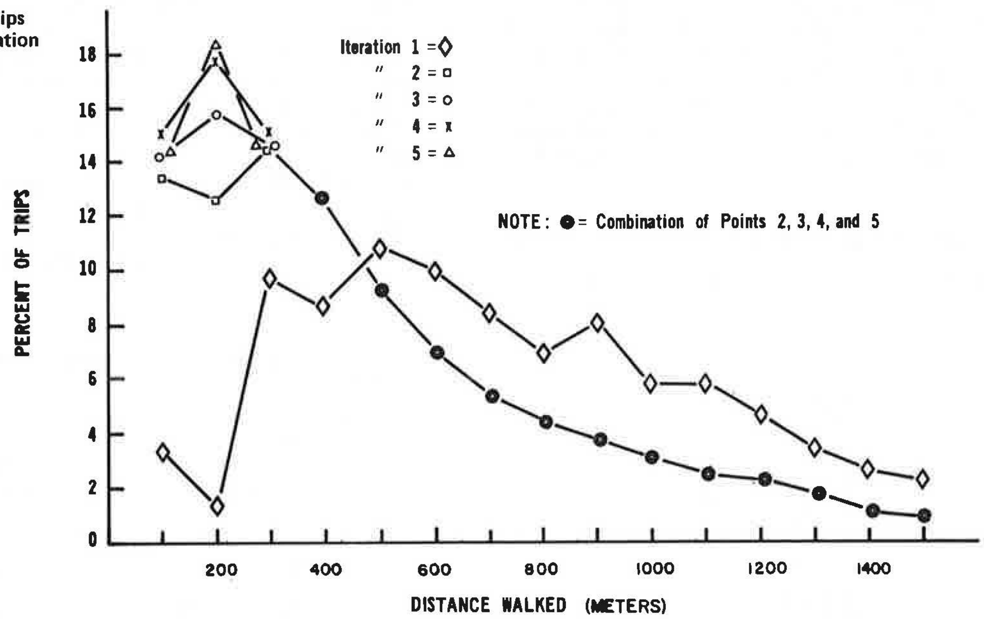
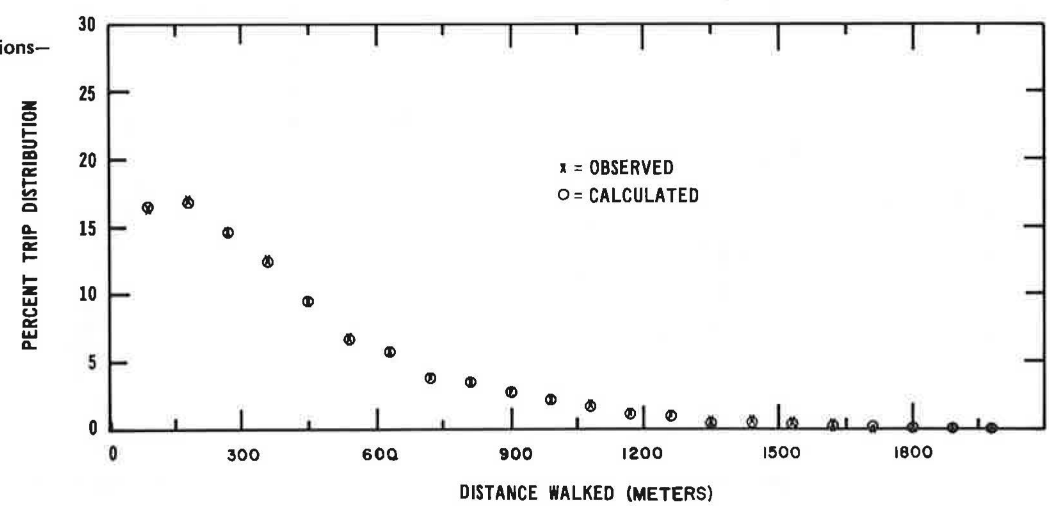


Figure 2. Observed and calibrated length distributions—all purposes and all trips.

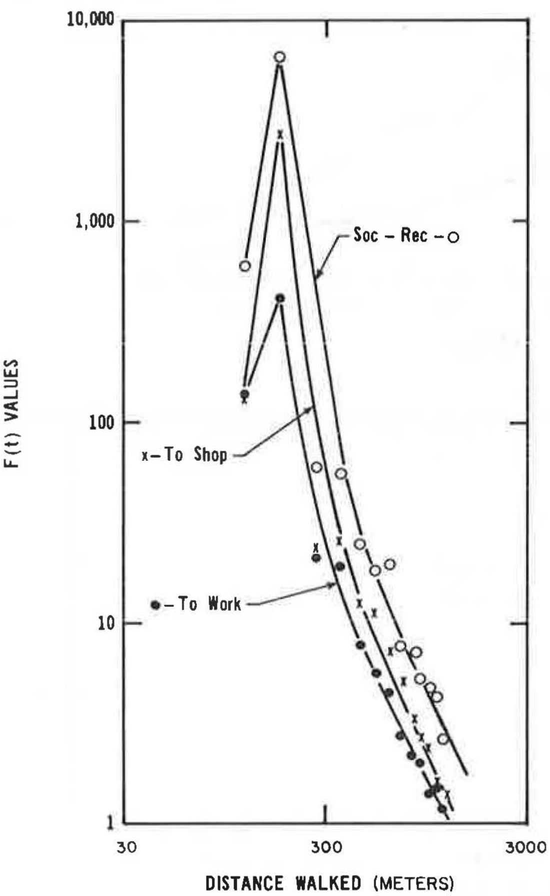


length distributions similar to "to work" (which were to home, work-related business, and personal-business) calibrated very rapidly, generally after one additional iteration over the initial three; it was necessary, however, to perform four additional iterations in order to establish stabilized F(t) values for purposes "to shop" and "social-recreation". It is evident from this experience that one can save a great deal of calibration time by starting with a realistic set of friction factors. Figure 3 shows the set of calibrated F(t) values for three representative trip purposes on a log-log scale. As expected, the shopping and social-recreation trips have more peak distributions, which indicates a propensity for shorter trips.

After the F(t) values have been calculated, one can plot and estimate a curve based on the points. From this curve, new F(t) estimates can be made that ensure that the values will decrease monotonically; this was not done in this study since the calibration values were essentially monotonically decreasing without further adjustment.

After the tables and curves are reviewed and, recalling that the basis for calibration was the observed trip-length distribution, one can conclude that the model has been successfully calibrated. A better test of the model is its ability to reproduce the observed trip interchanges between blocks. One check is available at this point, and that is to compare the friction factors calculated from the model with those found in a study done in Toronto (6). By using an average walking speed for downtown Chicago of 1.386 m/s (4.55 ft/s) so that the results here can be plotted on the Toronto study graph, the F(t) values for the "to home" trip (due to their association with transportation facilities) are plotted along with Toronto's values associated with terminals and appear in Figure 4. Keeping in mind that Chicago's "to home" values include trips to all modes, the Chicago values are generally within the curves that describe Toronto's envelope for trips to transportation facilities. This shows that the curves are generally similar and increases confidence in the calibration of the gravity model for Chicago.

Figure 3. Log-log plot F(t) versus distance.

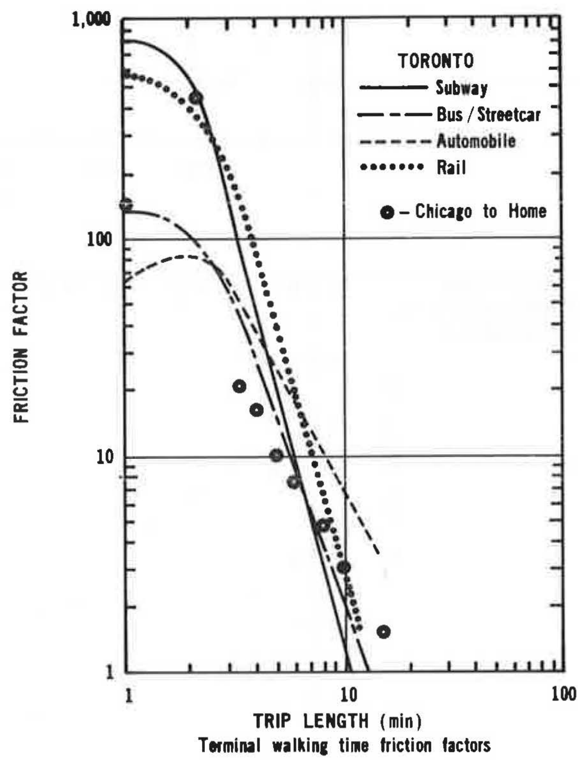


Possible Improvements to the Calibration Process

Numerous factors influence the trip-length distribution that was used as a basis for calibration of the gravity model. These factors include purpose of trip, time of day, employment status, and area of trip origin. It seems likely that the inclusion of these factors in the calibration process would result in a better description of travel. The inclusion of any of these items in the calibration process is quite easy; all one has to do is run the calibration separately for each factor in the same manner as was done for the six purposes. For example, one might decide to calibrate a separate model for various CBD areas, for employees and non-employees, and for the six purposes. Should this be done, the model would undoubtedly be improved, but the cost of calibration would rise significantly and problems involved in forecasting these disaggregate values in the future would be difficult to surmount. The most reasonable adjustment to the model (for Chicago) would be to subdivide the trips by employee group and by two areas (Loop and fringe). This scheme, although it includes many factors found to affect trip length, would be less expensive to calibrate than the previous suggestion.

These extensions were not included with the current

Figure 4. Chicago F(t) values compared to those of the Toronto study.



research for several reasons: (a) it was felt that available resources could be better used in extending the applications of the model rather than fine tuning it for downtown Chicago; (b) once calibrated for the factors listed above, the model then used for the distribution process will again be more expensive since the distribution must be done for each trip group (a typical trip group might be trips by Loop employees for the purpose of work); and (c) it was felt the model was generally valid based on its overall calibration. Therefore, the model appears to be calibrated satisfactorily with respect to the observed length distributions, and the results compare favorably to another pedestrian study.

Figure 5. Observed destinations—all purposes.

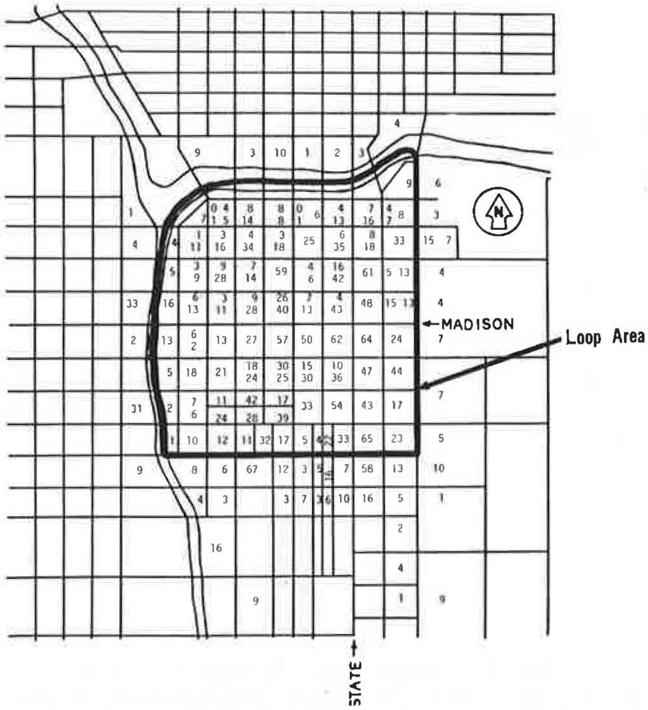
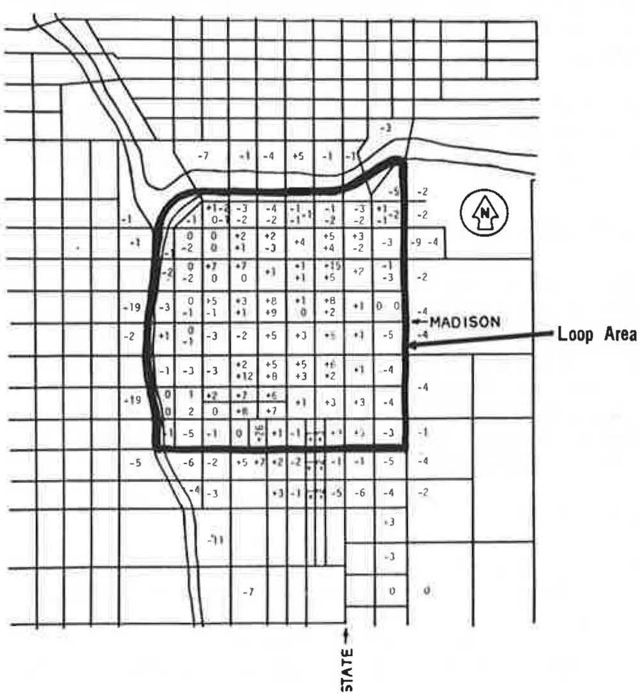


Figure 6. Model destinations minus observed destinations.



APPLICATION AND EVALUATION OF THE GRAVITY MODEL

In order to evaluate the performance of the gravity-distribution model, it is necessary to see whether or not it can reproduce the observed trip interchanges. The basis of evaluation for the calibration of the model was the reproduction of trip-length distributions; it did that nearly perfectly. The task at hand is to evaluate the model's ability to distribute trips to the blocks in the Loop in a similar manner as they were observed (i.e., Can this model send trips to blocks in the same numbers that were surveyed?).

The model results can best be presented by comparing the observed destinations per block with the destinations predicted by the gravity model (summed over all purposes). This comparison is made by observing Figure 5, which shows observed destinations, and Figure 6, which shows the difference between the calculated trip destinations and destinations observed. Agreement is fairly close; however, one can see that, in general, the model distributes too many trips to the central area and too few to the fringe. This indicates that the model cannot distribute trips adequately to the fringe areas. This is not a surprising result since the same set of F(t) values was used for fringe trips as for central trips. Further analysis showed that trips that originate in the fringe were much longer, since only external trips were surveyed and internal trips ignored, and would thus have different F(t) values. This can be seen in Figure 7, which shows the difference between length distribution for the Loop and fringe.

In particular, the commuter railroad stations located in the fringe did not get an adequate number of trips distributed to them. The observed destinations to the blocks west and south of the Loop with commuter stations totaled 70 000 trips, and the model only distributed a total of 21 000 trips. This may indicate that special generators, such as those on the periphery, must be treated differently.

The fringe area as a whole had a total of about 220 000 trips according to the observed data analysis, whereas the model distributed about 56 000 trips or only one-fourth of the observed total; the missing trips were distributed to the Loop area, which caused the totals there to be larger than observed. An adjustment of some sort is clearly needed and it seems clear that, as in models of vehicle trips, external trips must be modeled separately.

Another comparison can be made by relating the observed and distributed trips in Table 1. This table lists the distribution error by categories that represent the magnitude of trip attractions. One would expect more error for blocks that have large magnitudes and smaller errors for those with less (i.e., the percentage of error should be nearly constant over all the blocks). The results viewed from Table 1 are somewhat inconclusive since blocks in the 3000-9000 range had a larger percentage of error than other blocks. This probably reflects the poor distribution to the fringe blocks, which generally fell into this range. The error in blocks with larger values was quite small.

It seems that the gravity model produced a reasonable replication of the observed trip attractions, except for the fringe areas. It is important to note that these results were obtained without any special adjustments to the basic theoretical equation. In practical planning

Figure 7. Trip-length distributions for Loop and fringe—all purposes.

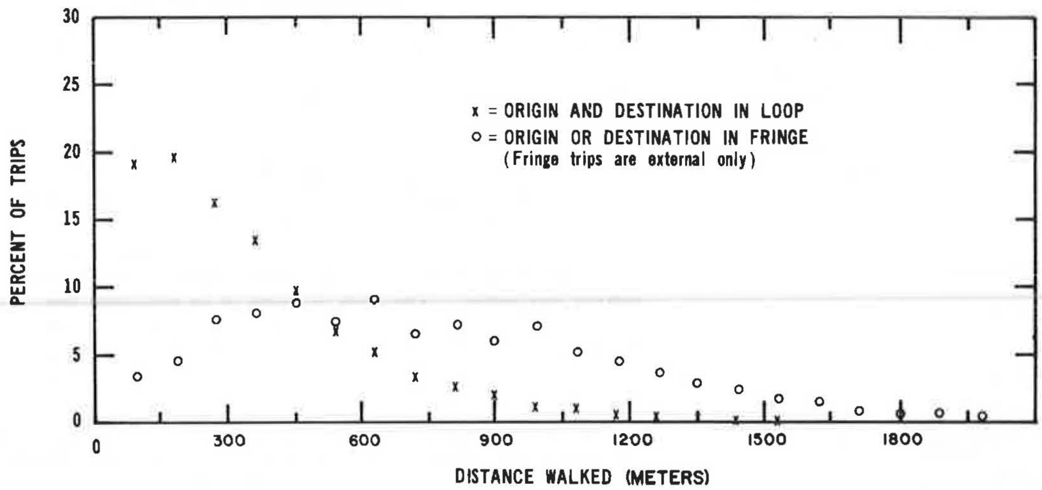


Table 1. Error for distribution by volume range.

Block Attraction	Error[a] ±1000	±1000-2000	±2000-5000	±5000-10 000	±10 000	Total
0-3 000	35	12	5			52
3 000-6 000	6	12	14	1		33
6 000-9 000	5	5	17	1		28
9 000-12 000	2	3	3	5		13
12 000-20 000	9	2	10	5	2	28
20 000-40 000	3	7	9	6	5	30
40 000	1	3	9	3	–	16
Total	61	44	67	21	7	200

[a]Error = observed-distributed trips.

efforts, such models usually go through a considerable amount of fine tuning (i.e., parameter adjustment) before reproducing observed results within reasonable limits.

Many applications of the gravity model for prediction of vehicular travel have used an iterative approach to ensure that the number of trips attracted to each zone is equal to the initially estimated trip attraction. The application of that approach in this research might have eliminated some of the problems discussed above. However, there is considerable uncertainty in the measured trip attractions and productions. Forcing the model to conform to the measured values of attractions, therefore, does not have strong appeal. (Productions, by definition, conform to the initial survey estimates.)

In a forecasting mode, some applications of the gravity model to vehicular travel prediction have foregone the step of balancing attractions on the grounds that, indeed, the gravity model is about as likely to give a good estimate of attractions as is the trip attraction model itself. This is a rather indirect way of letting accessibility assist in the determination of trip attractions: the gravity model attraction estimates are determined both by accessibility provided by the transport system and by the initial attraction estimated. Given the uncertainty in the input data and in spite of the lack of knowledge about accessibility-trip generation relationships, this latter approach was adopted for this research.

CONCLUSION

This paper has shown that pedestrian trip distributions are predicted fairly accurately by using a standard gravity model, and with a few simple modifications the accuracy can be greatly improved. This model outputs block-to-block interchanges that could be used as a basis to begin testing the impact of various CBD improvements, such as downtown people movers, which would compete with walking for patronage. A distribution model is central to any transportation-planning analysis. This study demonstrates that the gravity model (an institution in itself) can be easily adapted to pedestrian travel, and, therefore, provide an alternative framework for analyzing improvements to travel in CBDs.

ACKNOWLEDGMENT

This work is based on dissertation research made while I was a student at Northwestern University in Evanston, Illinois. I am deeply indebted to Joseph L. Schofer for a great deal of guidance and advice. Funding for this research was provided in part by Northwestern's Urban Systems Engineering Center, the National Science Foundation, and through a Royal E. Cabell Fellowship.

REFERENCES

1. J. D. Orzeske. How to Collect Data on Special Generators. Chicago Area Transportation Study, 1967.
2. G. S. Rutherford and J. C. Schofer. Analysis of Some Characteristics of Pedestrian Travel. TRB,

Transportation Research Record 605, 1976, pp. 29-34.
3. Calibrating and Testing the Gravity Model for Any Size Urban Area. Bureau of Public Roads. U.S. Department of Commerce, Washington, D.C., 1965.
4. K. E. Heanue and C. E. Pyers. A Comparative Evaluation of Trip Distribution Procedures. HRB, Highway Research Record 114, 1966, pp. 20-37.
5. Urban Transportation Planning. Federal Highway Administration, March 1972.
6. J. F. Morrall and others. Traffic Prediction Models for Central Business District Planning. Proc., Fourth Conference, Australian Road Research Board, Melbourne, 1968.

Publication of this paper sponsored by Committee on Passenger Travel Demand Forecasting.

**G. S. Rutherford conducted the research on which this paper is based while at Northwestern University.*

Population Segmentation in Urban Recreation Choices

Peter R. Stopher and Gökmen Ergün, Department of Civil Engineering, Northwestern University, Evanston, Illinois

The paper describes an investigation of various segmentation bases for capturing the behavioral differences in urban recreation demand. The analysis and evaluation of the segmentation bases were mainly achieved through the calibration of disaggregate quantal choice models (by using the multinomial logit technique) for each population segment and statistical comparison of these models and their estimated coefficients. After a preliminary elimination, three segmentation bases were selected for detailed evaluation: stage in the family life cycle, recreation-activity attractiveness, and geographic location. For each of the categories of these bases, a recreation-activity choice (a detailed trip-purpose) model was calibrated. These segment models were then compared with the pooled model both in terms of the overall goodness of fit and in terms of the differences in their coefficient estimates. Each of the segmentation schemes that was tried revealed significant differences and most of these differences bear plausible relation to the segmentation variables. Significant behavioral variations, which may result from differences in tastes, motivations, and personalities, may be captured through population segmentation.

Recreation is a broad and diverse area of human activity, encompassing a wide range of pursuits. Increased demand for participation in these activities creates, in varying degrees, increased use of transportation facilities. Visits to national parks alone have increased at an annual growth rate of about 7.5 percent in the period from 1957 through 1976 (1, 2). This is considerably higher than the population growth rate during the same period and also implies a very considerable growth rate in the consumption of fossil fuels for recreation activities.

The concern of the research in this paper is urban recreation and cultural activities. Most work on demand for recreation has concentrated on nonurban recreation and vacation activities (3-5), although many government units in urban areas are becoming increasingly concerned about issues of policy and investment in recreation facilities. If in the future transportation fuels are less available or the costs of such fuels are increased significantly, urban recreation facilities will probably receive the impacts of resulting changes in travel behavior. This will occur because travel to recreation is one type of travel most likely to be reduced or diverted from far sites to near ones (urban) in the event of high price or low availability of fuel. From a policy viewpoint, freedom to participate in a wide range of recreation activities may be considered to be one element of the high living standards enjoyed in the United States and Canada. Thus, substitution of local (urban) recreation activities for long-distance ones may be one way to prevent energy scarcity or high prices from eroding living standards.

This research introduces market segmentation as a means to understand and analyze recreation travel behavior. However, the paper deals only with recreation-activity choice (i.e., a detailed trip purpose) for a variety of reasons:

1. The reasons why people engage in recreation activities are much more complex, diverse, and numerous compared to other trip purposes. Recreation activities can be undertaken simply for fun or to fulfill various other complex psychological matters such as needs, motivations, and values. Hence, the consequences of recreation travel can only be understood after recreation behavior, per se, is understood. This is perhaps more crucial than for any other trip purpose.
2. Recreation is a gross trip purpose. The activities covered include a wide variety of activities and widely varying needs for travel, ranging from skiing to watching television. Thus, activity choice becomes an important issue, especially for the resulting travel implications.
3. We believe that the differences in individual tastes, motivations, and perceptions are the greatest influences on activity choice and, hence, concentrating on this choice can show the effects of segmentation more clearly.
4. The passage to recreational travel demand from recreation demand is a relatively trivial matter.

The basic demand-modeling hypotheses, which are described elsewhere (6), assume that both characteristics of the individual and attributes of the alternatives affect the choice process. Several mechanisms may be argued for the process by which these characteristics influence choices. One possibility is to use these characteristics as linear, additive terms in the utility function of the recreation activities. In this case, the effect of the characteristics is marginally to add to or subtract from the utility of activities and to affect the relative tastes of individuals for different attributes. Watson and Stopher (7), inter alia, argue that this is not the most appropriate manner in which to portray the effects of these variables. Rather, they argue that the appropriate manner to enter the variables is to use them as a basis for population (market) segmentation. This has also been argued extensively as a basis for improving the capability and responsiveness of individual choice models (8, 9).

The data for this research consist of 812 cases from two suburbs of Chicago: Evanston and Des Plaines. They provide information on the perceptions of attributes, availabilities, attractiveness, and annual and seasonal participation for selected recreation activities. In addition, data were obtained on socioeconomic characteristics of respondents. Some of the questions in the survey pertain to a list of 17 activities that were determined to represent a majority of urban recreation pursuits; however, perceptions of the attributes were obtained for only three activities, which were selected by each respondent as his or her most frequent recreation activities. The attributes include physical measures, such as distance traveled to the site, fee paid, and duration, and 23 conceptual items, which were rated on a five-point Likert scale that covers a range of agreement from strongly agree to strongly disagree.

One of the principal tasks of this research was to determine the feasibility of transferring the technology of individual choice modeling from travel demand to recreation demand by using the multinomial logit model (10-12). This technique can be expressed mathematically as

$$P(i; A_t) = \exp[V(Z_i, S_t)] \Big/ \sum_{j \in A_t} \exp[V(Z_j, S_t)] \qquad (1)$$

where

- $P(i; A_t)$ = the probability that recreation alternative i is chosen by consumer t from his or her choice set (A_t),
- $V(Z_i, S_t)$ = systematic (nonrandom) part of the utility,
- Z_i = vector of attributes of recreation alternative i, and
- S_t = vector of characteristics of individual t.

In this project, further support for segmentation is provided by the models built on the Evanston and Des Plaines data sets, which revealed substantial differences; however, these differences were also found in the distributions of various characteristics of respondents from the two locations. It seems reasonable to postulate that the observed differences may, therefore, be due to different distributions of tastes for recreation-activity attributes in the two suburbs. Also note that McFadden, Tye, and Train (13) have shown that treatment of a heterogeneous population as a homogeneous one results in case 2 violations of the independence of irrelevant alternatives property of multinomial logit models and leads to biased coefficient estimates and a pattern of overprediction and underprediction. Hence, population segmentation is necessary in order to reduce the likelihood of bias in the fitted models. (Of course, if no differences are found in the fitted coefficients of models from different segments, it may be postulated that the population is homogeneous and that case 2 violations from this cause are not present.)

HYPOTHESES OF SEGMENTATION

A number of hypotheses relating to population segmentation can be tested. First, a number of variables may be considered as bases for segmentation, including available socioeconomic characteristics (income, age, sex, and stage in the family life cycle) and situational or taste variables (geographic location, importance of recreation activities, and activity attractiveness, subjectively rated). In travel-forecasting work, results have been rather inconsistent with socioeconomic variables (7, 14-16). Nevertheless, it seems appropriate to test such variables because some can readily be hypothesized to have an effect on participation in recreation activities. The first hypothesis is, therefore, that socioeconomic and situational or taste variables can be used as a basis for population segmentation and will reveal significant differences in recreation-choice behavior. This hypothesis can be tested partially by analyzing variations in participation rates for different activities over the ranges of selected segmentation variables. Methods for this include simple graphical and cross-tabular presentations and analysis of variance.

The second hypothesis arises from the treatment of the ratings of the 23 conceptual attributes of recreation activities. These fundamental attributes should not be used in modeling because their individual reliabilities are very low, as has been established in psychometric theory (17); because they relate to a few underlying salient concepts that are formed by groups of the fundamental attributes; and because the evaluative space of an individual is believed to be quite limited in its number of dimensions, and these dimensions represent the salient concepts. The salient concepts can be identified by multidimensional scaling, individual scaling, and factor analysis. Previous work (18, 19) has shown factor analysis to be an acceptable procedure that is cheaper and less subject to limitations than the scaling procedures, and it was therefore used in this study (6). Three-factor solutions were used for all analytical work because these solutions appeared to meet all of the criteria set for selecting the most efficient space.

The second hypothesis, which arises from this, is that different population segments operate with different perceptual spaces and, hence, different factor structures. Although a statistical test for different factor structures has been suggested recently (20), this hypothesis was not tested in this research for three reasons: (a) Allaire (21) and Hauser (22) have shown that in consumer marketing it is reasonable to assume homogeneous perceptual spaces but with heterogeneous preference parameters; (b) some preliminary investigations of heterogeneity on two of the segmentation variables failed to reveal any apparent differences in the perceptual spaces for the data of this project; and (c) the adoption of an assumption of heterogeneous perceptual spaces would invalidate the use of the other statistical tests of comparison used in this research. Therefore, a homogeneous perceptual space was assumed for all segments.

It may be postulated that different segments will weigh various attributes differently in the recreation-participation model. This hypothesis may be tested by building models of the same specification for each selected population segment. Statistical tests, using Student's t-distribution, may be conducted on the coefficients of different segments by using Equation 2.

$$t = (a_k^m - a_k^n) / \sqrt{(\sigma_k^m)^2 + (\sigma_k^n)^2 - 2\mathrm{cov}(a_k^m, a_k^n)} \qquad (2)$$

where

- a_k^m, a_k^n = coefficients for attribute k from the m th and n th segments,

Table 1. Geographic segmentation models.

	Segments					
	Pooled Model (812 cases)		Des Plaines (395 cases)		Evanston (414 cases)	
Variable[a]	Coefficient	t-Value	Coefficient	t-Value	Coefficient	t-Value
ACHV	0.283	15.3	0.335	9.9	0.258	10.4
EXTR	0.102	5.3	0.231	7.1	0.017	1.0
PAST	0.061	3.5	0.278	9.0	-0.062	2.8
ATTR	0.249	17.9	0.338	15.1	0.190	10.2
AVAIL	-0.057	5.2	-0.136	7.9	0.018	1.1
FEINC	-0.088	1.8	-0.046	0.8	-0.307	3.1
DISTLDA	-0.0002	1.9	-0.0006	4.0	0.0004	2.3
CARLDA	0.041	1.8	0.051	1.4	0.038	1.1
GOLFAGE	0.008	4.1	0.015	5.0	-0.002	0.8
EDCULT	0.117	2.1	-0.148	1.6	0.197	2.4

[a] The alternative-specific constants have been excluded for space considerations.

σ_k^m, σ_k^n = standard errors of the coefficients, and

$\mathrm{cov}(a_k^m, a_k^n)$ = covariance of the coefficients a_k^m and a_k^n.

If the segments can be considered to be independent samples, the covariance term can be ignored (and, in practice, usually is).

In addition, likelihood-ratio tests can be performed between the pooled results of the segments and an unsegmented model. Minus twice the logarithm of the likelihood ratio ($-2 \log \lambda$) has been shown by Theil to be distributed like chi-square, with degrees of freedom equal to the difference between the sum of the number of fitted parameters of the segmented models over all segments and the numbers in the unsegmented model [i.e., $N_p(N_g-1)$, where N_p is the number of parameters and N_g is the number of segments or groups used].

The likelihood-ratio test, in this case, establishes whether or not the segmented models succeed in explaining more of the behavior than does the single unsegmented model. If the value of $-2 \log \lambda$ for the unsegmented model and the segmented models exceeds the table value of chi-square at a given significance level, then the null hypothesis (that segmentation provides no improvement in explanation of the phenomenon) can be rejected at that confidence level.

It may also be postulated that different segments of the population have different choice mechanisms, as would be shown if models with different specifications provide the best fit for different segments. This hypothesis is somewhat more difficult to test than was the preceding one. Rigorous statistical tests can be made only if the specification of the best model contains variables that represent a subset of those used under the preceding hypothesis or if the model from the preceding hypothesis is a subset of the best model. Otherwise, judgment would have to be on the basis of predictive performance and other similar properties.

In this research it was assumed that the perceptual spaces were common to all groups of the population and that all segments have the same choice mechanisms. Thus, it was necessary to find the best specification for a model to test for different weights on given attributes.

The search for the best model was done on the pooled data of the Evanston and Des Plaines suburbs. The steps in model development can be found elsewhere (6), and this model is reported later in this paper. The same model specification was used in all segmentation tests to facilitate the statistical testing of the hypotheses.

POPULATION SEGMENTATION

Seven segmentation variables were examined initially: income, age, sex, importance of recreation, stage in family life cycle, location, and attractiveness. Before the models were tested, however, cross-tabulations and one-way analysis-of-variance tests were made to detect interactions between the variables and activity-participation rates and to determine the levels at which to segment the variables. A constraint on the segmentation was imposed as a result of the relatively small size of the entire sample. A minimum sample of 100 cases was thought desirable, and a maximum of 812 cases was available from the entire data set. From this initial analysis, the most promising segmentation bases were found to be life-cycle stages, attractiveness, and geographic location.

In all of the models reported, a pooled three-factor structure was used in the best specification that was found for the unsegmented data. The dependent variable used in the model was summer participation (number of days on which the respondent had participated) for each of 10 reported activities—bowling; bicycling; swimming; playing tennis; playing golf; fishing; going to movies; going to theater, opera, or concerts; watching sports; and participating in team sports. The selection of these activities is reported elsewhere (6). The independent variables are listed and defined below.

ACHV—Achievement factor;
EXTR—Extroversion factor;
PAST—Pastoralism factor;
ATTR—Reported attractiveness of the activity;
AVAIL—Reported availability of the activity;
FEINC—Participation fee divided by annual gross income;
DISTLDA—Distance traveled to the activity for long- and medium-distance activities (i.e., swimming; playing golf; fishing; attending theater, opera, or concerts; and watching sports; 0 for other activities);
CARLDA—Number of automobiles available for long- and medium-distance activities, 0 otherwise;
GOLFAGE—Age for golf, 0 otherwise; and
EDCULT—Level of education for attending theater, opera, or concerts; 0 otherwise.

Geographic Segmentation

The models for geographic segmentation are shown in Table 1. The log-likelihood test between the geographic segments and the pooled model produces a value of 306 (this is the adjusted value for the difference in the numbers of observations for chi-square with 19 degrees of freedom. At 99 percent, the table value of chi-square is 36, so that the segmentation has clearly improved the model significantly. t-tests were also made for differences between individual coefficient values. The results of these tests are given below. It can be seen that all but two of the variables are significantly different at better than 95 percent confidence (t-value of the difference is less than 1.96).

Variable	t-Value	Variable	t-Value
ARCHV	2.49	FEINC	2.23
EXTR	5.06	DISTLDA	4.28
PAST	8.97	CARLDA	0.27
ATTR	5.11	GOLFAGE	0.32
AVAIL	6.59	EDCULT	2.80

In summary, geographic segmentation shows significant differences and improves the performance of the models. Variations are found both in the weights given to different recreation factors and to the weights of situational variables for the two segments.

Attractiveness Segmentation

As noted, the attractiveness segmentation appeared likely to be reasonably useful and represents the best approximation to a personality segmentation that can be achieved from these data.

Segmentation on attractiveness was undertaken through further analysis of the attractiveness scores on each activity. First, activities were grouped in terms of attractiveness. This is necessary for a number of reasons. Pragmatically, to use the 10 separate activities would generate a minimum of 100 (10^2) segments, where these would be defined according to a low or high attractiveness rating on each activity. Clearly, the data set is inadequate in size to support such a segmentation. All of those segments would probably not be populated because several activities would have sufficient segments in common that similar ratings would be given to activities that fall in particular groups. Also, the reliability of attractiveness scores for individual activities will probably be relatively low and would be improved by grouping similar activities and by using an aggregate rating for each group and individual. If activity groups are used, people would then be grouped according to the attractiveness scores that they gave to the different activity groupings. Activity groupings were obtained by subjecting the raw attractiveness scores to cluster analysis. These clusters are shown in Table 2. The first two clusters are considered to be reasonably consistent internally and also represent intuitively plausible clusters: The first cluster is social-cultural activities, and the second is outdoor (pastoral) sports. The third cluster is less consistent and less easily identified and was not used for segmentation.

The next step in the process was to find values of each attractiveness cluster that could be used for segmentation purposes. To do this, attractiveness scores were summed for each individual for the activities in each of the two clusters and then plotted on a scatter diagram, from which the data were divided into four approximately equal-sized groups (quadrants) for population segmentation, as shown in Table 3.

By using the same model specification as for the geographic segments, models were built for each of the four attractiveness segments, as shown in Table 4. The same likelihood-ratio test was carried out to determine if the segmented models together were able to explain more of the choice variation than was the pooled model. The adjusted value of -2 log λ for the test was found to be 492, which is substantially larger than the 99.5 percent table value of chi-square (of 88) for 57 degrees of freedom. Hence, the attractiveness segmentation can again be said to offer a significant improvement in the model performance.

Table 5 shows the results of t-tests for similarity of coefficients. It can be seen that 4-7 of the 10 vari-

Table 2. Attractiveness activity clusters for segmentation of pooled data.

Cluster	Activity	Internal Consistency (α)
Social-cultural (SOCATT)	Visit museum or art gallery; attend theater, opera, or concert; visit zoo; go to movies; picnic; and dance	0.70
Outdoor-sports (SPATT)	Bicycle, swim, play tennis, jog, and sail	0.65
Recreational activities	Watch sports, play team sports, bowl, fish, golf, and motorboat	0.58

Table 3. Attractiveness segments for pooled data.

	Attractiveness Score	
Segment	Social-Cultural	Outdoor Sports
1	Low ≤ 19	Low ≤ 13
2	Low ≤ 19	High > 13
3	High > 19	Low ≤ 16
4	High > 19	High > 16

Table 4. Attractiveness segmentation models.

			Segment Models							
	Pooled Model (812 cases)		Low Social and Cultural, Low Pastoral Sports (154 cases)		Low Social, High Pastoral Sports (222 cases)		High Social, Low Pastoral Sports (225 cases)		High Social and Cultural, High Pastoral Sports (187 cases)	
Variable[a]	Coefficient	t-Value	Coefficient	t-Value	Coefficient	t-Value	Coefficient	t-Value	Coefficient	t-Value
ACHV	0.283	15.3	0.407	5.4	0.470	12.7	0.157	4.9	0.193	5.1
EXTR	0.102	5.3	0.249	4.1	0.010	0.3	0.129	3.6	0.133	3.5
PAST	0.061	3.5	-0.169	2.8	-0.0005	0.01	0.093	2.8	0.188	5.0
ATTR	0.249	17.9	0.372	9.2	0.373	12.5	0.196	8.3	0.280	6.6
AVAIL	-0.057	5.2	-0.133	4.0	-0.014	0.7	-0.035	1.7	-0.106	4.4
FEINC	-0.088	1.8	0.699	5.8	-0.062	0.8	-0.489	4.2	-0.131	1.0
DISTLDA	-0.0002	1.9	-0.0002	0.8	-0.0004	1.5	-0.023	1.3	-0.0005	1.8
CARLDA	0.041	1.8	-0.254	3.5	0.071	1.6	0.013	0.3	-0.085	1.6
GOLFAGE	0.008	4.1	0.023	4.3	0.014	4.3	0.006	1.1	-0.011	1.7
EDCULT	0.117	2.1	0.284	1.2	0.189	1.4	-0.176	2.3	0.286	2.0

[a]Alternative-specific constants have been excluded, for space considerations.

Table 5. T-tests of the differences in coefficient estimates for the attractiveness segments.

	T-Values for Coefficient Differences									
Segment	ACHV	EXTR	PAST	ATTR	AVAIL	FEINC	DISTLDA	CARLDA	GOLFAGE	EDCULT
1 and 2	0.75	3.21	2.44	0.03	2.98	5.65	0.51	3.84	1.40	0.35
1 and 3	3.08	1.68	3.82	3.76	2.46	7.13	0.11	3.21	2.29	1.90
1 and 4	2.56	1.62	5.06	1.57	0.66	4.56	0.81	1.89	4.07	0.01
2 and 3	6.38	2.15	1.93	4.66	0.73	2.68	0.48	0.95	1.35	2.28
2 and 4	5.24	2.17	3.70	1.80	2.83	0.29	0.31	2.23	3.47	0.48
3 and 4	0.73	0.06	1.89	1.72	2.19	1.99	0.81	1.44	2.00	2.78

ables produce coefficients that are significantly different between segments. The least distinction is found between groups 3 and 4, and the greatest differences are between groups 1 and 3 and groups 2 and 4. This suggests that the attractiveness of social-cultural activities gives the strongest segmentation, and the attractiveness of outdoor sports gives a rather poor segmentation.

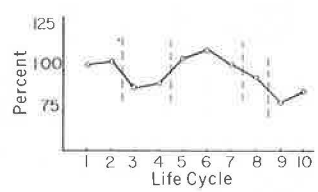
In conclusion, it may be stated that segmentation by attractiveness ratings has produced significantly different models, wherein most of the differences have intuitively meaningful interpretations.

Stage in the Family Life Cycle

The final segmentation variable used is a compound socioeconomic variable, which is given in the table below. It was felt that this compound variable would be a more useful segmentation variable than any of the simple socioeconomic variables considered in the preliminary work. For our purposes, married is interpreted as implying a household of two adults who live together. The compound variable has been found to be useful for travel-demand segmentation (14), as well as in other social science areas (23-25).

Stage	Definition
1	Young (≤35 years), unmarried, living alone
2	Young, unmarried, living with others
3	Young, married, no children
4	Married, oldest child <5 years
5	Married, oldest child between 5 and 12 years
6	Married, oldest child between 12 and 17 years
7	Married, oldest child over 17 years
8	Older (>35 years), married, no children at home
9	Older, unmarried, living alone
10	Older, unmarried, living with others

Figure 1. Percentage of mean participation versus life cycle.



One may suggest, a priori, how the life-cycle variable will affect recreation behavior. For example, people in stages 1 and 2 are likely to be more active because of a lack of various responsibilities and independence from other people; whereas in stages 3 and 4, which constitute a home-making stage, they would tend to be less active because of the existence of preschool children or because of the need for extra money, which leads to extra working hours and sacrifices from leisure time. A number of similar arguments can be advanced to suggest other groupings among life-cycle stages.

Because of some apparent similarities among some cycles and for pragmatic reasons, all 10 stages were not retained for population segmentation. It was therefore decided to group various stages to form segments. Initially, a graph was produced to show average activity-participation rates for each life-cycle stage. This is shown in Figure 1 and suggests that a reasonable grouping of stages would be (1, 2), (3, 4), (5, 6, 7), (8), and (9, 10). These are numbered as segment numbers 1, 2, 3, 4, and 5, respectively. Separate models were estimated for each segment by using these groupings and the same model specification as for the two previous segmentation procedures. The results of the segmented modeling are shown in Table 6.

The first test made on the segmented models is the likelihood-ratio test, which, after adjustment, produces a value of $-2 \log \lambda$ of 902 with 76 degrees of freedom. The table value of chi-square at 99.9 percent is approximately 112, from which one may again conclude that the segmented models perform significantly better than the unsegmented model. The results of t-tests of the coefficient differences among the five segments are shown in Table 7. All segments exhibit some significant differences from any other segment; the maximum number (7) was between segments 1 and 2 (stages 1 and 2 and stages 3 and 4), and the minimum (3) was between segments 4 and 5 (stage 8, and stages 9 and 10). Note, however, that segments 4 and 5 each contain very small samples, which has the effect of reducing significantly the reliability of the coefficients, so that only four have significant coefficients in both segments. This may,

Table 6. Models of life-cycle segments for Des Plaines and Evanston pooled data.

	Life Cycle 1 (125 cases)		Life Cycle 2 (189 cases)		Life Cycle 3 (324 cases)		Life Cycle 4 (57 cases)		Life Cycle 5 (90 cases)		Pooled (812 cases)	
Variable[a]	Coefficient	t-Value	Coefficient	t-Value	Coefficient	t-Value	Coefficient	t-Value	Coefficient	t-Value	Coefficient	t-Value
ACHV	0.315	6.0	0.402	9.8	0.229	7.5	0.593	5.7	0.398	5.1	0.283	15.3
EXTR	0.142	2.5	-0.123	2.9	0.192	6.5	0.298	2.0	0.361	4.7	0.102	5.3
PAST	0.073	1.6	0.002	0.1	0.144	5.4	0.845	6.1	0.167	2.2	0.061	3.5
ATTR	0.329	7.0	0.171	5.5	0.399	19.7	0.568	5.5	-0.230	4.3	0.249	17.9
AVAIL	0.035	1.1	-0.104	4.5	-0.084	5.0	0.162	2.0	0.011	0.3	-0.057	5.2
FEINC	-0.512	3.0	0.042	0.3	-0.373	4.2	-0.228	1.0	-0.646	2.8	-0.088	1.8
DISTLDA	-0.0009	1.9	0.001	3.6	-0.0003	2.4	0.0015	1.9	-0.0003	1.2	-0.0002	1.9
CARLDA	-0.243	3.8	0.103	1.5	0.026	0.8	-0.046	0.4	0.300	2.2	0.041	1.8
GOLFAGE	-0.023	1.3	0.033	1.8	0.014	3.8	0.067	4.2	0.0007	0.1	0.008	4.1
EDCULT	-0.138	1.1	0.108	0.6	-0.430	4.4	0.769	4.1	0.427	2.9	0.117	2.1

[a] Alternative-specific constants have been excluded, for space considerations.

Table 7. T-tests of the differences in coefficient estimates for the life-cycle segments.

	T-Values for Coefficient Differences									
Segment	ACHV	EXTR	PAST	ATTR	AVAIL	FEINC	DISTLDA	CARLDA	GOLFAGE	EDCULT
1 and 2	1.30	3.78	1.15	2.00	3.60	2.51	3.60	3.70	2.19	1.07
1 and 3	1.41	0.78	1.33	0.10	3.86	0.72	1.18	3.73	2.02	3.65
1 and 4	1.35	0.99	36.48	0.42	1.49	1.02	2.44	1.36	3.73	4.08
1 and 5	0.88	2.30	1.07	7.81	0.47	0.46	1.03	3.55	1.20	2.96
2 and 3	3.40	6.16	2.89	6.16	1.15	2.51	4.74	0.99	1.01	1.48
2 and 4	1.72	2.73	5.85	3.69	3.23	1.04	0.23	1.01	1.42	2.46
2 and 5	0.04	5.51	1.92	6.44	2.55	2.52	3.54	1.27	1.63	1.31
3 and 4	3.39	0.71	4.98	1.62	3.05	0.61	2.11	0.54	0.44	1.62
3 and 5	2.02	2.06	0.29	10.92	2.25	1.09	0.00	1.97	1.51	0.02
4 and 5	1.51	0.38	4.32	6.80	0.61	1.30	1.99	1.82	3.70	1.44

therefore, be the major cause of a lack of significant differences between coefficients.

The full interpretation of the differences among coefficient estimates is not given here because of space considerations; they can be found elsewhere (6). In summary, however, segmentation by life-cycle stages has revealed a significant number of plausible differences in the weights attached to the variables in the recreation-participation models. With the exception of two of the alternative-specific situational variables, all significant differences in weights point to expected differences in tastes and constraints. It seems appropriate to conclude, therefore, that this segmentation scheme is a worthwhile scheme that has identified a number of underlying differences in behavior, although the life-cycle variable may be operating as a proxy for a complex set of constraints and for personality maturation.

CONCLUSIONS

The results reported in this paper indicate that there exist significant variations in tastes and behavior that can and should be captured through population segmentation. Each of the tested segmentation schemes has revealed significant differences, and most of these differences bear plausible relationships to the segmentation variables. The use of a single, unsegmented model offers advantages of simplicity but will result in significant inaccuracies in the representation of recreation behavior, and may result in misdirected policies with respect to urban recreation facilities.

Separate prediction tests were not carried out, as this would have required reserving at least one-half of the already small sample for such tests. However, our experience is that the significant differences found in the models generally lead to poorer predictions if the unsegmented models are used in predictions.

It must also be noted that this research makes no claim to have identified optimum segmentation schemes. No attempts have been made to examine alternative groupings within segmentation schemes, to examine multiple segmentation (i.e., segmentation on more than one variable), or to seek optimal model specifications within segments. Until such efforts are made, we can only conclude that segmentation will improve model accuracy and that the segmentation schemes reported here will at least provide some gains in both policy insights and model accuracy.

ACKNOWLEDGMENT

The work reported here was done under grant from the Division of Research Applied to National Needs of the National Science Foundation.

REFERENCES

1. L. Gapay. National Parks Battle to Deal with Crowds but Seem to Be Losing. Wall Street Journal, Sept. 15, 1976, pp. 1, 18.
2. P. Stoler. Bumper to Bumper in the Wilderness. Time, July 4, 1977.
3. C. J. Cichetti, A. C. Fisher, and V. K. Smith. Economic Models and Planning Outdoor Recreation. Operations Research, Vol. 21, Sept. 1973, pp. 1104-1113.
4. M. Clawson and J. L. Knetsch. Economics of Outdoor Recreation. Johns Hopkins Press, Baltimore, 1966.
5. V. R. Desimone and A. Zierenberg. Transportation Strategies for Parks and Recreation. Journal of the Urban Planning and Development Division, Proc., ASCE, May 1978, pp. 7-19.
6. A. Anas and others. Urban Recreation. Division of Research Applied to National Needs, National Science Foundation, Final Draft Rept., Grant APR 76-19086, 1978.
7. P. L. Watson and P. R. Stopher. The Effect of Income on the Usage and Valuation of Transport Modes. Transportation Research Forum Proceedings, Vol. 15, No. 1, 1974, pp. 460-469.
8. J. J. Louviere, L. M. Ostresh, Jr., D. H. Henley, and R. J. Meyer. Travel-Demand Segmentation: Some Theoretical Considerations Related to Behavioral Modelling. In Behavioral Travel-Demand Models (P. R. Stopher and A. H. Meyburg, eds.), Lexington Books, D. C. Heath and Company, Lexington, MA, 1976, Chapter 14.
9. D. A. Hensher. Use and Application of Market Segmentation. In Behavioral Travel-Demand Models (P. R. Stopher and A. H. Meyburg, eds.), Lexington Books, D. C. Heath and Company, Lexington, MA, 1976, Chapter 15.
10. D. McFadden. Quantitative Methods for Analyzing Travel Behavior of Individuals: Some Recent Developments. In Behavioural Travel Modelling (D. A. Hensher and P. R. Stopher, eds.), Croom Helm Ltd., London, 1978, Chapter 13.
11. T. Domencich and D. McFadden. Urban Travel Demand: A Behavioral Analysis. North-Holland, Amsterdam, 1975.
12. P. R. Stopher and A. H. Meyburg. Survey Sampling and Multivariate Analysis for Social Scientists and Engineers. Lexington Books, D. C. Heath and Company, Lexington, MA, 1978.
13. D. McFadden, W. B. Tye, and K. Train. An Application of Diagnostic Tests for the Independence from Irrelevant Alternatives Property of the Multinomial Logit Model. TRB, Transportation Research Record 637, 1977, pp. 39-46.
14. E. Aldana, R. DeNeufville, and J. H. Stafford. Microanalysis of Urban Transportation Demand. HRB, Highway Research Record 446, 1973, pp. 1-11.
15. P. R. Stopher and A. H. Meyburg. The Effect of Social and Economic Variables on Choice of Travel Mode for the Work Trip. In Transportation and Traffic Theory (D. J. Buckley, ed.), A. H. and A. W. Reed, Sydney, 1974, pp. 685-713.
16. L. L. Lucero. Towards Improved Demographic Groupings in Disaggregate Behavioral Modelling of Travel Demand. Department of Civil Engineering, Northwestern Univ., M.S. Thesis, Aug. 1978.
17. J. C. Nunnally. Psychometric Theory. McGraw-Hill, New York, 1967.
18. F. S. Koppelman and J. Hauser. Effective Marketing Research: An Empirical Comparison of Techniques to Model Consumers Perceptions and Preferences. Paper presented at TRB meeting, Washington, DC, 1978.
19. J. N. Prashker. Development of a "Reliability of Travel Modes" Variable for Mode-Choice Models. Northwestern Univ., Evanston, IL, Ph.D. dissertation, 1977.
20. J.-M. Choffray and G. L. Lilien. Industrial Adoption of Solar Air Conditioning: Measurement Problems, Solutions and Marketing Implications. Alfred P. Sloan School of Management, Massachusetts Institute of Technology, Cambridge, Working Paper WP 894-76, 1976.
21. Y. Allaire. The Measurement of Heterogeneous Semantic Perceptual and Preference Structures. Massachusetts Institute of Technology, Cambridge, Ph.D. dissertation, 1973.

22. J. R. Hauser. A Normative Methodology for Predicting Consumer Response to Design Decisions: Issues, Models, Theory and Use. Massachusetts Institute of Technology, Cambridge, Ph.D. dissertation, 1975.
23. D. Wells and G. Gubar. Life Cycle Concept in Marketing Research. Journal of Marketing Research, Vol. 3, 1966, pp. 355-363.
24. G. Katona, E. L. Mueller, J. W. Schmiedeskamp, and J. A. Sonquist. Survey of Consumer Finances, 1966. Institute for Social Research, Univ. of Michigan, Ann Arbor, 1967.
25. J. B. Lansing and J. N. Morgan. Consumer Finances over the Life-Cycle. In Consumer Behavior (L. H. Clark, ed.), New York Univ. Press, New York, Vol. 2, 1955.

Publication of this paper sponsored by Committee on Traveler Behavior and Values.

Sampling Vehicle Kilometers of Travel

Herbert S. Levinson, Wilbur Smith and Associates, New Haven, Connecticut
A. L. Roark, Commissioner of Environmental Protection, Frankfort, Kentucky
J. S. Guhin, Federal Highway Administration

This paper develops sampling procedures for estimating vehicle kilometers of travel on urban streets. It shows how simple and stratified random sampling techniques can be applied to estimate sample-size requirements for estimating freeway, arterial-collector, and local-street vehicle kilometers of travel. The paper also presents and provides ranges in the parameters associated with the variations in traffic volumes in space and time. These estimates are then used as part of a practical, operational procedure.

Reliable estimates of urban vehicle kilometers of travel are important for many transportation planning and policy purposes. They help assess the effectiveness of safety programs. They provide a basis for allocating highway-user revenues and establishing highway financing programs. They help validate urban transportation planning models and monitor urban travel growths. They provide a means to assess the effectiveness of transportation system management, air quality, and energy conservation programs.

More than 40 years of research on traffic volume characteristics and variations (1-3) has shown that:

1. Urban traffic follows daily and hourly variation patterns that are generally consistent and often predictable. Urban traffic patterns exhibit relatively little weekday and seasonal variation. The percentage of total traffic in any given period is approximately the same along any route.
2. The more counts at a given location, the greater is the reliability. Similarly, the heavier the traffic volumes at a particular location, the greater is the reliability of the estimated volume.
3. The distribution of counts throughout the day is more significant than the total time during which the traffic is counted. Therefore, the number of separate and independent observations is more important than the duration of each observation.
4. Five- to six-minute short-counts are entirely satisfactory where traffic is not light or unduly erratic.

BASIC CONCEPTS AND VARIABLES

The most reliable method for developing traffic volume and vehicle kilometer information is to count each section of roadway for each day throughout the entire year. Such a procedure is neither practical nor possible. Consequently, it is necessary to apply sampling procedures.

Sampling urban vehicle kilometers of travel involves (a) identification of the basic variables and how they relate, (b) quantification of them, (c) statistical application of them, and (d) development of simplified procedures for practical use. This last step involves applying observed ranges in parameters to various sampling formulas to simplify computational steps.

Traffic volumes on the urban street system vary by time and space. Where estimates of vehicle kilometers are involved, the length of roadway section becomes a third variable. A link is defined as a section of roadway that has a uniform traffic volume. Sampling of vehicle kilometers of travel thus involves the following three basic sources of variation or error:

1. The variation in traffic volumes from one link to another (this is defined as the spatial variation among the population of traffic counts),
2. The variation in volumes on any given link resulting from day-to-day changes in traffic flow (this is defined as the temporal variations in traffic counts), and
3. The variations in the lengths of links.

These variations exist for volumes along any urban road system. The three types of variations are essentially independent of each other with zero correlation. This results in the following formula for the first two sources of variations.

$$S_x^2 = S_1^2 + S_2^2 + S_{1,2}^2 \qquad (1)$$

Since we assume that $S_{1,2}^2 = 0$

$$S_x^2 = S_1^2 + S_2^2 \qquad (2)$$

where

S_1^2 = spatial variance,
S_2^2 = temporal variance,
$S_{1,2}^2$ = covariance of S_1 and S_2, and
S_x^2 = composite variance in the population of traffic volume counts at a given point in time.

Estimation of the vehicle kilometers of travel per link is somewhat analogous to estimation of the area of a rectangle with errors in both the length and width. The variance in the vehicle kilometers of travel per link reflects variations in (a) the volume per link and (b) length of links. This variance can be approximated by the following formula:

$$S^2 \simeq S_x^2 \bar{\ell}^2 + S_L^2 \bar{X}^2 \tag{3}$$

where

$\bar{\ell}$ = mean length of link (km),
S_L^2 = variance in mean length of link,
$\bar{X}$ = mean volume per link (vehicles),
S^2 = variance (vehicle-km/link), and
S_x^2 = variance in population of counts (volumes).

Figure 1 shows how variations in volumes per link and link length affect this formula.

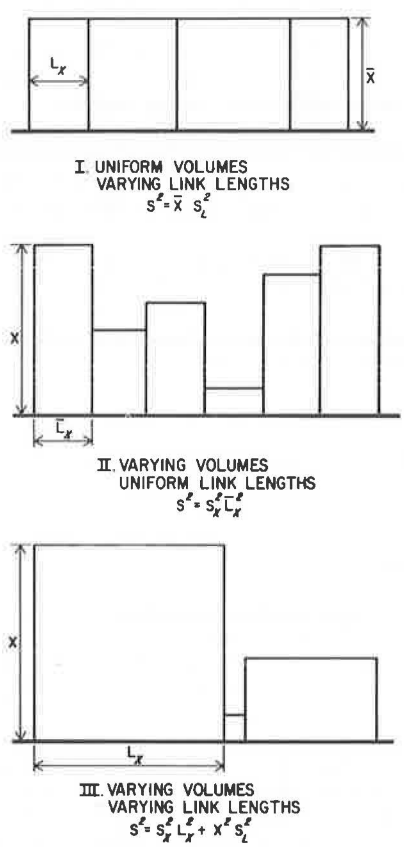
Case 1—when links vary widely in length but volumes are uniform, a sampling error is introduced in the estimate of vehicle kilometers of travel. The variance (S^2) equals the mean volume per link $(\bar{X})^2$ times the variance in link length (S_L^2).

Case 2—when links are of uniform length, the variance in vehicle kilometers of travel is proportionate to the square of the mean length of link (that is, $S^2 = S_x^2 \bar{L}^2$). This is a restatement of a long-established statistical relationship.

Case 3—the variance in vehicle kilometers of travel per link increases when there is a variation in both link lengths and vehicle kilometers of travel.

Each of these sources of variations should be reduced in practice. Links can be stratified by volume group to reduce the variance among counts (volumes). This should be done where the engineer or planner is familiar with the road system. Links of nearly equal length can be used, although it is not practical to have links of exactly equal length. Slight variations in link length can produce major differences in the variance of the vehicle kilometers of travel per link.

Figure 1. Theoretical formulation of variance in vehicle kilometers per link.



ESTIMATING PARAMETERS

In applying various sampling formula, it is necessary to know how traffic volumes vary in space and time. Accordingly, estimates were made of these variations based on a literature review and analysis of urban traffic volume patterns.

Temporal coefficients of variation of one-day traffic counts at continuous urban counting locations are shown in Table 1. Similar data were obtained from analysis of two-, three-, four-, and five-day counts. These analyses produced the following coefficients of variation for weekday traffic volumes.

Average Weekday Traffic	Coefficient of Variation 1-Day Count (%)	5-Day Count (%)
1 000	30	25
5 000	16	12
10 000	13	9
20 000	10	7
50 000	7.5	5
80 000	6	4

Typical spatial variations of urban traffic counts, based on minimum stratification of road types, are shown in Table 2. This current experience suggests the following ranges in the coefficient of variation: arterial streets—80-120 percent; freeways—50-80 percent.

Grouping of roadway sections into relatively uniform volume strata substantially reduces the coefficients of variation. The results of such groupings for traffic volumes in six urban areas are given in Table 3. These data suggest that the distribution of counts within any class or strata can be approximated by a uniform distribution. Thus, the spatial standard deviation approximates 30 percent of the range. To illustrate, for a 5000-volume range, the spatial standard deviation approximates 1500 vehicles.

Composite measures of spatial and temporal variation for freeways, arterials, and local streets are given in Tables 4 and 5. These values are based on a 2000-vehicle range for local streets, 5000- and 10 000-vehicle ranges for arterial streets, and broader ranges for urban

Table 1. Temporal coefficients of variation.

Location	Weekday Traffic[a]	Coefficient of Variation (V)[b]
Connecticut		
CT-124, New Cannan	8 420	0.126
Charter Oak Bridge, Hartford	16 975	0.106
Bissell Bridge, S. Windsor	9 640	0.107
Putnam Bridge, Wethersfield	13 285	0.097
Chicago		
Dan Ryan Expressway, Congress	98 470	0.051
Dan Ryan Expressway, Garfield	113 970	0.060
Dan Ryan Expressway, 95th (west spur)	31 175	0.086
Calumet Expressway, south of 95th	33 945	0.108
Stevenson Expressway, Pulaski	46 740	0.065
Kennedy Expressway, west of Edens	56 045	0.051
Lake Shore Drive, Foster	42 015	0.060
Dade County, Florida		
Northwest 27th Avenue	21 125	0.066
South Dixie Highway	30 155	0.054
Massachusetts		
Northeast Expressway, Revere	26 905	0.062
Southeast Expressway, Boston	59 965	0.076

Note: Values are based on weekday counts for the entire year except for Illinois, where values are based on counts for one week each month.
[a]One-direction.
[b]$V = \sigma/x$.

Table 2. Reported spatial variations in urban traffic volumes.

Place	Class	Coefficient of Variation (%)
Great Britain[a]	Motorway	56
	Trunk road	81
	Class 1 road	106
	Class 2 road	92
	Class 3 road	158
	Unclassified	149
Australia[b]	Class 6—urban arterial street	103
	Class 7—urban subarterial street	117
	Class 8—urban residential street	163
	Class 9—urban special purpose street	139
United States[c]		
Denver	6-lane freeway (N = 50)	46
	4-lane freeway (N = 11)	50
Tulsa	4-lane freeway (N = 27)	48
Winston-Salem	4-lane freeway (N = 34)	95
	All facilities (N = 622)[d]	108
Bergen County, NJ	All facilities (N = 984)[d]	160
Washington, DC	All non-central business district (CBD) facilities (N = 2082)[d]	99
	All CBD facilities[d]	69
Nassau County, NY	All facilities (N = 829)[d]	81
Syracuse, NY	All facilities (N = 748)[d]	74
Tulsa	Arterial street only (N = 665)[d]	82
	All facilities (N = 700)[d]	87
Hartford	All facilities (N = 1576)[d]	115

[a] Classification by British roadway class (urban facilities) (4).
[b] National Association of Australian State Road Authorities classification (5).
[c] (6).
[d] Excludes local streets.

Table 3. Summary of spatial parameters—urban traffic volumes by strata.

Traffic Volume	Tulsa[a] SD	Tulsa[a] V	Hartford SD	Hartford V	Washington, DC CBD SD	CBD V	Non-CBD SD	Non-CBD V	Bergen County SD	Bergen County V	Nassau County SD	Nassau County V	Syracuse SD	Syracuse V
0-1 000	230	0.46	240	0.33	60	0.06	90	0.09	260	0.37	120	0.15	230	0.31
1 000-5 000	1080	0.40	1 140	0.37	1110	0.37	1 020	0.33	940	0.27	1 130	0.33	1 030	0.29
5 000-10 000	1460	0.20	1 400	0.19	1540	0.20	1 450	0.20	1 450	0.19	1 360	0.18	1 560	0.21
10 000-15 000	1390	0.11	1 420	0.12	1320	0.10	1 490	0.12	1 500	0.12	1 340	0.11	1 440	0.12
15 000-20 000	1440	0.08	1 390	0.08	1640	0.09	1 520	0.09	1 390	0.08	1 470	0.09	1 480	0.09
20 000+	3250	0.13	19 300	0.46	5960	0.20	15 500	0.48	52 900	1.17	15 500	0.59	11 700	0.39
All	8950	0.82[b]	11 730	1.15	9860	0.69	13 600	0.99	29 700	1.60	11 500	0.93	8 900	0.74

[a] Arterials only.
[b] All facilities: 0.87.

Table 4. Suggested composite measures of variation by length of count cluster and volume—5000-vehicle strata, arterial streets.

Range	Mean[a]	Standard Deviation 1 Day	SD 2 Days	SD 3 Days	SD 5 Days	Coefficient of Variation (%) 1 Day	CV 2 Days	CV 3 Days	CV 5 Days
Local streets									
0-2000	1 000	670	660	660	650	67	66	66	65
Arterial streets									
0-5 000	2 500	1 580	1 570	1 560	1 550	63	63	62	62
5 000-10 000	7 500	1 830	1 790	1 750	1 710	24	24	23	23
10 000-15 000	12 500	2 120	2 020	1 940	1 850	17	16	16	15
15 000-20 000	17 500	2 300	2 250	2 130	1 980	13	13	12	11
20 000-25 000	22 500	2 700	2 450	2 310	2 110	12	11	10	9
25 000-30 000	27 500	2 890	2 660	2 480	2 270	11	10	9	8
30 000-35 000	32 500	2 930	2 890	2 670	2 410	9	9	8	7
35 000-40 000	37 500	3 350	3 090	2 830	2 500	9	8	8	7
Freeways									
4 lanes 20 000-60 000	40 000	12 410	12 290	12 250	12 170	31	31	30	30
6 lanes 40 000-120 000	80 000	24 470	24 410	24 290	24 210	31	31	30	30
8 lanes 60 000-200 000	130 000	42 540	42 430	42 310	42 190	33	33	32	32

Note: Data are rounded.
[a] Mean volume is assumed to be the midpoint of the range.

Table 5. Suggested composite measures of variation—10 000-vehicle strata, 1-day counts, arterial streets.

Volume Range	Mean[a]	Spatial SD	Temporal SD	Composite SD	Composite V (%)
0-10 000	5 000	3000	800	3100	62
10 000-20 000	15 000	3000	1650	3420	23
20 000-30 000	25 000	3000	2380	3830	15
30 000-40 000	35 000	3000	2980	4230	12

[a] Mean volume is assumed to be the midpoint of the range.

freeways. They provide a guide in establishing a first-year sampling plan, and they also allow for further simplification of estimating procedures.

ESTIMATING VEHICLE KILOMETERS OF TRAVEL

The general approach to estimating urban vehicle kilometers of travel by major functional class of roadway is shown in Figure 2. It calls for grouping the urban street system into three basic types of roadways—freeways, arterial-collector streets, and local streets. Both simple and stratified sampling can be used, depending on the extent of volume information available. Where traffic volume information exists within each road class, further stratification can be used to reduce the variance. This stratification can also be based on the number of travel lanes.

Sampling Approach

Sampling can be based on the population of road segments of 1-km length. This approach obviates the need to estimate the distribution of link length and represents a reasonable approximation.

Estimates can be made of the variation patterns in the distribution of vehicle kilometers per kilometer. By using these estimates of variation, a sample size of n kilometers can be obtained for any road type for specified precision and confidence levels. Since links are on the average $\bar{l}$ kilometers long (in any given volume, group, or strata), $n/\bar{l}$ links should be counted. The more uniform the link lengths, the greater the accuracy of this method.

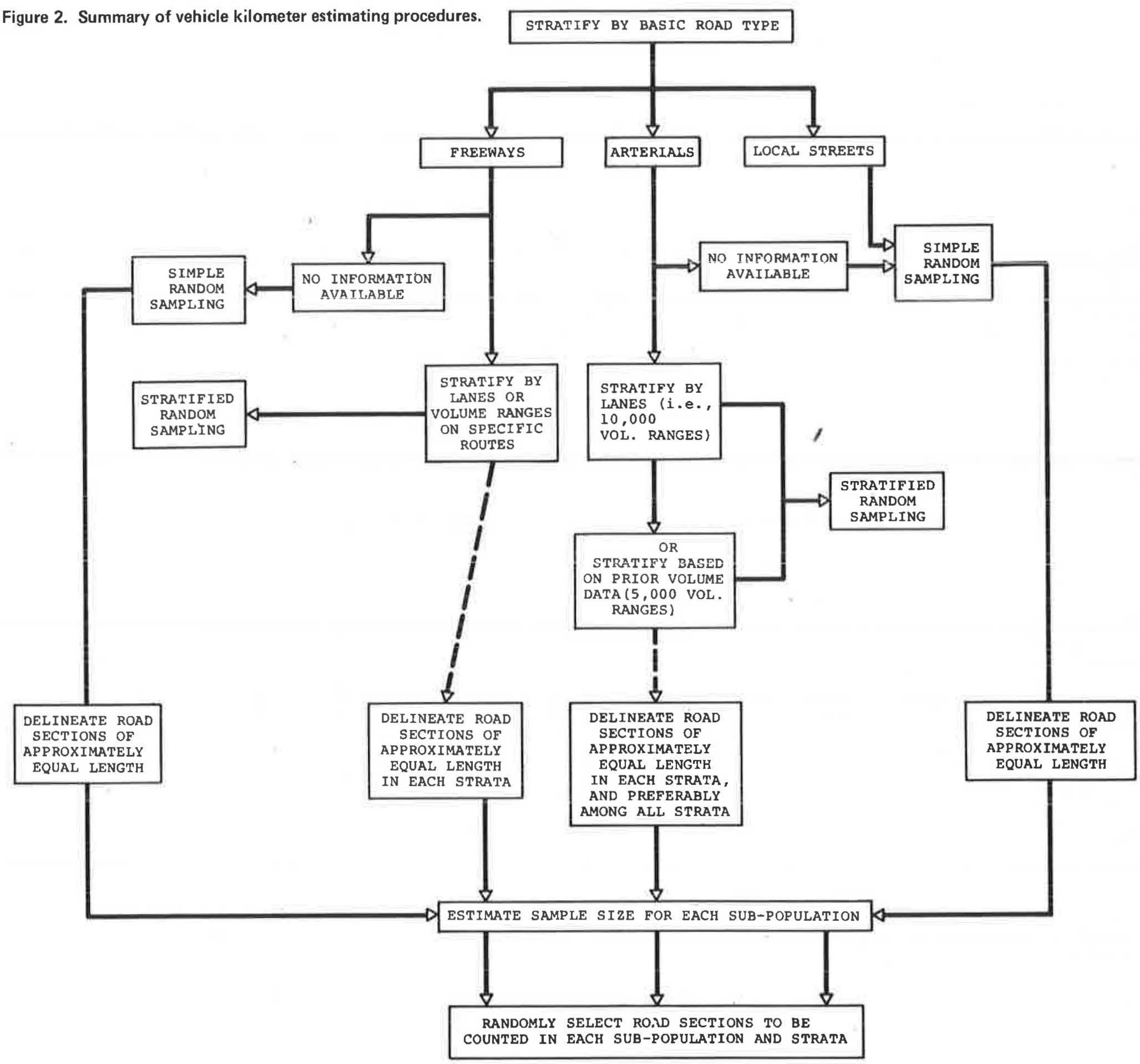
This approach assumes that the variability in vehicle kilometers is proportional to the variation in counts. The variance in the mean vehicle kilometers per kilometer of roadway can be assumed equal to variance in traffic counts.

$$S^2(\text{vehicle-km of travel/km}) = S^2(\text{average volume/count}) \qquad (4)$$

This relation appears valid, even where link sizes vary.

The sample size (n) required to estimate the mean

Figure 2. Summary of vehicle kilometer estimating procedures.



vehicle kilometers per kilometer can be assumed to represent the number of kilometers of roadway that should be measured. [Thus, if there are 100 km of road in a given strata and $n = 15$, then 15 km of road (15 percent of the road system) should be measured.] This implies that the same number of kilometers of road section would be measured, regardless of the size or distribution of links. The number of links counted would then increase inversely with link length.

The random selection of links equivalent to the desired kilometers of road is a conservative approach as compared with sampling all links within each kilometer selected. In application, the maximum link length should not exceed 1 km.

The sampling plan implies that n kilometers of road in each strata (or road class) would be measured and that all links found in these kilometers would be counted.

A reasonable approximation is to estimate the number of links in each volume stratum (h) from the relationship

$$n_h/\ell_h \tag{5}$$

where n_h = kilometers of road in stratum h, and ℓ_h = average link length (km) in stratum h ($0 < \ell < 1$).

Each roadway link should be described uniquely and should represent a section of relatively homogeneous volume. Where there are great variations in link lengths, dummy sections may be appropriate to try and equalize link lengths. Links can be selected randomly within each type of road or strata.

Each section of roadway should be counted only once. The durations of count may range from a one-day (weekday) count to a five-day Monday-Friday count. This is based on the finding that spatial variations are greater than temporal variations.

Sampling Formula

The following formulas are based on probability sampling theory with appropriate adjustments for the finite aspects of the population sampled (7).

Simple Random Sampling

$$n = [Z^2(S_1^2 + S_2^2 K)]/[E^2 + (S_1^2 Z^2/N_1)] \tag{6}$$

$$n \simeq [Z^2(S_1^2 + S_2^2)]/[E^2 + (S_1^2/N_1)Z^2]$$
$$\cong [Z^2(C_1^2 + C_2^2)]/[e^2 + (C_1^2/N_1)Z^2] \tag{7}$$

Ignoring the finite population correction factor,

$$n = [Z^2(C_1^2 + C_2^2)]/e^2 \tag{8}$$

where

- S_1^2 = spatial variance,
- S_2^2 = temporal variance,
- Z = intercept on normal distribution,
- K = finite population correction factor for temporal variance $\simeq 1$,
- N_1 = number of kilometers in road class (if links average 1 km in length, then N_1 = number of links in road class),
- E = absolute error in mean vehicle kilometers per kilometer,
- C_1 = spatial coefficient of variation,
- C_2 = temporal coefficient of variation,
- e = relative error = $E/\overline{X}$ where $\overline{X}$ = mean vehicle kilometers of travel per kilometer, and
- n = number of kilometers of roadway to count.

Application of these formulas assumes only one count per link. This approach is reasonable since spatial variations are substantially greater than temporal variations. The formulas should be applied separately to (a) local streets, (b) arterial streets, and (c) freeways.

Stratified Random Sampling—Optimal Allocation

$$n \simeq (\Sigma W_h S_h)^2/[(E_2^2/Z^2) + (1/N)(\Sigma W_h S_h^2)] \tag{9}$$

$$n_0 \simeq [Z^2(\Sigma W_h S_h)^2]/E_2^2 \tag{10}$$

where

- N = total kilometers of streets,
- W_h = weight of stratum h = N_h/N,
- N_h = number of kilometers of streets in stratum h,
- S_h = assumed composite standard deviation of vehicle kilometers per kilometer in stratum h $\simeq$ composite standard deviation of population of counts,
- Z = normal variate,
- E_2 = absolute error in average vehicle kilometers per kilometer = {relative error (e)} × {total vehicle kilometers ÷ total kilometers of roadway},
- n_0 = number of kilometers to count (without finite population correction), and
- n = number of kilometers to count (with finite population correction).

The second term in the denominator of Equation 9 represents the finite population correction factor. This term generally will be insignificant for arterial streets but should be computed for freeways.

After the sample size (n) is determined, the number of kilometers to count in any stratum h (i.e., n_h) can be computed according to the following formula:

$$n_h = n(N_h S_h/\Sigma N_h S_h) = n(W_h S_h/\Sigma W_h S_h) \tag{11}$$

ARTERIAL STREET GUIDELINES

The stratified sampling formula can be simplified for application to arterial streets. An equivalent constant standard deviation can replace S_h in each stratum in Equations 9 and 10 to simplify computations. It would take into account both spatial and temporal variations.

$$n_0 = (Z^2K^2S_1^2)/(e\overline{X})^2 \tag{12}$$

and

$$n = KS_1^2/[(e\overline{X}/Z)^2 + (1/N)K^2S_1^2] \tag{13}$$

In these formulas

$$n = n_0 f \tag{14}$$

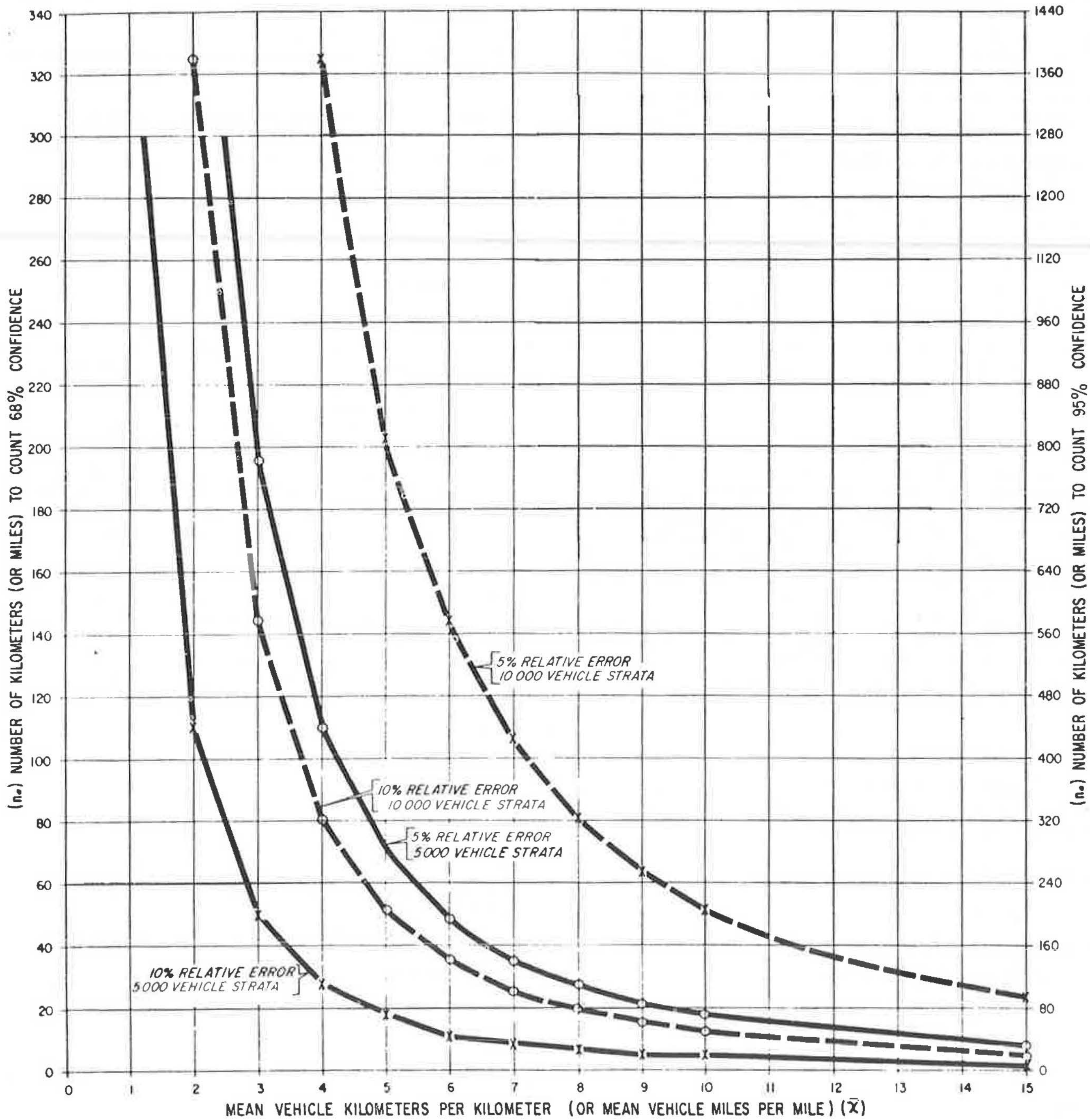
and

$$f = 1 + (Z/e)^2(1/\overline{X}^2)[(1/N)K^2S_1^2] \tag{15}$$

where K = equivalency factor. KS_1 represents the equivalent standard deviation, and S_1 equals 30 percent of the width of volume strata; f, of course, represents the finite population correction factor, and $\overline{X}$ represents the mean vehicle kilometers of travel per kilometer of road. (It can be shown that, where S_1 is always $> S_2$, $K \cong 1.414$.)

Values of K were obtained empirically by computing the composite spatial and temporal variations in six cities. These computations produced K values of about 1.4 for 5000-vehicle strata and 1.2 for 10 000-vehicle

Figure 3. Approximate sample sizes for arterial-street vehicle kilometers of travel estimates—stratified sampling.



strata based on one-day counts. This, in turn, led to the following formulas for 5000- and 10 000-vehicle classes:

5000-volume strata

$$n_0 \simeq Z^2[(2100)^2/(e\bar{X})^2] \quad (16)$$

10 000-volume strata

$$n_0 \simeq Z^2[(3600)^2/(e\bar{X})^2] \quad (17)$$

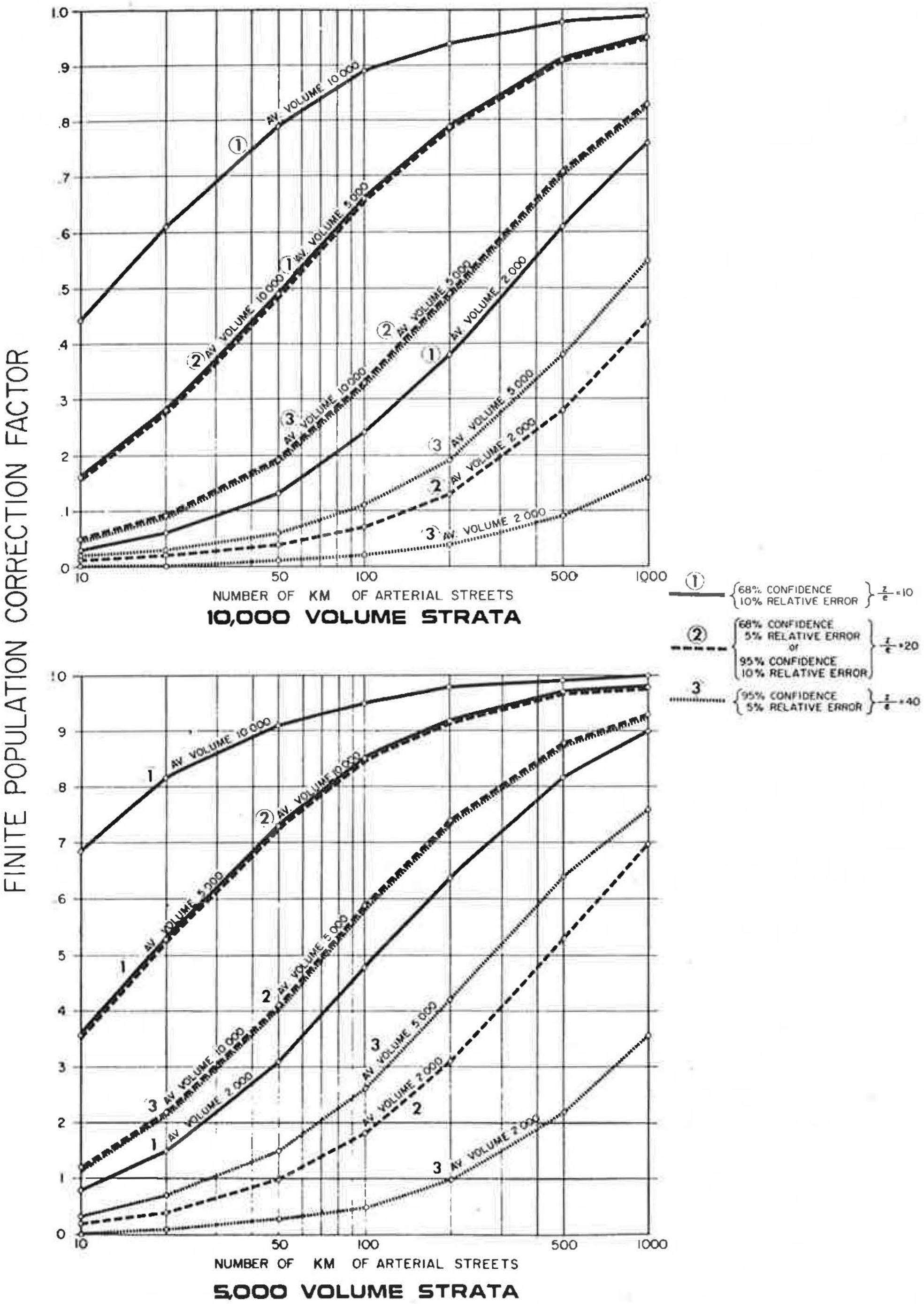
Figure 3 graphs the results of Equations 16 and 17, and Figure 4 graphs Equation 14. These two charts provide a practical guide for the estimation of sample size along arterial streets.

An illustrative application is as follows: Assume that, based on 5000-vehicle strata, there are about 5000 arterial vehicle-km of travel/km of roadway and about 500 km of road in the system. If we want to estimate the mean vehicle kilometers of travel per kilometer within a 5 percent relative error at 95 percent confidence, Figure 3 suggests that the preliminary sample size should be 280. Since 280 is large (relative to 500), it is necessary to correct for population size.

Figure 4, for 5000-vehicle strata, is used, assuming a Z/e of (2/0.05 or 40). The reductive factor is found to be 0.64. Thus, 0.64 × 280 or 180 km should be measured.

If the links average 0.5 km, then, 180/0.5 or 360 links should be counted.

Figure 4. Finite population correction factors.



FURTHER CONSIDERATIONS

The objective of establishing a vehicle-kilometer-of-travel sampling program is to provide meaningful estimates of travel by road class. The estimation of sample size is only a point of departure. The next step is to count locations, expand counts to represent total population sampled, and assess the accuracy of the results. Thus, there should be a minimum practical sample size in each stratum. This minimum should take precedence over values obtained by formula. Consequently, unduly small sample sizes should be avoided.

1. In theory, there should be at least two observations per stratum. However, to obtain reliable variance estimates, 6-10 observations are desirable.
2. Where classification of roads is by type of road, a minimum sample size of 30 is suggested. This will result in a sample standard error of estimate to within ±15 percent at 68 percent confidence.
3. Finally, after the sample is collected, its standard error should be computed.

Levels of confidence and allowable error should be established by participating agencies. These should reflect specific urban transportation planning needs and growth trends.

The suggested vehicle kilometers of travel procedures are practical and are based on sampling theory. The random selection of days and locations to count obviates the need to apply specific adjustment factors. It is hoped that cities and states will apply these procedures in the development of their initial sampling plan. As information on the reliability of vehicle kilometers of travel estimates are assembled, it will be possible to refine methods and pinpoint parameters.

Further research is desirable to obtain information on the variance of the distribution of freeway, arterial, and local street links. This information will make it possible to directly estimate the vehicle kilometers per link, thereby allowing greater clarity in sampling frames and procedures.

ACKNOWLEDGMENT

This paper is based on research undertaken for the Federal Highway Administration. The views represented are ours. We especially appreciate the insight of Nathan Lieder, former statistician of the Office of Planning, Federal Highway Administration.

REFERENCES

1. M. McClintock. Short Count Traffic Surveys and Their Application to Highway Design. Portland Cement Assn., Chicago, Highway Planning and Design Series No. 3, 1934.
2. V. H. Peltz and S. T. White. Short Count Results, 1946-1948. Traffic Engineering, Aug. 1948.
3. Wilbur Smith and Associates. General Analysis of Traffic Auditing Procedures for Outdoor Advertising. Outdoor Advertising of America, 1958.
4. J. A. Timbers. Traffic Survey at 1300 Sites. Ministry of Transport, London, Road Research Laboratory Rept. LR 206, 1968.
5. Wilbur Smith and Associates. Traffic Counting Program for Australia. Commonwealth Bureau of Roads, Melbourne, Jan. 1970.
6. H. S. Levinson and A. Roark; Wilbur Smith and Associates. Urban Traffic Volume Counting Manual. Federal Highway Administration, Sept. 1974.
7. W. C. Cochran. Sampling Techniques. 2d Ed. Wiley, New York, 1963.

Publication of this paper sponsored by Committee on Passenger Travel Demand Forecasting.

Abridgment

Empirical Comparison of Various Forms of Economic Travel Demand Models

Chong K. Liew and Chung J. Liew, Department of Economics, University of Oklahoma, Norman

Transportation planners are interested in assessing future conditions of intercity travel demands and in knowing the passenger's response to a fare hike. Planners need to be able to forecast correctly how a reduction in air fare would affect the passenger demand for airlines and other competing modes such as rails and buses.

To answer these questions, we introduce several different forms of demand equations, which were developed by many economists (1, 2). To be consistent with the theory of consumer behavior, all demand equations should satisfy three basic properties: homogeneity, summability, and symmetry. Traditional demand analysis in intercity travel demand (3-5) has never explicitly introduced the three properties in the formulation of demand equations. In many cases, the final form of reduced equations becomes a double-log form. The own and cross elasticities are a popular tool for the evaluation of the passenger's response to the price hike. When the signs of the estimated parameters are inconsistent with their experiences, an inequality-constrained double-log equation is often introduced to impose correct signs (4, 6). The market cross elasticities fail to correctly measure the substitutability among alternative travel modes because those elasticities include the income effects.

We adopt five popular demand models:

1. A double-log demand model,
2. An inequality-constrained double-log demand model,
3. A weighted Stone model (7),
4. The Rotterdam system of demand equations (2), and
5. A homogeneous translog demand model (8).

The last three models have firm foundations in the theory of consumer behaviors, and the parameter estimation of the models has been done by imposing three basic properties (i.e., homogeneity, summability, and symmetry). The first two models have very loose ties with the theory of consumer behavior.

One feature of our analysis is a comparative study that answers the following questions: (a) Does the choice of functional form matter in predicting substitutability among intercity travel demand? (b) Is it necessary to tie the model to the theory of consumer behavior to get a reliable result? and (c) Are market cross elasticities proper indicators of substitutability?

Another interesting feature of our model is the use of a compensated demand concept. Conventional intercity travel demand models (1, 3-5, 9, 10) fail to introduce this theoretically important and useful concept. The compensated demand concept can be used to correctly measure both consumer surplus and substitutability.

Our demand analysis differs from conventional intercity modal-split models in one important aspect. The conventional models employ trips as the variable of interest whereas our model employs the distance of travel. Use of travel distance instead of trip simplifies conceptual understanding of intercity travel-demand behavior by excluding trip-related variables, such as trip origin, destination, and length. Furthermore, travel distance, which is a continuous variable, directly ties with many policy-related variables, such as energy consumption in transportation, accident frequency rates, and pollution control measures.

THE MODELS

We assume that a consumer has an additively separable utility function in terms of several group commodities such as food, intercity travel, clothing, energy, and leisure. The consumer maximizes his or her utility in terms of these group commodities, which have money and time constraints. From the first-stage maximization, the consumer decides how much money (M) is required for the intercity travels and how much time (T) he or she can allocate for the intercity travels.

At the second stage, the consumer allocates the intercity travel budget (M) and travel time (T) on various travel modes so as to maximize his or her utility. The usual Lagrangian solutions provide the derived demand equations, which are expressed in terms of unit costs (p_i, $i = 1, \ldots, n$), speed (t_i, $i = 1, \ldots, n$), time (T), and money (M) budgets for the intercity travel demands:

$$x_i = x_i (p_1, \ldots p_n, t_1, \ldots t_n, M, T) \tag{1}$$

where $i = 1, \ldots, n$. Equation 1 is the usual starting point of empirical demand equations. We have selected four popular forms of demand equations.

Double-Log Form

The double-log demand equation provides various market elasticities directly from the estimated coefficients. However, there is no guarantee that all estimated coefficients will have a right sign. To avoid such difficulty, we introduce the inequality-constrained double-log demand equation.

Inequality-Constrained Double-Log Form

We may impose positive signs on all cross elasticities and negative signs on all own elasticities. However, they may not satisfy the basic properties of the theory of consumer behaviors.

Weighted Stone Model

We impose the summability, homogeneity, and symmetry conditions on the Stone model (7). This can be done by multiplying the budget share to Stone's demand equation and properly restricting the values of parameters:

$$s_i \log x_i = \sum_{j \in C} (b_{ji} \log p_j) + b_{mi} \log M^* + b_{ti} t + b_{si} \log SR + b_{oi} \tag{2}$$

with the restrictions

$$\sum_{i \in C} b_{mi} = 1 \text{ (summability)} \tag{3}$$

$$b_{ji} = b_{ij} \text{ (symmetry)} \tag{4}$$

$$\sum_{i \in C} b_{ji} = 0 \text{ (homogeneity)} \tag{5}$$

where

$b_{mi} = s_i a_{mi}$,
$b_{ji} = s_i a_{ji}$,
$b_{ti} = s_i a_{ti}$,
$b_{si} = s_i a_{si}$,
$b_{oi} = s_i a_{oi}$,
$\log M^* = \log M - \sum_{j \in C} (s_j) \log p_j$,
C = air, bus, or rail mode,
s_j = the money budget share of j th mode,
SR = the ratio of airline speed to the bus-rail speed, and
t = time trend.

a_{ji}, a_{mi}, a_{ti}, a_{si}, and a_{oi} are the parameters of Stone's demand equation and a_{ji} for j, $i \in C$ are compensated elasticities.

Rotterdam System of Demand Equations

$$s_i d \log x_i = \sum_{j \in C} (b_{ji} d \log p_j) + b_{mi} d \log M^* + b_{ti} + b_{si} d \log SR \tag{6}$$

(with the same restrictions as in Equations 3-5)

Homogeneous Translog Demand Equations

Alternatively, we may begin with a specific form of indirect utility function. We assume that the consumer has a homogeneous translog indirect utility function. By using Roy's identity, we have the following budget share equations:

$$s_i = -\sum_{j \in C} [b_{ji} \log (p_j/M)] + a_{si} \log (SR) + a_{ti} t + a_{oi} \tag{7}$$

The symmetry and homogeneity in Equations 4 and 5 and normalization ($\Sigma a_j = -1$) are imposed.

DESCRIPTIONS OF DATA AND EMPIRICAL RESULTS

Intercity passenger kilometers, prices per passenger kilometer, and number of passengers by each mode (airline, bus, and railroad) for 1947-1974 were collected from Transportation Facts and Trends (11). The average annual speed of the airline service is obtained from the Handbook of Airline Statistics (12). The average speed of bus and rail is gathered from Federal Highway Administration (FHWA) and Amtrak, respectively. The speed data of the rail mode include not only intercity trains but also suburban trains, including waiting time, whereas the speeds of airline and bus are the average maximum speed of the trip, not including waiting time. Because of this factor, the difference in speed between the bus and rail modes is not great. Hence the speed of the airline mode versus the speed of the bus-rail mode is considered.

An ordinary least-squares estimation was used to estimate the parameters of the double-log model. Some of the estimated coefficients in the double-log model turn out to have wrong signs. For example, several market cross elasticities turn out to be negative.

We impose correct signs on the parameters of the double-log demand model and estimate the parameters. We use an inequality-constrained least-squares estimation method (13, 14).

These two models fail to satisfy the homogeneity, symmetry, and summability conditions. We impose the three conditions on Stone's model, the Rotterdam system of demand equations, and the homogeneous translog demand models. Parameters of these models are estimated by the nonlinear maximum-likelihood estimation method. Table 1 gives the results of parameter estimations.

The latter three models—weighted Stone (WS), Rotterdam (RD), and the homogeneous translog (HTD)—provide fairly consistent empirical results, particularly in compensated cross elasticities, income elasticities, and own price elasticities. The first two models—double-log (DL) model and inequality-constrained double-log (ICDL) model—provide rather inconsistent empirical results. The market cross elasticities, which include the income effects, substantially differ, depending on the choice of the models. The results of demand elasticities are given in Table 2.

All five models predict the own market price elastic-

Table 1. Parameter estimation.

	Double Log		Inequality-Constrained Double Log		Weighted Stone		Rotterdam		Homogeneous Translog	
Equation	Parameter	t-Value	Parameter	t-Value	Parameter	t-Value	Parameter	t-Value	Parameter	t-Value
Air										
Air fare	-1.08	-20.4	-1.04	-7.53	-0.139	70.9	-0.120	-16.0	-0.045 7	-2.87
Bus fare	0.084 1	1.36	0.212	1.32	0.039 0	16.3	0.031 2	5.05	0.012 7	1.94
Rail fare	-0.063 2	-2.42	0.010 0	-[a]	0.099 8	75.0	0.089 3	11.6	0.033 0	2.53
Money budget	1.15	13.3	1.02	5.12	0.836	141.0	0.857	55.3		
Time trend	-0.006 77	-5.71	-0.008 45	-2.23	-0.003 96	-5.86	-0.006 09	-2.88	0.003 35	4.97
Speed ratio	0.153	5.14	0.186	2.19	0.120	5.35	0.130	2.75	-0.136	6.20
Intercept	0.005 93	0.035 0	0.397	0.892	-0.277	-10.7			-0.642	-20.2
R^2	0.998		0.997		0.996		0.859		0.607	
D-W statistic	1.34		1.48		0.646		1.34		0.784	
Bus										
Air fare	0.446	1.39	0.338	1.13	0.039 0	16.3	0.031 2	5.05	0.012 7	1.94
Bus fare	-0.859	-2.29	-0.965	-1.93	-0.044 7	-6.44	-0.032 3	-4.02	-0.005 54	-0.786
Rail fare	0.326	2.06	0.298	1.49	0.005 73	1.09	0.001 10	0.157	-0.007 19	-1.93
Money budget	-0.207	-0.394	0.010 0	-[a]	0.046 7	2.27	0.031 8	2.30		
Time trend	0.002 91	0.403	0.003 75	0.311	0.000 059 5	0.141	0.000 553	1.36	0.000 195	0.833
Speed ratio	0.508	2.81	0.459	1.84	-0.032 8	-2.81	0.001 66	0.194	-0.015 7	-2.79
Intercept	-5.01	-4.87	-4.98	-3.00	-0.082 1	-4.62			-0.027 5	-2.71
R^2	0.712		0.711		0.914		0.041 7			
D-W statistic	1.16		1.13		0.845		1.68			
Rail										
Air fare	0.537	1.59	0.463	1.54	0.099 8	75.0	0.089 3	11.6	0.033 0	2.53
Bus fare	-0.650	-1.64	0.010 0	-[a]	0.005 73	1.09	0.001 10	0.157	0.007 19	-1.93
Rail fare	-0.666	-4.00	-0.585	-2.23	-0.106	-25.3	-0.090 4	-7.90	-0.025 8	-2.26
Money budget	0.271	0.491	0.010 0	-[a]	0.117	7.26	0.111	5.33		
Time trend	0.048 8	6.45	0.003 76	4.01	-0.003 79	-5.82	0.003 05	2.76	0.003 54	-6.17
Speed ratio	-1.26	-6.61	-1.13	-3.65	0.136	6.56	-0.075 8	-3.12	0.152	7.96
Intercept	-2.84	-2.63	-1.20	-0.980	-0.415	-16.3			-0.330	-12.2
R^2	0.869		0.848		0.803		0.386		0.706	
D-W statistic	1.53		1.20		0.784		1.62		0.979	
Log of likelihood function					401.0		388.8		248.3	

[a] Denotes the bounded value.

Table 2. Demand elasticities.

Elasticities	Double-Log	Inequality Constrained	Weighted Stone	Rotterdam	Homogeneous Translog
Own price					
Air	-1.08	-1.04	-1.00	-1.01	-0.945
Bus	-0.859	-0.965	-1.15	-0.830	-0.863
Rail	-0.666	-0.585	-0.979	-0.846	-0.790
Income					
Air	1.15	1.02	1.00	1.03	1.00
Bus	-0.207	0.010 0	1.15	0.785	1.00
Rail	0.271	0.010 0	0.951	0.902	1.00
Speed					
Air	0.153	0.186	0.144	0.156	0.163
Bus	0.508	0.459	-0.810	0.041 0	0.390
Rail	-1.26	-1.13	1.11	-0.616	-1.24
Cross elasticity					
Air-bus fare	0.084 1	0.212	0.006 2	-0.004 40	-0.015 2
Air-rail fare	-0.063 2	0.010 0	-0.004 0	-0.020 0	-0.039 4
Bus-air fare	0.446	0.338	0.002 0	0.114	-0.315
Bus-rail fare	0.326	0.298	0.000 45	-0.069 4	0.178
Rail-air fare	0.537	0.463	0.016 0	-0.028 0	-0.268
Rail-bus fare	-0.650	0.010 0	0.008 1	-0.027 6	0.058 5
Compensated cross elasticities					
Air-bus fare	0.131	0.253	0.046 7	0.037 3	0.025 2
Air-rail fare	0.077 8	0.135	0.119	0.107	0.083 5
Bus-air fare	0.273	0.346	0.963	0.770	0.522
Bus-rail fare	0.301	0.299	0.141	0.027 2	0.301
Rail-air fare	0.764	0.471	0.811	0.726	0.568
Rail-bus fare	-0.639	0.010 4	0.046 6	0.008 94	0.098 9

ity of airline demand to be near 1: -1.08 for DL model, -1.04 for ICDL model, -1.00 for WS model, -1.01 for RD model, and -0.945 for HTD model. The bus and rail own price elasticities vary, depending on which model we choose. WS predicts as high as -1.15 for bus and -0.979 for rail services, whereas ICDL predict as low as -0.585 for rail service. RD predicts -0.830 for bus. All five models correctly predict the sign of the own price elasticities of all three modes. The income elasticities of intercity airline demand range from 1.00 to 1.15. WS, RD, and HTD predict the income elasticities of railroad demand to be very close to 1.0. DL and ICDL appear to predict very low income elasticity of rail. The income elasticity of bus demand varies from 1.15 by WS to 0.785 by RD.

A change in price of a transportation mode affects not only the demand for that mode but also the demand for the alternative modes. The latter is measured by the cross elasticities. However, the market cross elasticities do not correctly measure the substitutability among alternative modes since the elasticities include the income effects. We notice that the empirical results of the market cross elasticities become substantially different depending on the choice of the model. For example, the cross elasticity of airline price with respect to railroad demand varies from 0.537 by DL model to -0.268 by HTD model. The unstable nature of the market cross elasticities seems to originate from the unstable income effects. As we take income effects out and constrain the homogeneity, summability, and symmetry conditions, the compensated cross elasticities (predicted by different models) become relatively close values.

The compensated cross elasticities that exclude the income effect correctly measure the degree of substitution among alternative transportation modes. The RD model has all negative cross elasticities except one case. However, when we take the income effect out, all compensated cross elasticities become positive, which implies that airline, bus, and rail intercity demands have a substitutional relationship. Similar results are obtained in the WS and HTD models.

Since the market cross elasticities do not correctly provide the degree of substitutability, our discussion concentrates on the compensated cross elasticities. WS, RD, and HTD predict that a change in air fare causes a significant change in demand for bus and rail. The models predict that a reduction in air fare could severely reduce the demand for intercity rail and bus travel. For example, WS model predicts that a 1 percent decrease in air fare could cause a 0.963 percent decrease in bus travel demand and 0.811 percent decrease in rail travel demand. RD model predicts that the decreases will be 0.770 percent for bus travel demand and 0.726 percent for rail travel demand when the air fare drops by 1 percent. The HTD model predicts slightly lower percentages, 0.522 for bus demand and 0.568 for rail demand.

Neither the change in bus fare nor the change in rail fare significantly affects the intercity travel demand for airline services. Passengers are more sensitive to air fare change than to bus fare or rail fare change.

The speed ratio elasticity of the airline equation ranges from 0.144 by the WS model to 0.186 by the ICDL model. This implies that, as the speed of airline service increases 1 percent faster as compared with the bus-rail speed, it attracts more passengers to the airline industry by 0.144 percent, according to the WS model. As is expected, the rail industry loses its passengers. All models except the WS model forecast that the rail industry is the major victim of air speed increases. For example, a 1 percent speed increase in the airline industry decreases the rail passenger demand by 1.26 percent (DL model), by 1.13 percent (ICDL model), by 0.616 percent (RS model) and by 1.24 percent (HTD model), respectively. However, the air speed impact on bus becomes positive except for the WS model. The elasticities range from 0.508 by DL to 0.041 by RD. Possible sources of the wrong sign are due to the model specification errors or partly due to the errors from aggregation. Bus is mainly used for shorter trips. A data stratification by trip distance or by trip purpose or inclusion of automobile could improve the empirical results.

CONCLUSION

Our empirical results indicate that the family of demand equations that are imposed by the homogeneity, summability, and symmetry conditions provides more stable results on the compensated cross elasticities than those equations that do not have such conditions imposed. In general, the market cross elasticities are very unstable and they vary depending on the choice of functional forms, even when we impose the three basic conditions on their demand equations.

Our conclusion is that it is desirable to impose the homogeneity, summability, and symmetry conditions. The market cross elasticities are theoretically improper and empirically unstable in measuring the substitutability of intercity travel modes. Theoretically sound and empirically stable indicators of substitutability are the compensated cross elasticities. When the three conditions are imposed, the choice of functional forms yields minimal variation on the compensated cross elasticities.

ACKNOWLEDGMENT

The Division of Economics and Center for Economics and Management Research provided administrative support for this research. We are grateful to Dale W. Jorgenson, Daniel Brand, Antti Talvitie and other referees for their valuable comments on the earlier version of the paper.

REFERENCES

1. J. A. C. Brown and A. S. Deaton. Models of Consumer Behavior: A Survey. Economic Journal, Vol. 82, Dec. 1972, pp. 1145-1236.
2. H. Theil. Theory and Measurement of Consumer Demand, Vol. 1. North-Holland, Amsterdam, 1975.
3. G. Kraft; System Analysis and Research Corp. Demand for Intercity Passenger Travel in the Washington-Boston Corridor, Part V. Office of the Undersecretary of Transportation, U.S. Department of Commerce, 1965. NTIS: PB 166 884.
4. J. M. McLynn and T. Woronka. Passenger Demand and Modal Split Models. U.S. Department of Transportation, Northeast Corridor Transportation Project Rept. 230, 1969. NTIS: PB 190 946.
5. R. E. Quandt and W. T. Baumol. The Abstract Mode Model: Theory and Measurement. Studies in Travel Demand, Mathematica, Vol. 2, Sept. 1966.
6. A. Talvitie. A Direct Demand Model for Downtown Work Trips. Transportation, Vol. 2, 1973.
7. R. Stone. The Measurement of Consumer Expenditure and Behavior in the United Kingdom, 1920-38. Cambridge Univ. Press, London, 1954, Chapter 18.
8. L. R. Christensen, D. W. Jorgenson, and L. J. Lau. Transcendental Logarithmic Utility Functions. The American Economic Review, Vol. 65, No. 3, June 1975.
9. J. C. Bennett, R. H. Ellis, J. C. Prokopy, and M. D. Cheslow. A Comparative Evaluation of In-

tercity Modal-Split Models. TRB, Transportation Research Record 526, 1974, pp. 83-92.
10. J. W. Billheimer. Segmented, Multimodal, Intercity Passenger Demand Model. HRB, Highway Research Record 392, 1972, pp. 47-57.
11. Transportation Facts and Trends, 12th Ed. Transportation Association of America, Washington, DC, 1975.
12. Handbook of Airline Statistics. Civil Aeronautics Board, 1974.
13. C. K. Liew. Inequality Constrained Least-Squares Estimation. Journal of the American Statistical Association, Vol. 72, Sept. 1976, pp. 746-751.
14. C. K. Liew and J. K. Shim. A Computer Program for Inequality Constrained Least-Squares Estimation. Econometrica, Vol. 46, No. 1, Jan. 1978, p. 237.

Publication of this paper sponsored by Committee on Passenger Travel Demand Forecasting.

Abridgment

Estimation of Demand for Public Transportation

Yehuda Gur, Urban Systems, Inc., Chicago
Don Kopec, Elizabeth Lowe, Eugene Ryan, and Anant Vyas, Chicago Area Transportation Study

This paper describes the mode-split model used by the Chicago Area Transportation Study (CATS). The model was formulated in 1972 and was first used in mid-1973 for the evaluation of the 1995 regional transportation plan (1, 2). Since then, the model has been used as an operational tool (3) and has been refined, recalibrated, and validated. The current operational version of the model is described in this paper.

CHARACTERISTICS OF THE MODEL AS A PLANNING TOOL

The mode-split model operates as an integral part of the CATS transportation planning process. The major products of the model are estimates of the number of trips by automobile and transit for each origin zone and trip purpose. If necessary, these estimates are further processed through the CATS planning models, including mode-specific trip-distribution models, to provide estimates of volumes on specific roads and transit lines. In other cases, where such detailed information is not required, estimated changes in transit and highway demand (in response to proposed policies) are directly applicable to the evaluation of those policies.

As with most mode-split models, the CATS model is sensitive to the levels of service provided by various transportation modes and to the socioeconomic attributes of travelers. However, the model is unique in its emphasis on the effect of access and egress service on the demand for transit. The model provides for an accurate description of the access and egress service, considering both the availability of various submodes and the variations in the level of service within zones due to spatial dispersion of trip ends. The model's capabilities make it possible to describe accurately a wide range of policies related to improvements in access and egress service and to estimate the effects of those policies on travel demand.

STRUCTURE AND OPERATION OF THE MODEL

The CATS mode-split model may be described as an application of Monte Carlo simulation principles to travel-demand analysis. It may also be described as an aggregation procedure, which facilitates the application of disaggregate mode-choice models (4, 5) by the use of aggregate data.

Straightforward applications of mode-choice models in planning are done in the following way. Data are collected on a sample of the population under analysis, including (for each trip) relevant socioeconomic characteristics, service attributes, and chosen mode. The sum of the individual choices, properly weighted, provides an unbiased estimate of the population's modal shares. The sample is used to estimate a mode-choice model. A policy to be analyzed is introduced into the sample as changes in the level of the attributes that are affected by the proposed policy, and the resulting changes in mode-choice probabilities are calculated. The changes in the sum of those probabilities are used as estimators of the expected changes in modal shares of the population.

Many successful applications along these lines have been documented (6-8); however, this method has a number of deficiencies that seriously limit its applicability. The most obvious are the cost and time required to collect the data and the inability to sample future populations. Other deficiencies include the difficulties of identifying the population affected by a given policy and of selecting an effective sample.

The CATS mode-split model uses the same conceptual approach; the difference is that a pseudosample rather than a real sample is used. The pseudosample is generated by sampling the frequency distribution of the attributes of the population under analysis. This approach permits full exploitation of the power of disaggregate models without a need for a real sample. The procedures for creating the sample are designed to operate not only within the limitations imposed by considerations of data availability and analysis costs but also with the provision of means for accurately describing a wide range of proposed policies.

Operation of the Model

The heart of the CATS mode-split model is a procedure that repeatedly generates individual samples and mode-choice responses. Each sample corresponds to a single simulated traveler and his or her response to the modal alternatives available for a simulated trip. The attributes of the simulated trip maker and the characteristics of the trip are generated by a random Monte Carlo sampling process. The sequence of operations for generating a sample is described in Figure 1.

The following data are needed to generate the sample:

1. The person-trip table;
2. Matrices that show the zone-to-zone line-haul travel characteristics (such as times and costs for transit and highway);
3. Frequency distributions of household income by zones;
4. Frequency distributions by zone of the distance from a trip end to the nearest rail transit or commuter rail station, nearest bus stop, and nearest feeder bus service; and
5. Walking distance versus parking cost for the zones in the Chicago central business district (CBD) and parking access for the remainder of the region.

All those attributes that do not have high intrazonal variability (for example, line-haul travel times) are assigned to the trip directly from the zonal data. The attributes that do vary substantially within the zone are assigned by random sampling of the corresponding frequency distributions.

When the trip characteristics are estimated, the probability that the trip will use each of the available modes is estimated via a logistic mode-choice model. The process is then repeated a reasonable number of times for each zone (or interchange), and the estimated modal probabilities are accumulated. The resulting mode shares are used as estimators of the total population's mode split.

Description of Transit Service Attributes

A transit trip is subdivided into three portions: origin

Figure 1. CATS mode-split procedure.

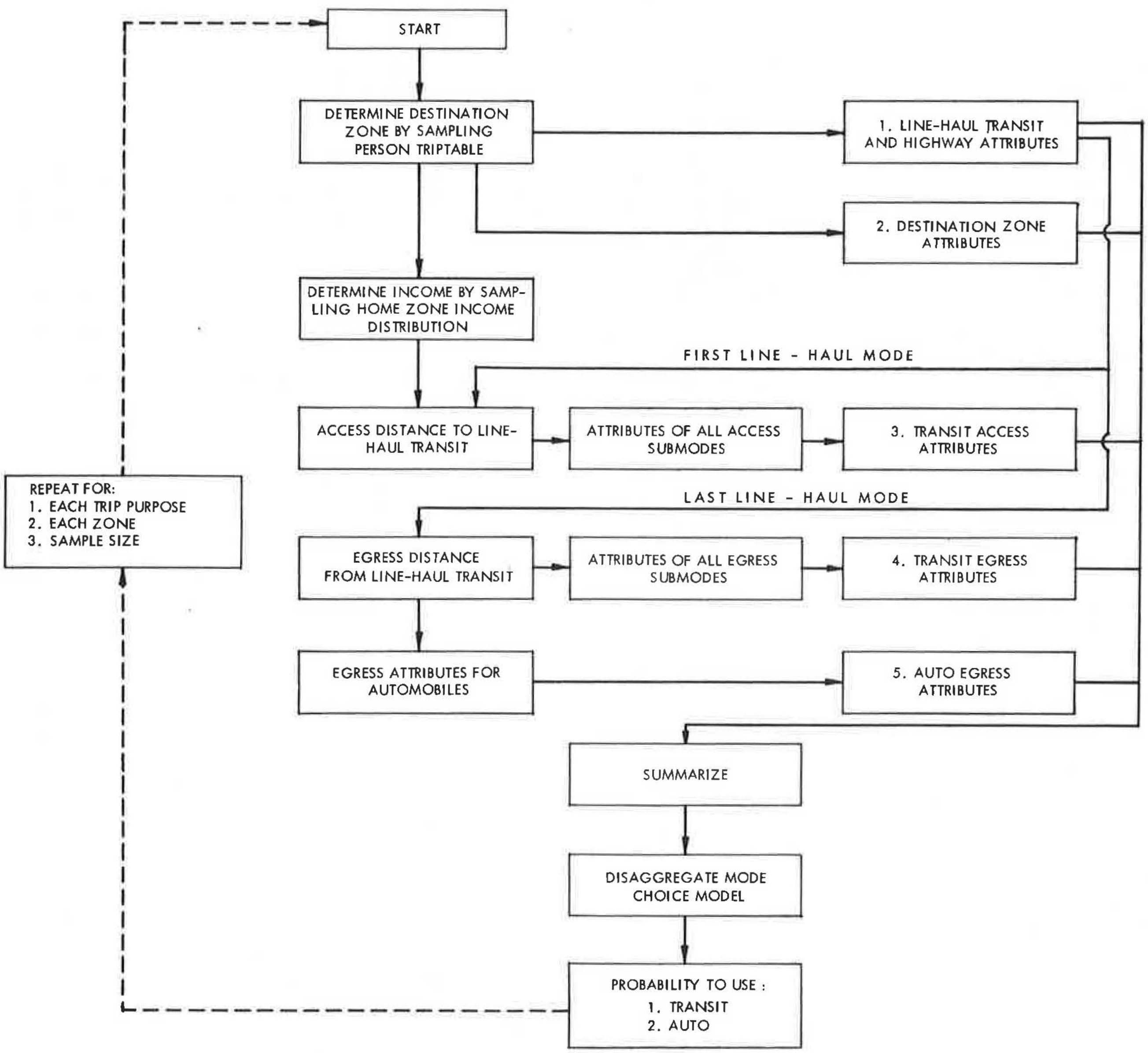


Table 1. Model variables and parameters.

	Work Trips								
	CBD			Non-CBD			Short Trips		
Coefficient	Final Value	Standard Error	T-Ratio	Final Value	Standard Error	T-Ratio	Final Value	Standard Error	T-Ratio
In-vehicle time	0.0276	0.0091	3.04	0.0024	0.0096	0.25	0.0077	0.0022	3.53
Cost	0.0121	0.0020	5.94	0.0152	0.0045	3.37	0.0187	0.0070	25.37
Excess time	0.0302	0.0098	3.08	0.0786	0.0214	3.68	0.0530	0.0030	17.49
Out-of-vehicle time	0.1139	0.0246	4.64	0.0616	0.0226	2.72	0.0070	0.0047	1.49
One-half first headway	0.0233	0.0160	1.46	0.0997	0.0341	2.92	0.0232	0.0040	5.84
Transit bias	-0.3594	0.1978	-1.82	0.3454	0.2065	1.67	1.4111	0.0382	36.97

Table 2. Mode-split: trips by transit.

		Work Trips					
		CBD		Non-CBD		Short Trips	
County	Population	Observed (%)	Predicted (%)	Observed (%)	Predicted (%)	Observed (%)	Predicted (%)
Cook	5 492 369	63.5	63.4	17.5	17.3	11.5	11.5
DuPage	491 882	69.9	60.8	2.1	4.7	1.0	2.0
Kane	251 005	72.3	66.5	0.9	2.5	0.5	0.9
Lake	382 638	68.5	61.0	1.1	2.7	0.6	0.9
McHenry	111 555	73.1	75.0	5.1	2.4	0.2	0.6
Will	249 498	52.1	55.4	0.0	1.3	0.0	0.4

approach (access), line-haul, and destination approach (egress). The attributes of the line-haul portion of the trip are determined from network analysis by using the Urban Transportation Planning System (UTPS) program UPATH (9). UPATH provides a description of the whole path from the origin to the destination, including access and egress. In order to separate the line-haul service characteristics, a special program strips the approach portions from the path. Approach is defined as walk and drive modes and, if the primary trip mode is rail, short bus legs. The program also records the first and last line-haul modes.

A special procedure is used to determine the approach impedance, given the line-haul modes. As part of the input, the program is provided with a description of the frequency distribution of the distance between the trip ends in each zone and the closest station. Separate distributions are provided for feeder bus, bus, rapid transit, and commuter rail. Those distributions can be estimated either mechanically (as a function of the service density) or manually.

The approach distance is determined by random sampling of the distribution corresponding to the first line-haul mode. In cases where feeder bus is a valid approach submode, the distance to the closest feeder stop is also determined by random sampling of the corresponding distribution. Given the approach distance, an estimate is made of the travel impedance by each of the available approach submodes. Submode impedances are estimated in terms of in-vehicle time, out-of-vehicle time, and cost. By using the parameters of the mode-choice equation, the total disutility by each of the submodes is computed, and the best submode is assigned to the trip. (Note that these data can be used readily to estimate submode split, if necessary.) The individual impedance measures of the selected approach submode are added to the line-haul measures for estimating the disutility for the whole trip and the mode-choice probabilities.

THE DISAGGREGATE MODE-CHOICE MODEL

The mode-choice model has been specified as a binary logit-choice model between automobile and transit:

$$P(t) = \exp(-\text{Transit disutility})/[\exp(-\text{Transit disutility}) + \exp(-\text{Automobile disutility})] \quad (1)$$

where P(t) = probability of selecting transit.

Table 1 lists the independent variables used for calculating the respective disutilities. Separate choice parameters have been specified for each of the following trip types:

1. Long residential trips (home-to-work trips) with CBD destination,
2. Long residential trips with non-CBD destination, and
3. Short trips (all trips other than home-to-work).

Model Estimation

A subsample of person trips from the 1970 home-interview data was used for model estimation. The home-interview data provided the origin zone, destination zone, and the mode used. Service attributes for the various modes were assigned by using engineering estimates based on the coded network and the mode-split model description of the access and egress service characteristics. One thousand observations were selected for each trip type. Table 1 summarizes the final estimated parameters by trip type.

Note that, in all three models, the parameters for in-vehicle time are rather low. This can be explained partly by the fact that the differences between automobile and transit in-vehicle travel times are generally not large. More important, this result of the calibration underscores the dominance of ease of access to transit as a determination of transit ridership.

Model Validation

The model was applied to the whole Chicago region by using 1970 data. The results were compared with the mode-split rates from the 1970 home-interview survey; a summary of the results appears in Table 2. Generally the model performed rather well in the developed parts of the region but tended to overestimate mode-split rates in the outlying areas. The reason for this tendency to overestimate is still being investigated; preliminary observations indicate that more accurate determination of the frequency distribution of distance to bus in large zones that have scant transit service will correct this tendency of the model. Another possible reason is the lack of socioeconomic variables (e.g., income or automobile ownership) in this calibration. More detailed analysis of the model estimates for the developed portions of the region indicate satisfactory performance, even for small areas.

CONCLUSIONS

The paper describes a methodology for mode-split analysis that possesses a number of highly desirable attributes: (a) it is compatible with the conventional transportation planning process, (b) it permits the application of disaggregate mode-choice models, and (c) it permits a detailed description of the access and egress transit service and a realistic account of its effect on transit ridership. The method for describing the service is flexible enough to support analyses of nonstandard services, such as dial-a-ride.

The model is fully operational and has been proven applicable for analysis of large-scale regional problems as well as for small-scale, subregional projects, including transportation system management strategies. The resources required for data preparation and analysis are reasonable.

REFERENCES

1. A. D. Vyas and E. V. Ryan. CATS' Network Sensitive Mode Split Model: User's Manual. Chicago Area Transportation Study, Chicago, In-House Rept., March 1974.
2. Technical Evaluation of the Recommended 1995 Transportation System in the Chicago-Gary Region. Chicago Area Transportation Study, June 1974.
3. E. V. Ryan. Calibration of the Network Sensitive Mode-Split Model. CATS Research News, Dec. 1974.
4. T. E. Lisco. The Value of Commuters' Travel Time. Univ. of Chicago, Ph.D dissertation, 1967.
5. D. McFadden. The Measurement of Urban Travel Demand. Journal of Public Economics, Vol. 3, Nov. 1974, pp. 303-328.
6. N. Tahir. Feeder Buses as an Alternative to Commuter Parking: An Analysis of Economic Tradeoffs. Office of Research and Development, Illinois Department of Transportation, Springfield, Technical Papers and Notes 15, Feb. 1974.
7. Cambridge Systematics, Inc. Urban Transportation Energy Conservation—Short Range Generalized Policy. U.S. Department of Energy, 1977.
8. B. D. Spear. Application of New Travel Demand Forecasting Techniques to Transportation Planning: A Study of Individual Choice Models. Federal Highway Administration, March 1977.
9. UTPS Reference Manual. Urban Mass Transportation Administration, 1975.

Publication of this paper sponsored by Committee on Passenger Travel Demand Forecasting.

Abridgment

Second Role of the Work Trip–Visiting Nonwork Destinations

Clinton V. Oster, Jr., Graduate School of Business Administration, Harvard University, Cambridge, Massachusetts

On a typical weekday in a major California urban area, about one-third of the households in which the head is employed visit one or more nonwork destinations as part of a trip to or from the workplace. Many transportation analysts find this number surprisingly high because conventional models of urban travel behavior make the assumption that the sole function of the work trip is to get the worker to and from the workplace. In fact, the analysis presented in this paper found that intermediate stops during trips to and from the workplace are an important means of visiting nonwork destinations and account for about 17 percent of nonwork destinations visited per household per weekday.

These figures are based on an analysis of home-interview origin-destination data collected in 1971 as part of the Fresno-Clovis area transportation study. An initial reaction to these numbers is to ask why Fresno is so at variance with the conventional wisdom. Fresno may be an unusual case; however, the use of complex patterns of travel found in Fresno is consistent with studies by Ginn and Horowitz of complex travel patterns in other cities (1, 2).

More likely, the conventional wisdom is no longer consistent with actual travel behavior. Cross-sectional evidence presented later in this paper implies that, if current demographic trends continue, the use of workplace-related trips to visit nonwork destinations will increase from the already substantial levels found in 1971. The conventional wisdom may be based on earlier data, collected when these types of trips were less important than they are now.

CHARACTERISTICS OF THE DATA

Before turning to the results of the analysis, briefly consider the strengths and weaknesses of the data on which they are based. The Fresno survey was used because the data were collected and organized in a disaggregated manner that permitted the analysis of complex travel patterns. These data reflect travel behavior before the oil emargo and subsequent increases in the price of gasoline. The data refer only to trips made by vehicle by persons age 5 and older. Thus, walking trips were excluded. Trips by vehicle include trips made by automobile, pickup truck, and public transit. Public transit, however, was relatively unimportant in Fresno; it accounted for only about 1 percent of trips to work.

The data were recorded in terms of person trips rather than vehicle trips. If a second person rode along on a trip made by another person, the trips were recorded as two separate trips.

Some additional restrictions were placed on the data used for this analysis. Only weekday travel was included because the character of weekend trips to the workplace might be substantially different. Destinations visited for related business purposes, such as a salesperson's call on customers, were excluded because these trips were considered to be an attribute of the person's job rather than a dimension of travel over which a person had much daily choice. Because the focus is on workplace-related trips to visit nonwork destinations and because the data were collected during July and August, when a relatively high incidence of vacation travel behavior could be expected, only the travel of household members when the head of the household made a work trip on the travel day was included. Only household members who had one or more automobiles (or pickup trucks) available were included. Finally, destinations refer only to nonwork destinations.

SAVINGS IN TRAVEL RESOURCES

A reasonable assumption is that a principal incentive for visiting a nonwork destination during a workplace-related trip (a trip from home to work, a trip from work to home, or a trip that originates and terminates at the workplace) is to obtain a savings in the time and money cost of travel, thereby lowering the total costs of the goods and services obtained via travel. The savings realized from making an intermediate stop during a workplace-related trip rather than making a separate single-destination trip from the home to accomplish the same purpose can be substantial.

To get an estimate of the order of magnitude of these savings, assume that, instead of visiting nonwork destinations via workplace-related trips, the household member visits different destinations for the same purposes via single-destination trips from the home. Further assume that the travel time and travel distance for these single-destination trips are the same as the average for all single-destination trips made for the same purposes by members of other households living in the same census tract. These assumptions represent the situation where close substitutes for the goods obtained via workplace-related trips are available at many locations throughout the urban area.

Under these assumptions, each time a household member visits a nonwork destination via a workplace-related trip an estimated 3.9 person-km (2.4 person miles) and 4.8 person-min of travel resources are saved. Recall that each weekday about one-third of the household members visit one or more nonwork destinations in this manner. Without such use of these trips, household members would use about 5 percent more travel resources per weekday.

If the destinations visited via workplace-related trips offer highly specialized goods or services not widely available, then an alternative method for calculating the travel savings would be to assume that the household member would visit the same destinations via single-destination trips. The savings calculated under this assumption are substantially higher. Each destination visited via workplace-related trips results in a savings of 6.3 person-km (3.9 person miles) and 6.6 person-min. Without such trips, household members would use about 7.6 percent more travel resources. This method is reasonable for some destinations but, in general, it almost surely overstates the savings from workplace-related trips and can be regarded as an upper bound.

VARIATIONS IN HOUSEHOLDS' USE OF WORKPLACE-RELATED TRIPS

If the use of workplace-related trips to visit nonwork destinations did not vary greatly among households, then we might argue that the omission from transportation analysis of consideration of this function of these trips is not serious. If, however, this use of workplace-related trips does vary systematically with differences in the characteristics of households, then ignoring this use may introduce systematic error into urban transportation analysis.

The analytical tool used to examine the relations between the characteristics of the household members and their use of workplace-related trips is ordinary least-squares multiple-regression analysis by using the daily travel of each household as the unit of observation. A linear additive functional form is used in all cases. If all of the household's characteristics are included in each equation, then the estimated effects are the separate effects of each individual characteristic, holding all other characteristics constant. The dependent variables measure the number of nonwork destinations visited during workplace-related trips and not the number of workplace-related trips themselves.

Table 1 contains the significant coefficients from the regression equations that describe the effects of differences in household characteristics on the total number of nonwork destinations per household per weekday, the number of these destinations, the fraction of these destinations, and the probability of visiting one or more of these destinations via workplace-related trips.

In addition to the independent variables presented in Table 1, the equations also contained variables for the number of household members less than five years of age, accessibility indices for the residence, the workplace of the head of the household, the workplace of a second worker, and dummy variables for the availability of a second or third automobile (or pickup truck). None of these variables had a significant impact on any of the dependent variables.

Household Income

Household income is total annual income for all household members measured in hundreds of 1971 dollars. As might be expected, an increase in household income increases the total number of nonwork destinations. We can presume that members of higher-income households have more goods and services in their consumption bundle and thus visit more destinations to obtain these goods and services. The effect of a $1000 increase in household income is not large; total nonwork destinations per household per weekday increase by 0.06 for an income elasticity of 0.17.

In addition to visiting more nonwork destinations, members of higher-income households place greater reliance on workplace-related trips to visit these destinations. Not only does the number of nonwork destinations visited via workplace-related trips increase with an increase in income, but the fraction of total nonwork destinations visited via such trips increases as well. A $1000 increase increases the number of destinations by 0.03, for an income elasticity of 0.51 and the fraction by 0.01 for an elasticity of 0.38.

The use of workplace-related trips has been seen to offer a savings in travel time and distance over the use of single-destination trips to visit nonwork destinations.

Table 1. Principal effects on the use of workplace-related trips.

Household Characteristics	Total Nonwork Destinations	Nonwork Destinations via Workplace-Related Trips	Fraction of Nonwork Destinations via Workplace-Related Trips	Probability of at Least One Nonwork Destination via Workplace-Related Trips
Annual household income				
$1000 increase	0.06	0.03	0.01	0.01
t-statistic	3.59	5.18	3.99	4.82
Elasticity	0.17	0.51	0.38	0.34
Household size age 16+				
One-person increase	3.34	0.08	-0.11	-0.15
t-statistic	25.9	1.79	-7.68	-7.64
Elasticity	1.65	0.24	-1.15	-0.88
Household size ages 5-15				
One-person increase	0.43	-0.06	-0.03	-0.04
t-statistic	6.73	-2.78	-5.47	-5.17
Elasticity	0.11	-0.10	-0.19	-0.13
Presence of second worker				
Second worker	-1.84	0.57	0.24	0.38
t-statistic	-2.56	2.25	3.49	4.02
Elasticity	-0.13	0.23	0.34	0.31
Total destinations				
One-destination increase	-[a]	-[a]	0.01	0.04
t-statistic			3.17	12.5
Elasticity			0.15	0.47

[a]Total destinations was not used as an independent variable in this equation.

The observed greater reliance on workplace-related trips by members of higher-income households suggests that the higher value of time usually associated with higher income increases the incentive to economize on the use of travel time and has an important impact on the composition of travel.

Household Size and Structure

The variables that characterize household size and structure are the number of household members age 16 and older and the number ages 5-15. Household members age 16 and older are eligible to have a driver's license and, if an automobile is available, have the potential for independent automobile travel.

An increase in household members age 16 and older (holding the number of workers constant) increases total nonwork destinations per household per weekday by 3.34. Only 0.08 of this increase comes via workplace-related trips so that the fraction via such trips declines by 0.11.

An increase in household members ages 5-15 increases total nonwork destinations per household per weekday by 0.43, but the number visited via workplace-related trips actually declines by 0.06 so that the fraction via such trips declines by 0.03.

An increase in household size increases the total number of nonwork destinations visited, in part because members of larger households are likely to have more goods and services in their consumption bundle and in part because there are more people to go places. Neither of these reasons provides an incentive for the household members to make a large change in the number of nonwork destinations visited via workplace-related trips.

Two-Worker Households

The presence of a second worker in the household (recall that all households in the sample have at least one worker) decreases the total number of nonwork destinations per household per weekday by 1.84 in spite of increasing the number of nonwork destinations visited via workplace-related trips by 0.57.

Because of the additional time spent at work, a two-worker household has less time available for travel than an otherwise identical one-worker household. As a result, the time savings realized from using workplace-related trips are more valuable and the incentive to make more use of such trips is greater. In addition, there are more opportunities to make such trips.

IMPLICATIONS FOR FUTURE TRAVEL PATTERNS

The systematic variation of household members' use of workplace-related trips with differences in household characteristics suggests that, as the demographic characteristics of the population change, the aggregate travel patterns of the population may also change. By projecting recent trends in changing demographic characteristics and using the elasticities of the coefficients presented in Table 1, changes in future aggregate travel patterns can be predicted.

If recent trends continue (3), over a five-year period household income would increase 10.4 percent, household size age 16 and older would decrease by 2.9 percent, household size ages 5-15 would decrease by 14.6 percent, and the fraction of households that have two workers would increase 8.8 percent. As a result, total nonwork destinations per household per weekday would decrease by 5.7 percent in spite of an increase in nonwork destinations visited via workplace-related trips of 8.2 percent. The fraction of nonwork destinations visited via workplace-related trips would increase 12.7 percent over a five-year period. The probability of visiting one or more nonwork destinations via a workplace-related trip, which in the aggregate can be interpreted as the fraction of households who would visit at least one such destination on a typical weekday, would increase 9.7 percent.

The magnitudes of these changes are subject to a great deal of uncertainty; however, the direction of change is clear. Workplace-related trips, already a frequently used means of visiting nonwork destinations in 1971, are likely to become increasingly more important if current demographic trends continue.

IMPLICATIONS FOR TRANSPORTATION PLANNING

The systematic variation in the use of workplace-related trips to visit nonwork destinations suggests that household members have more mechanisms by which they can adjust their travel behavior in response to changes in household characteristics than conventional transportation analysis recognizes. Household members would use these same adjustment mechanisms to respond to changes in the price of travel or changes in transportation policy. If so, then transportation analysis, which fails to account for these mechanisms, may have difficulty predicting the magnitude or even the direction of changes in travel patterns in response to these price or policy changes.

Empirical evidence suggests that household members used such mechanisms during the gasoline shortages of 1973-1974. A study by Peskin, Schofer, and Stopher of travel patterns of households in the suburbs north of Chicago found that the combining of nonwork destinations with work trips increased sharply during the shortage, as did the combining of single-destination trips into multiple-destination tours (4).

The analysis of transportation policies intended to divert commuters from the private automobile to other modes to help achieve air pollution or energy conservation goals must recognize the advantages of the private automobile in visiting nonwork destinations as part of workplace-related trips and the increasing importance of such use of these trips for many households. Incentive schemes that subsidize transit or penalize private automobile may not be as effective in diverting commuters as conventional, generalized cost analysis would imply.

The assumption that the sole function of the work trip is to get people to and from the workplace may have once been reasonable. However, as demographics change and emphasis on using transportation policy to help achieve air pollution goals and energy conservation goals increases, this assumption is becoming increasingly untenable (5). An understanding of the extent to which and the reasons why household members use workplace-related trips to visit nonwork destinations seems essential for effective transportation planning.

ACKNOWLEDGMENT

This paper is based on my Ph.D. dissertation, which was financed through a grant from the Motor Vehicle Manufacturers Association and through funds from the U.S. Environmental Protection Agency. I would like to thank John F. Kain, John R. Meyer, and Gregory K. Ingram for their guidance and comments and the Division of Planning of the California Department of Transportation for providing computer tapes of the Fresno-Clovis area data.

REFERENCES

1. J. R. Ginn. Transportation Considerations in an Individual's Activity System. Civil Engineering Department, Northwestern Univ., Evanston, IL, Ph.D. dissertation, 1969.
2. J. Horowitz. Effects of Travel Time and Cost on the Frequency and Structure of Automobile Travel. TRB, Transportation Research Record 592, 1976, pp. 1-5.
3. Statistical Abstract of the United States: 1977. 98th Ed. Bureau of the Census, U.S. Department of Commerce, 1977.
4. R. L. Peskin, J. L. Schofer, and P. R. Stopher. The Immediate Impact of Gasoline Shortages on Urban Travel Behavior. Federal Highway Administration, April 1975.
5. C. V. Oster, Jr. Household Tripmaking to Multiple Destinations: The Overlooked Urban Travel Pattern. Traffic Quarterly, Oct. 1978, pp. 511-529.

Publication of this paper sponsored by Committee on Passenger Travel Demand Forecasting.

Generalized Attributes and Shopping Trip Behavior

Patrick S. McCarthy, Department of Economics, Purdue University, West Lafayette, Indiana

Attitudinal data obtained from an impact travel survey of the San Francisco area was analyzed to determine the composition of generalized attributes that identify an individual's cognitive structure of shopping behavior. Once it was determined (by employing two measures of factorability) that factor analysis was an appropriate analytical tool, the data (stratified by residence and trip destination) were factor analyzed. The results indicate that each population's cognitive structure is unique, although in all cases a common set of generalized attributes was found to be important. For the respective populations, an index of satisfaction was developed for each of the generalized attributes. The index was used to investigate the relation between a population's cognitive structure and its socioeconomic profile. Based on tests of independence and gamma measures of association, the following attributes were significantly related to a population's satisfaction relative to alternative attributes of the shopping excursion: travel, mode, length of residence at current address, and age distribution. Among the implications of the analysis is that a set of attributes exists, independent of residence or trip destination, that should be incorporated into travel-demand models if shopping travel behavior is to be forecast accurately. Moreover, the extent of travel incurred in a shopping journey appears to significantly affect an individual's attitude structure of shopping activities.

Recent emphasis in transportation research has focused on the development of travel-demand models that seek to explain and subsequently predict, as accurately as possible, individual travel behavior (1-4). Concomitant with the shift toward disaggregate modeling has been the recognition that individual attitudes are important inputs into the decision process (5-11). As a result of its explanatory and predictive potential for travel behavior, therefore, attitudinal modeling and its associated analytical techniques are of widespread interest to transportation analysts.

In general, attitudinal modeling serves the travel forecaster in two ways:

1. Univariate or multivariate psychometric scaling techniques can be applied to define multifaceted transportation attributes, such as comfort and convenience,

that have hitherto been difficult to quantify and incorporate into travel-demand studies.

2. Attitudinal modeling can be employed as a preliminary analytic procedure to segment the travel market under study into homogeneous populations, according to the similarity of individual perceptions or preferences; separate travel-demand models are then estimated for the partitioned populations.

This study, whose domain primarily falls into the first category, seeks to determine the cognitive structure of shopping trip behavior that, accordingly, can be employed to define generalized attributes for use in travel-demand models. In addition, statistical tests of independence and measures of association are employed to determine whether the underlying dimensions are related to the socioeconomic characteristics of a population, thus providing implications for market segmentation.

Individuals in the San Francisco Bay area were asked to rate various shopping characteristics, which reflect attributes of traveling from home to a shopping area and attributes of the area itself, on a five-point Likert scale. After the data were stratified by residence (central city or suburb) and shopping trip destination [downtown San Francisco or local central business district (CBD)], the responses of each subpopulation were factor analyzed. The results suggest that each group's cognitive structure of shopping behavior is unique. However, for each subpopulation, a common set of generalized attributes exist on which individuals assess their shopping excursions.

When these results were used to investigate the relation between the common dimensions of a population's cognitive structure and its socioeconomic characteristics, travel mode, length of residence, and age were significant determinants of individual satisfaction relative to alternative attributes of the shopping activity. In addition, education, ethnicity, sex, and marital status were variously related to perceptions of shopping trips, although no general pattern was evident.

METHODOLOGY

The frame of reference for this study is conveniently illustrated with Golob and Dobson's (7) general schematic representation of the transportation decision process, reproduced in Figure 1. G_i is individual group i, T_j reflects a set of transportation alternatives, A_k $(k = 1, \ldots, n)$ represent those attributes of T_j described from a priori considerations, $\tilde{A}_k$ $(k = 1, \ldots, m)$ are those attributes derived from a set of judgments expressed by G_i, C_{ij} is the choice of alternative T_j by G_i, and D_{ij} is the realization of a decision by G_i towards T_j, which differs from C_{ij} due to intervening variables, IV_p. In this context, this study is concerned with the $\tilde{A}_k \leftarrow A_k$ interrelationships, where T_j $(j = 1,2)$ is the set of shopping trip destinations (downtown San Francisco and local CBD) and G_i $(i = 1,2)$ are individuals who reside in central city and suburban locales.

Factor analysis was employed to determine the relationships that exist between the A_k $(k = 1, \ldots, n)$ and $\tilde{A}_k$ $(k = 1, \ldots, m)$ and, to facilitate interpretation, the solution was rotated according to the VARIMAX criterion. The factor loadings were estimated by using Joreskog's maximum likelihood procedure (12), which also provides two measures of factorability (13). Bartlett's test of sphericity tests the hypothesis that the sample correlation matrix came from a multivariate normal population in which the variables of interest are independent. According to this test, the statistic

$$-[(N - 1)(1/6)(2p + 5)] \ln |R| \tag{1}$$

where N is sample size, p is the number of variables, and R is the correlation matrix, is approximately chi-square distributed as N becomes large, with $(1/2)p(p-1)$ degrees of freedom. Rejection of the null hypothesis indicates that the data are appropriate for factor analysis.

A second measure of factorability is the Kaiser-Meyer-Olkin measure of sampling adequacy, which assesses whether the variables belong together psychometrically and, therefore, whether factor analysis is a suitable analytical tool. A value of this index (which varies between 0 and 1) below 0.5 is a clear indication, according to Kaiser (14), that the data are not appropriate for factor analysis.

There also exists a goodness-of-fit test to determine whether the number of factors extracted is sufficient. It can be shown that the null hypothesis (that m factors are adequate for the generation of the observed correlation matrix) is based on a statistic that, as sample size increases, is approximately chi-square distributed with $[(1/2)(p - m)^2 - p - m]$ degrees of freedom (15). Following Green and Rao (16), eight factors were extracted initially and, if this null hypothesis could not be rejected at the 0.05 level, additional factors were extracted until the criterion was met. Only in one instance, discussed below, was this condition not satisfied.

After the latent dimensions of a group's cognitive structure are identified from the factor-analytic results, the next task is to determine whether any relationship exists between those generalized attributes common to each group and its socioeconomic characteristics. This was achieved in two steps.

First, given a set of generalized attributes $[\tilde{A}_k\ (k = 1, \ldots, m)]$, a dichotomous variable $[S_{ik}\ (i = 1, \ldots, I;\ k = 1, \ldots, m)]$ was defined for each individual such that S_{ik} received a value of one if individual i's shopping activity was satisfactory and zero if not satisfactory with respect to $\tilde{A}_{ik}$ $(i = 1, \ldots, I;\ k = 1, \ldots, m)$. The definition was made operational by the following assignment:

$$S_{ik} = 1 \quad \text{if } \tilde{A}_{ik} = \sum_{j=1}^{n} A_{ikj}/n > 3 \tag{2a}$$

and

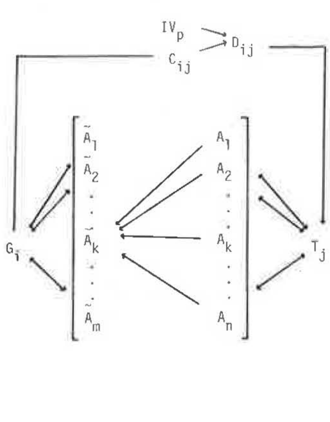
$$S_{ik} = 0 \quad \text{if } \tilde{A}_{ik} = \sum_{j=1}^{n} A_{ikj}/n \leq 3 \qquad (i = 1, \ldots, I;\ k = 1, \ldots, m) \tag{2b}$$

where

$\tilde{A}_{ik}$ = value of the kth generalized attribute assigned to individual i;

A_{ikj} = jth component of individual i's generalized attribute k and is the rating, on a five-point Likert scale, that individual i assigned to

Figure 1. Transportation decision process.



this shopping characteristic;

n = number of components in the kth generalized attribute;

m = number of generalized attributes;

3 = midpoint of Likert scale on which individuals assessed characteristics of their shopping activities.

Thus, for generalized attribute k, if an individual's mean rating for k exceeded 3, it is assumed that his or her shopping activity was satisfactory with respect to the generalized attribute; if less than or equal to 3, his or her shopping activity is assumed to be unsatisfactory in this regard.

Second, an individual's satisfaction rating was cross-tabulated with various socioeconomic characteristics. Let O_f be observed frequency, E_f be expected frequency (assuming no association between satisfaction rating and the socioeconomic characteristic), c and r be the number of columns and rows, respectively, in the contingency table, and L be the number of cells in the table. Then the statistic

$$\sum_{i=1}^{L} (O_f - E_f)^2/E_f \qquad (3)$$

has a chi-square distribution with (c - 1) (r - 1) degrees of freedom and has an associated null hypothesis that the satisfaction rating of a particular generalized attribute and a socioeconomic characteristic are statistically independent (17). If the null hypothesis is rejected at the 0.05 level (that is, the variables are not independent), then the direction and strength of the relationship is investigated by using a gamma measure of association. Gamma is defined as

$$\gamma = (n_s - n_d)/(n_s + n_d) \qquad (4)$$

where n_s and n_d are the number of concordant and discordant pairs, respectively, and are particularly useful in this context because, in absolute value, they reflect the proportion by which error in prediction of generalized attribute satisfaction is reduced from knowledge of the particular socioeconomic characteristic (18).

Table 1. Shopping activity attributes rated by respondents.

Attribute Number	Statement
V1	A person like me can dress informally when shopping in....
V2	A good variety of merchandise I like can be found in....
V3	The merchants stand behind goods they sell and provide reliable repair service in....
V4	A person will find the walkways and sidewalks uncrowded when shopping in....
V5	A person like me will find the area clean when shopping in....
V6	A person like me can easily get from store to store when shopping in....
V7	Low prices can be found for the merchandise I want in....
V8	Persons who drive will find it easy to park when shopping in....
V9	Shoppers will find the stores open evenings and weekends when shopping in....
V10	A person is safe from accidents when traveling as I do to shop in....
V11	Transportation, plus any parking, doesn't cost much when going as I do to shop in....
V12	I can start and return when convenient for me when going as I do to shop in....
V13	A person is safe from robbery or assault when going as I do to shop in....
V14	A person has a clean, attractive passenger area to ride in when traveling as I do to shop in....
V15	Getting there doesn't take as much time when going as I do to shop in....
V16	I know for sure when I will get there when traveling as I do to shop in....
V17	A person has a comfortable ride when traveling as I do to shop in....
V18	A person is protected from bad weather when traveling as I do to shop in....
V19	It's easy to stop at other places on the way when traveling as I do to shop in....

DATA

Data for this analysis were obtained from a 1973-1974 Bay Area Rapid Transit (BART) impact travel survey (19). The survey included four areas throughout the San Francisco Bay area, two in the East Bay and two in the West Bay; sample sizes were 814 and 910, respectively. In each bay one study area was selected to represent a central city, comprised of a sizable minority population, that had bus service available, and the second study area represented a suburban environment, predominantly white, that had little or no bus service, that is, automobile oriented. In the East Bay, Oakland's Fruitvale District and adjoining hill areas represented the central city, and the area in and around the city of Walnut Creek and extending south through Danville represented the suburban study area. In the West Bay, the Mission District in San Francisco was selected for the central city area, and the suburban study area began in the southwest corner of San Francisco and extended south through Daly City and Pacifica. For this analysis, the central city and suburban areas in each bay were combined to form aggregated central city and suburban populations, for total sample sizes of 807 and 917, respectively.

In the survey individuals were asked to rate, on a five-point Likert scale, 19 attributes that represent various aspects of a trip from home to a shopping center (located in downtown San Francisco or in the local CBD) and of the shopping area itself. Table 1 presents the statements that individuals were asked to evaluate. Only individuals who, in the previous 12 months, had made a shopping excursion to downtown San Francisco or to their local CBD to buy or look for major appliances and who rated the relevant shopping area on each of the 19 items were included in the analysis. The sample sizes for each model run are summarized below.

Population	Downtown San Francisco	Local CBD
Central city	285	417
Suburban	247	576

RESULTS

Tables 2-5 summarize the factor-analytic results obtained for central city and suburban populations who undertook shopping journeys to downtown San Francisco and local CBDs. Relative to the measures of factorability, it is observed from the tables that, in all cases, the Bartlett test of sphericity resoundly rejects the hypothesis that the sample correlation matrix came from a multivariate normal population whose variables are stochastically independent. This conclusion is buttressed by the Kaiser-Meyer-Olkin measure of sampling adequacy, which, according to Kaiser's calibration of the index (14), is in the meritorious range ($\geq$0.8 and <0.9) in all instances. With respect to the goodness-of-fit test, in all but one of the runs we could not, at the 0.05 level, reject the hypothesis that the number of factors extracted was adequate for generating the observed correlation matrix. The one exception was local CBD shopping trips made by the respondents in the central city areas, in which nine factors were ex-

Table 2. Factor analysis for shopping trips to downtown San Francisco—central city population.

Rotated Factor		Attribute	
Number	Description	Number	Factor Loading
1	Trip comfort	V18	0.772
		V14	0.674
		V17	0.648
		V19	0.528
		V15	0.319
2	Shopping area attraction	V7	0.615
		V6	0.417
		V15	0.359
		V9	0.352
		V3	0.339
		V8	0.331
3	Trip convenience	V15	0.651
		V16	0.621
		V12	0.492
		V11	0.302
4	Shopping congestion	V5	0.670
		V4	0.536
		V8	0.443
5	Range of merchandise	V2	0.975
6	Trip safety	V13	0.783
		V10	0.310
		V11	0.303

Notes: Percentage of total variance explained = 70.4; Bartlett test of sphericity = $\chi^2(171)$ = 1511.32 (significant at 0.01 level); Kaiser-Meyer-Olkin measure of sampling adequacy = 0.85; goodness-of-fit test = $\chi^2(47)$ = 39.06 (not significant at 0.05 level).

Table 3. Factor analysis for shopping trips to downtown San Francisco—suburban population.

Rotated Factor		Attribute	
Number	Description	Number	Factor Loading
1	Trip comfort	V17	0.793
		V14	0.766
		V18	0.678
		V19	0.468
		V12	0.309
2	Trip convenience	V16	0.750
		V12	0.531
		V15	0.473
		V11	0.423
3	Trip safety	V10	0.721
		V11	0.512
		V13	0.427
		V15	0.368
		V16	0.302
4	Range of merchandise	V2	0.977
		V3	0.470
5	Shopping congestion	V5	0.689
		V4	0.643
		V6	0.387
6	Availability of merchandise	V7	0.583
		V9	0.552
7	Trip flexibility	V19	0.494

Notes: Percentage of total variance explained = 71.3; Bartlett test of sphericity = $\chi^2(171)$ = 1295.32 (significant at 0.01 level); Kaiser-Meyer-Olkin measure of sampling adequacy = 0.81; goodness-of-fit test = $\chi^2(47)$ = 45.19 (not significant at 0.05 level).

Table 4. Factor analysis for shopping trips to local CBD—central city population.

Rotated Factor		Attribute	
Number	Description	Number	Factor Loading
1	Shopping congestion	V8	0.720
		V5	0.611
		V4	0.593
		V10	0.405
		V9	0.324
2	Trip convenience	V12	0.650
		V11	0.622
		V15	0.573
		V16	0.330
3	Shopping area attraction	V6	0.695
		V2	0.505
		V7	0.501
		V3	0.363
		V9	0.355
4	Trip comfort	V18	0.602
		V14	0.425
		V17	0.370
		V19	0.316
5	Riding comfort	V17	0.847
6	Trip safety	V13	0.486
		V10	0.397
		V12	0.325
7	Quality of merchandise	V3	0.887
8	Dependability	V16	0.692
		V15	0.327

Notes: Percentage of total variance explained = 7.20; Bartlett test of sphericity = $\chi^2(171)$ = 2272.74 (significant at 0.01 level); Kaiser-Meyer-Olkin measure of sampling adequacy = 0.87; goodness-of-fit test = $\chi^2(47)$ = 66.49 (not significant at 0.02 level).

Table 5. Factor analysis for shopping trips to local CBD—suburban population.

Rotated Factor		Attribute	
Number	Description	Number	Factor Loading
1	Trip convenience	V12	0.703
		V15	0.643
		V11	0.595
		V16	0.424
2	Shopping congestion	V4	0.653
		V8	0.544
		V6	0.349
		V9	0.302
3	Trip flexibility	V18	0.889
		V19	0.459
		V17	0.323
4	Shopping area attraction	V2	0.743
		V3	0.400
		V7	0.341
		V9	0.340
		V5	0.313
5	Riding comfort	V14	0.921
		V17	0.420
6	Trip safety	V10	0.887
		V13	0.317
7	Shopping area appearance	V5	0.621

Notes: Percentage of total variance explained = 68.0; Bartlett test of sphericity = $\chi^2(171)$ = 2629.74 (significant at 0.01 level); Kaiser-Meyer-Olkin measure of sampling adequacy = 0.86; goodness-of-fit test = $\chi^2(47)$ = 54.74 (not significant at 0.05 level).

tracted and the significance level was 0.032. This could not be improved on because convergence could not be achieved when more than nine factors were extracted.

Only interpretable factors are displayed in Tables 2-5 and those variables that load highly on each factor are presented, where a loading is defined to be salient if its value equals or exceeds 0.3 (20). For shopping excursions to downtown San Francisco, Tables 2 and 3 reveal similarities and differences in the cognitive structures of central city and suburban residents. Both populations exhibit many of the same dimensions in their attitude structures, including trip comfort, trip convenience, shopping area attraction, shopping congestion, and trip safety, but their order of importance is not identical. Each population views trip comfort to be of primary importance; however, trip convenience and safety constructs are second and third in importance for the suburban occupants, whereas they rank third and sixth for central city residents. Travel considerations are more important to the suburban community because, in their shopping journeys to downtown San Francisco, they must incur more travel and overcome more spatial friction. This point is further illustrated by noting that the ability to make stops along the way (V19) constitutes a separate factor for suburban residents, so that individuals who want to satisfy other objectives will undertake a multipurpose trip when traveling to shop in downtown San Francisco. Note also that central city inhabitants emphasize the overall attractiveness of downtown San Francisco as a shopping area (as observed in factor 2), whereas suburban dwellers highlight specific characteristics, including the range of merchandise (factor 4) and the availability of merchandise (factor 6), where the latter dimension characterizes time and monetary constraints under which a shopper operates. This possibly alludes to the spatial proximity of the respective shopping areas to an individual's residence. If a shopping activity entails low travel investment, then an individual will be concerned about the overall attractiveness of a shopping area. Conversely, if high travel investment is required, then more planning will occur and specific features of the shopping center will be accentuated. Finally, both areas emphasize shopping congestion, although suburban residents are not concerned about the ease of parking in downtown San Francisco.

Tables 4 and 5 indicate that, in local CBD shopping trips, each population emphasizes trip convenience, shopping congestion, shopping area attraction, riding comfort, and trip safety, although, analogous to the previous case, the rank order of these dimensions is varied. In addition to the differential emphasis placed on these factors, the two groups are distinguished by stressing other aspects of the shopping excursion. For central city residents, the quality of merchandise (V3) and dependability (V15 and V16) are important components in their cognitive structures; suburban dwellers underscore the appearance of the shopping area and trip flexibility, where the presence of the latter dimension may again reflect the fact that, even for local CBD trips, suburban vis-a-vis central city residents incur more travel and accordingly are more prone to make a multipurpose trip.

In summary, the cognitive structures of central city and suburban householders, both for downtown San Francisco and local CBD shopping journeys, are differentiated in two respects:

1. The cognitive structure of each group is not represented by the same set of latent factors; and
2. The importance of the factors common to each population are varied.

Notwithstanding these differences, important similarities exist in the structures of the respective groups. Scrutiny of the results suggests that five factors, or generalized attributes, are relevant to each population's shopping trip perceptions. The table below lists these underlying dimensions and identifies those variables primarily associated with them.

Generalized Attribute	Associated Variables
Trip convenience—TCONV	V11, V12, V15, V16
Trip comfort—TCOMF	V14, V17, V18, V19
Trip safety—TSAFTY	V10, V13
Shopping area attraction—SATT	V2, V3, V6, V7, V9
Shopping congestion—CONGEST	V4, V5, V8

Generalized trip convenience (TCONV) encompasses travel time and travel cost (V11 and V15, respectively) as well as other time considerations (V12 and V16) associated with making a shopping trip. Generalized trip comfort (TCOMF) not only reflects riding and vehicle comfort but also weather exposure and trip stopovers. Note that V19, the ability to make other stops along the way, is consistently associated with comfort rather than convenience aspects of the trip. The third generalized attribute is trip safety (TSAFTY), which reflects both vehicular safety (that is, safety from accidents on the mode traveled) and personal safety from robbery or assault when making the trip. Fourth is generalized shopping area attraction (SATT), which includes variety and servicing of merchandise available (V2 and V3, respectively) as well as a shopper's ability to move easily from store to store (V6). Moreover, a shopping area's attraction will be enhanced if its major appliances are priced low (V7) and its stores have weekend and evening business hours (V9). The last generalized attribute common across populations is generalized shopping congestion (CONGEST), which includes ease of parking (V8), crowding (V4), and clean shopping environment (V5).

As observed in Tables 2-5, all of the primary variables associated with each of the generalized attributes do not consistently have salient loadings and, in some instances, the salient loadings are split between two factors. Nevertheless, since, in the majority of cases, these variables are grouped together, they are combined to form the generalized attributes listed in the preceding table. Also notice that most variables that comprise a particular generalized attribute also act to reinforce other generalized attributes. For example, a salient loading on V12 denotes an underlying trip convenience dimension. However, when associated with V14, V17, V18, and V19 [as in Table 2 (factor 1)], it acts to reinforce generalized trip comfort such that an increase in an individual's option to start and return when convenient enhances overall comfort of the trip. This highlights the multifaceted characteristic of the rated items and underlies the difficulty in obtaining clear-cut interpretations.

GENERALIZED ATTRIBUTES AND SOCIO-ECONOMIC CHARACTERISTICS

Once those attributes common to each population's cognitive structure are identified, we can investigate whether these latent dimensions are significantly related to a group's socioeconomic characteristics. If a systematic relationship is discovered, it can provide useful information for the segmentation of a travel market into homogeneous subpopulations. Following the procedure outlined, each individual was assigned a satisfaction rating on each of the generalized attributes such that if his or her rating exceeds three on a given attribute, the individual's shopping activity is assumed to be satisfactory in this regard; if less than or equal to three, the shopping activity is considered to be unsatisfactory for the given attribute. Once determined, the satisfaction ratings were cross-tabulated with various socioeconomic characteristics of the population, including gross annual family income, mode of travel to the shopping area, length of residence at current address, education, race, sex, age, and marital status. For the analysis, these variables are stratified as follows:

Variable	Name	Stratification
Gross annual family income	INCOME	<\$10 000 ≥\$10 000 but <\$20 000 ≥\$20 000
Mode of travel to shopping area	MODE	Automobile only Some form of public transit All others
Length of residence at current address	RESIDE	<2 years ≥2 years but <10 years ≥10 years
Education	EDUC	No college Some college
Ethnicity	RACE	White Nonwhite
Gender	SEX	Male Female
Age	AGE	≤30 years of age >30 years of age
Marital status	MARSTA	Married Not married

Table 6. Gamma measures of association, downtown San Francisco and local CBD trips.

Socioeconomic Characteristic	TCONV	TCOMF	TSAFTY	SATT	CONGEST
Downtown San Francisco Trips					
Central city population					
MODE	0.313	-0.476		0.181	
AGE	0.285	0.324	0.328	0.345	
RESIDE	0.257			0.305	
Suburban population					
MODE	0.339	-0.448			
EDUC	-0.389			-0.448	
RACE		0.448			
AGE	0.322	0.424	0.375		0.754
MARSTA					0.502
RESIDE	0.141	0.351	0.221		
Local CBD Trips					
Central city population					
MODE		-0.450			-0.339
EDUC	0.513		0.229		-0.261
RACE				-0.242	
AGE				0.279	0.242
RESIDE				0.192	0.270
Suburban population					
MODE	-0.351	-0.621			
RACE		0.557			
SEX				-0.289	
RESIDE			0.210		

As indicated, gamma is a measure of association that relates the order of one variable to that of another, where its sign is determined by the number of concordant (relative to discordant) pairs. Thus, for example, given our definitions of satisfaction ratings and age categories, a positive gamma measure of 0.285 between AGE and TCONV (Table 6) indicates that the number of same-ordered pairs (low age category—low satisfaction rating; high age category—high satisfaction rating) exceeds the number of reverse ordered pairs (low age category—high satisfaction rating; high age category—low satisfaction rating). For interpretive convenience, we can simply say that younger vis-à-vis older individuals are less satisfied by the convenience of a shopping trip. Moreover, the measure indicates that our prediction errors can be reduced by 28.5 percent, given the knowledge of the group's age stratification and that it is positively related to trip convenience satisfaction. Similar interpretations apply to all gamma measures reported.

Table 6 summarizes the results and reports the gamma measure of association for those relationships that are significant at the 0.05 level. In general, the information in these tables lends further support to the hypothesis that a population's cognitive structure of shopping behavior is related to its socioeconomic characteristics.

Travel mode was generally found to be significant relative to those attributes concerned with the trip to a shopping area. Thus, in the central city and suburban areas and for each trip destination, individuals who use automobiles in their shopping journeys are more satisfied with the comfort of the trip than are individuals who undertake this journey by some form of public transportation (or by some other mode of travel such as walking and bicycle).

Also note that, relative to the relationship between mode and travel convenience satisfaction, the sign of gamma is not constant. For central city and suburban populations that shop in downtown San Francisco, automobile vis-à-vis public-transit users perceive their trips to be less convenient. In their local CBD shopping activity, however, automobile users in the suburban population perceive their trips to be more convenient. This may reflect the greater degree of automobile travel in local CBD trips whereas excursions to downtown San Francisco characterize more extensive use of BART and other forms of public transit.

An individual's length of residence at his or her current address (RESIDE) is also significantly related to the cognitive structure and, in general, the shorter the duration at his or her current residence, the more dissatisfied he or she is with various aspects of the shopping excursion. Relative to downtown San Francisco shopping trips, RESIDE is significantly related to trip convenience and the attraction of a shopping area for central city residents; for suburban residents, on the other hand, it is related to all travel components of the activity, including TCONV, TCOMF, and TSAFTY. For shopping journeys to local CBDs, RESIDE is significantly associated with a central city resident's perceived satisfaction of shopping area attraction and congestion attributes; for a suburban resident, it is significantly related to trip safety.

Except for local CBD trips by suburban residents, age distribution in the population is an important influence. For shopping journeys to downtown San Francisco, the younger population (less than or equal to 30 years of age) are less satisfied with the convenience, comfort, and safety features of the trip. In addition, in the central city and suburban areas, these same individuals are less satisfied on the shopping area attraction and congestion aspects, respectively, of their shopping experience. For shopping trips to local CBDs (as observed in Table 6) a significant relationship does not exist between AGE and generalized travel attributes of the trip, although for the central city areas, the younger population is less satisfied with shopping congestion and attraction components of the activity.

In two instances, formal education is significantly related to cognitive structures. For suburban residents that undertake shopping activities in downtown San Francisco, those who have no college experience were more satisfied with the convenience aspects of their trips as well as with shopping area attraction. This is not the case, however, for central city residents making local CBD trips. In this case, those who have no college experience are less satisfied with the convenience and safety aspects of their shopping journeys. The same individuals, however, are more satisfied with local CBD shopping congestion.

Finally, RACE, SEX, and MARSTA are observed to be significantly related, in isolated cases, to satisfaction ratings of alternative attributes. For the suburban and central city populations who make trips to downtown San Francisco and the local CBD, respectively, nonwhites are more satisfied with the comfort aspects of the trip; for the central city populace that undertake shopping trips to downtown San Francisco, nonwhites are less satisfied with shopping congestion. Last, for the suburban population, males are more satisfied with the attraction of shopping in their local CBDs and, for downtown San Francisco trips, those not married are more satisfied with shopping congestion.

SUMMARY AND CONCLUSIONS

In this study, attitudinal data obtained from a 1973-1974 BART impact travel survey was factor analyzed to determine the cognitive structure of central city and suburban populations for shopping trips to downtown San Francisco and local CBDs. In general, the results indicate that, regardless of residence or trip destination, important constructs in a shopper's attitude structure are (a) generalized trip convenience, (b) generalized trip comfort, (c) generalized trip safety, (d) generalized shopping area attraction, and (e) generalized shopping congestion. This does not imply that the populations are perceptually homogeneous, however, since the importance of these attributes was not uniform across populations. Indeed, each population's cognitive structure was further differentiated by the presence of other factors unique to that population.

When the common attributes of each population's cognitive structure were related to its socioeconomic characteristics, mode of travel, length of residence, and age were important determinants of an individual's satisfaction rating. In general, travel mode was related significantly to trip comfort and trip convenience attributes of the shopping activity. Regardless of destination, automobiles were more satisfactory in providing a comfortable trip and, for shopping in local CBDs, more convenient. Automobile travel was less convenient, however, for making a shopping journey to downtown San Francisco.

In addition, length of residence was related significantly to an individual's attitude structure. Although no pattern was evident to demonstrate a relation between RESIDE and particular generalized attributes or specific populations, RESIDE was positively associated with an individual's satisfaction rating in each case so that, the longer one maintains a residence in a given locale, the more satisfied one is with the underlying dimensions of shopping activity. This may reflect the fact that, over time, an individual becomes more familiar with the transportation infrastructure, including routes, travel modes, and points of access, as well as with the general region, such that one is more knowledgeable regarding the advantages and disadvantages of alternative means of travel and alternative trip destinations. This effect will tend to lessen the dissatisfaction associated with various aspects of the shopping activity.

The results also characterize older individuals as more satisfied with all aspects of the shopping activity. However, depending on trip destination, the significant relationships vary. For downtown shopping excursions, trip and shopping area characteristics were significantly related to age, whereas, for local CBD trips, only shopping area attributes were significant.

Education, race, sex, and marital status were also significant in various circumstances, but no general pattern was evident. Finally, conspicuous in its absence was any significant relationship between income and satisfaction ratings.

From the analysis, four hypotheses for future research are suggested.

1. The generalized attributes listed are important dimensions for all shopping excursions and should be incorporated into travel-demand models if shopping trip behavior is to be forecast accurately. Moreover, care must be taken to develop separate models for each group that exhibits similar cognitive structures if biased predictions are to be avoided.
2. The distance from home to a shopping area critically affects individual perceptions of alternative shopping trips. Individuals who reside further from their shopping destination, for example, will undertake a multipurpose trip and, accordingly, emphasize in their attitude structure the ability to make stops along the way.
3. The more travel required to reach a shopping destination, the more will specific trip components be stressed. Conversely, if little travel is needed, individuals consider overall feasibility or desirability of the journey and do not highlight specific elements.
4. The results imply that the cognitive structures of a population segmented by residence and trip destination are unique, although they do include a common set of generalized attributes. Moreover, mode of travel, age distribution, and length of residence are important determinants of a population's attitude structure and act to further segment the travel market.

Continued research in this area is requisite to identify more completely the relation that exists between a population's cognitive structure and its socioeconomic characteristics and to determine, if a relation is defined, whether it is transferable among populations that have similar socioeconomic characteristics. Furthermore, the policy implications that emanate from these and other attitudinal modeling efforts need to be examined thoroughly.

REFERENCES

1. D. Brand. Travel Demand Forecasting: Some Foundations and a Review. In Urban Travel Demand Forecasting, TRB, Special Rept. 143, 1972, pp. 239-282.
2. F. X. DeDonnea. The Determinants of Transport Mode Choice in Dutch Cities. Rotterdam Univ. Press, Netherlands, 1971.
3. T. A. Domencich and D. McFadden. Urban Travel Demand: A Behavioral Analysis. North-Holland, Amsterdam, 1975.
4. R. G. McGillivray. Binary Choice of Urban Transportation Mode in the San Francisco Bay Region. Econometrica, Vol. 40, 1972, pp. 827-848.
5. R. M. Downs. The Cognitive Structure of an Urban Shopping Center. Environment and Behavior, Vol. 2, 1970, pp. 13-39.
6. T. F. Golob. Resource Paper on Attitudinal Models. In Urban Travel Demand Forecasting, TRB, Special Rept. 143, 1972, pp. 130-145.
7. T. F. Golob and R. Dobson. Assessment of Preferences and Perceptions Toward Attributes of Transportation Alternatives. In Behavioral Demand Modeling and Valuation of Travel Time, TRB, Special Rept. 149, 1974, pp. 58-81.
8. D. A. Hensher. Perception and Commuter Mode Choice—An Hypothesis. Urban Studies, Vol. 12, 1975, pp. 101-104.
9. G. C. Nicolaidis. Quantification of the Comfort Variable. Transportation Research, Vol. 9, 1975, pp. 55-66.
10. F. T. Paine and others. Consumer Attitudes Toward Auto Versus Public Transport Alternatives. Journal of Applied Psychology, Vol. 53, 1969, pp. 472-480.
11. P. R. Stopher. Development of Market Segments of Destination Choice. TRB, Transportation Research Record 649, 1977, pp. 14-22.
12. K. G. Joreskog and M. Van Thillo. New Rapid Algorithms for Factor Analysis by Unweighted Least Squares, Generalized Least Squares, and Maximum Likelihood. Educational Testing Services, Research Bull. 71-5.
13. C. D. Dziuban and E. C. Shirkey. When Is a Correlation Matrix Appropriate for Factor Analysis? Psychological Bull., Vol. 81, 1974, pp. 353-361.
14. H. F. Kaiser and J. Rice. Little Jiffy, Mark IV. Educational and Psychological Measurement, Vol. 34, 1974, pp. 111-117.

15. D. F. Morrison. Multivariate Statistical Analysis. McGraw-Hill, New York, 1967.
16. P. E. Green and V. R. Rao. Rating Scales and Information Recovery: How Many Scales and Response Categories to Use? Journal of Marketing, Vol. 34, 1970, pp. 33-39.
17. W. M. Leonard II. Basic Social Statistics. West Publishing Co., St. Paul, 1976.
18. H. L. Costner. Criteria for Measures of Association. American Sociological Review, Vol. 30, 1965, pp. 341-353.
19. W. L. Nicholls II. Sampling and Field Work Methods of the 1973-1974 BART Impact Travel Survey. Survey Research Center, Univ. of California, Berkeley, Document WP-1-3-74, 1974.
20. R. L. Gorsuch. Factor Analysis. W. B. Saunders Co., Philadelphia, 1974.

Publication of this paper sponsored by Committee on Passenger Travel Demand Forecasting.