HIGHWAY RESEARCH RECORD

Number 392

Transportation Demand and Analysis Techniques

18 reports
prepared for the
51st Annual Meeting

Subject Areas

55	Traffic Measurements
84	Urban Transportation Systems

HIGHWAY RESEARCH BOARD
DIVISION OF ENGINEERING NATIONAL RESEARCH COUNCIL
NATIONAL ACADEMY OF SCIENCES—NATIONAL ACADEMY OF ENGINEERING

Washington, D.C. 1972

NOTICE

The studies reported herein were not undertaken under the aegis of the National Academy of Sciences or the National Research Council. The papers report research work of the authors done at the institution named by the authors. The papers were offered to the Highway Research Board of the National Research Council for publication and are published herein in the interest of the dissemination of information from research, one of the major functions of the HRB.

Before publication, each paper was reviewed by members of the HRB committee named as its sponsor and was accepted as objective, useful, and suitable for publication by NRC. The members of the committee were selected for their individual scholarly competence and judgment, with due consideration for the balance and breadth of disciplines. Responsibility for the publication of these reports rests with the sponsoring committee; however, the opinions and conclusions expressed in the reports are those of the individual authors and not necessarily those of the sponsoring committee, the HRB, or the NRC.

Although these reports are not submitted for approval to the Academy membership or to the Council of the Academy, each report is reviewed and processed according to procedures established and monitored by the Academy's Report Review Committee.

ISBN 0-309-02064-6

Price: $4.60

Available from

Highway Research Board
National Academy of Sciences
2101 Constitution Avenue, N.W.
Washington, D.C. 20418

CONTENTS

FOREWORD. v

MODAL-CHOICE MODEL FOR RELATING DEMAND TO INVESTMENT
Salvatore J. Bellomo, Christopher G. Turner, and Denis K. Johnston . . . 1

CONSISTENCY IN TRANSPORTATION DEMAND AND EVALUATION MODELS
Dan G. Haney. 13
Discussion
Ezra Hauer 24
Author's Closure 25

FORECASTING TRAVEL DEMAND FOR NEW URBAN TRANSPORTATION SYSTEMS
Anthony R. Tomazinis. 26

TRAFFIC ASSIGNMENT WITHOUT BIAS
Alvin H. Benesh. 36

SEGMENTED, MULTIMODAL, INTERCITY PASSENGER DEMAND MODEL
John W. Billheimer. 47

EFFECT OF ZONE SIZE ON TRAFFIC ASSIGNMENT AND TRIP DISTRIBUTION
Bruno R. Wildermuth, D. J. Delaney, and K. E. Thompson 58

DEVELOPMENT AND APPLICATION OF DIRECT TRAFFIC ESTIMATION METHOD
Yehuda Gur 76

TOWARD AN EFFICIENT HIGHWAY SERVICE ESTIMATOR FOR CONGESTED NETWORKS
Hubert Le Menestrel and Bruno R. Wildermuth 89

ACTIVITY-ACCESSIBILITY MODELS OF TRIP GENERATION
T. Z. Nakkash and W. L. Grecco 98

COMPARISON OF PROBABILISTIC MODAL-CHOICE MODELS: ESTIMATION METHODS AND SYSTEM INPUTS
Antti Talvitie. 111

APPROACH TO PROBABILITY DISTRIBUTION OF VALUE OF WALKING TIME AND PEDESTRIAN CIRCULATION MODELS
Jean Francois Allouche 121

TRAVEL TO OUTDOOR RECREATION AREAS IN KENTUCKY (Abridgment)
John A. Deacon, Jerry G. Pigman, and Robert C. Deen 134

DEVELOPMENT AND TESTING OF A MULTIPATH-ASSIGNMENT TECHNIQUE (Abridgment)
P. M. Dalton and M. D. Harmelink 136

USE OF SOCIOECONOMIC INDICATORS IN TRIP ATTRACTION OF LARGE WORK CENTERS
George E. Mouchahoir and Paul H. Wright 139

ACCESS TO JOBS AND WILLINGNESS TO TRAVEL (Abridgment)
Charles B. Notess 143

STABILITY OF RECREATIONAL DEMAND MODEL
D. C. Robinson and W. L. Grecco. 147

RELIABILITY ANALYSIS IN LAND USE-TRANSPORTATION PLANNING (Abridgment)
Kumares C. Sinha 157

DISTRIBUTION PATTERNS OF AUTOMOBILE TRAVEL IN THE NEW YORK METROPOLITAN AREA (Abridgment)
J. David Jordan 159

SPONSORSHIP OF THIS RECORD. 161

FOREWORD

This RECORD will be of interest to those using conventional transportation forecasting techniques. The papers cover a wide variety of topics but emphasize improvements to procedures in the areas of trip-generation, modal-choice, and traffic-assignment techniques. Forecasting for recreational travel and pedestrian flow and for new urban transportation systems is also covered.

It is important to note that attention is again being paid to an important and somewhat neglected area: assigning travel to a network of facilities. Regardless of the methods used to obtain travel demand forecasts, the ultimate use of such forecasts is to assist responsible agencies in facility planning, evaluation, and design. The assignment of demand to a facility is the last and perhaps the most important step in this process. Multipath, multivariable traces and incremental loading techniques are discussed. In addition, direct assignment methods are evaluated.

The importance of including the effects of the transportation system in estimates of trip-making potential is covered in several papers. One researcher concludes that provision of a high level of highway service does not increase the use of automobiles, but he points out that there is a lower use of automobiles in areas of high density and good transit service. Others emphasize the importance of including accessibility variables in estimating person-trip generation. Both findings, then, indicate the importance of including transit system variables in trip-generation analysis. These findings point the way to significant improvements that can be made in present travel-forecasting models and procedures.

—George V. Wickstrom

MODAL-CHOICE MODEL FOR RELATING DEMAND TO INVESTMENT

Salvatore J. Bellomo, Christopher G. Turner, and Denis K. Johnston,
Alan M. Voorhees and Associates, Inc.

The development of a macromodel for modal choice is presented. The model relates investment to transit supply, supply to level of service, and level of service to demand. The model is developed at a city-wide level by using a simulation technique involving the analysis of 13 individual highway-transit system tests for a city of 3.5 million people. Land use activities are allocated on the basis of accessibility provided by both highway and transit systems. The generation of travel is sensitive to the level of service provided, and the distribution of trips is achieved by using weighted highway-transit skim trees and a standard gravity model. The modal-choice analysis employs a utilitarian model develped as part of a study of models calibrated for existing cities. The macromodel relates demand to mean travel-time difference and mean travel-cost difference between transit and highway for work and nonwork purposes for both peak and off-peak periods. The mean city-wide transit travel time is related to the supply of seat-miles of service per capita, and the highway travel time is related to capacity-miles of highway per capita. Transit supply is in turn related to the capital and operating costs of providing that service. In application, the model assumes a fixed level of highway supply and has as policy variables the absolute investment level in transit, the split of investment between bus or rail rapid transit and conventional bus transit, the transit fare, the split of service between peak and off-peak periods, and the parking cost.

•THERE IS a critical need in the urban transportation planning process for an analytical capability to permit the testing of multimodal and multiregional transportation-investment policies. In recognition of this need, the U.S Department of Transportation sponsored a series of studies aimed at improving that capability. One of these studies was the Transportation Resource Allocation Study that considered the quantity and the quality of the transportation system as well as the basic indicators of socioeconomic characteristics such as population and automobile ownership. In an earlier report, Kassoff and Gendell (1) described a process and travel-demand forecasting procedure to project future urbanized-area travel. This work has been further expanded in England by Lesley (2), who has attempted to establish a relation between the structure and operating performance of urban public transportation systems and the macroparameters of European cities.

This paper is concerned with the development of a modal-choice model that can be used to relate demand to investment on an urbanized-area basis. The model provides the planner with a simple analytical tool to estimate transit investment for a given demand, estimate transit demand for a given investment, or estimate the level of transit service for a given demand or investment.

In the development of this macromodel, a simulation approach was used. That is, the basic relations have been established by simulating the urban-activity characteristics for 2 large metropolitan areas, and an experimental design was developed to sketch feasible combinations of highway and transit levels of service for both areas based on the known theory of travel demand and urban activity. The reasons for adopting a simulation approach are as follows:

1. Supplying an adequate and consistent empirical data base across several urbanized areas is difficult;
2. Certain variables of concern (e.g., high levels of transit service) do not currently exist on a city-wide basis;
3. Sensitivities that emerge on a macrobasis can only be established through a microsimulation and then a city-wide summary; and
4. Interrelations among land use, transit, highway, and other policies for a metropolitan area make sensitivity testing very difficult.

EXPERIMENTAL DESIGN

The experimental design developed for the study is shown in Figure 1. The number of transportation alternatives (simulation tests) was set at 13, and the figure shows how these alternatives were developed to cover a wide range of highway and transit system service levels.

In each case, the design of the system was based on the characteristics of systems operating in, or proposed for, existing cities. A mix of transit systems is implied by the categories of conventional bus transit, bus rapid transit, and rail rapid transit because the operation of each system is dependent on supporting feeder and distribution systems.

Two base-year urban-activity patterns were developed in the study to simulate a range of existing cities. Specifically, patterns of central-city activity concentration and suburban-activity dispersion were tested. From an analysis of existing cities, each transportation system was matched with the urban-activity pattern to which it was most closely related, and as a result the 13 tests shown in Figure 1 were selected.

OPERATIONAL PROCEDURE

To provide inputs to the macromodel, an operational procedure was developed and applied to each of the 13 simulation tests shown in Figure 1. This procedure is shown in Figure 2, and basically it has involved the following major components.

1. The development of a procedure enabling the prediction of microutility modal-choice curve characteristics (3, 4) for a city, given a set of macroparameters describ- that city, involved a research study of the characteristics of existing microutility curves in order to relate those characteristics to the macroparameters of the city for which each model was developed. (For the aggregate transit percentages output from the study to be realistic, a modal-choice model that accurately reproduced the tripmaker's modal-choice decisions had to be utilized for each simulation test. Such models are usually developed and calibrated for individual cities on the basis of trip-interchange data. Because such data are not available for the simulated cities, it was necessary to develop a relation between the characteristics of microutility modal-choice models and aggregate city parameters.) A relation was developed between the microutility modal-choice curves calibrated for Los Angeles, Twin Cities, and Newcastle (England) and the mean city-wide automobile ownership stratified by the level of automobile ownership (5). The resulting family of curves provides the basis on which the microutility curve for different cities can be estimated directly from a knowledge of the mean automobile-ownership rate in the city.

The model is essentially a 3-dimensional surface, as shown in Figure 3, for each level of automobile ownership. However, because only mean automobile-ownership rates were predicted for the simulated cities, the model was aggregated over levels of automobile ownership for use in the study by using a unique relation between mean automobile ownership and the percentage of households in each ownership group.

2. The development of realistic urban-activity patterns and transportation system alternatives to use as input to the simulation study was achieved from an extensive analysis of existing cities and resulted in the development of 2 base-year distributions of population and employment representing a highly concentrated city typical of older cities and a more dispersed automobile-oriented city. Those cities were matched against the transportation systems as shown in Figure 1. Base-year distributions of

igure 1. Experimental design.

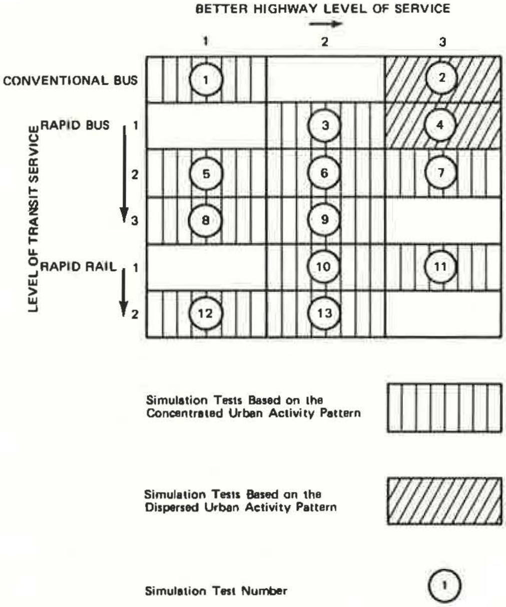


Figure 2. Operational procedure for development of variables for macromodel for modal choice.

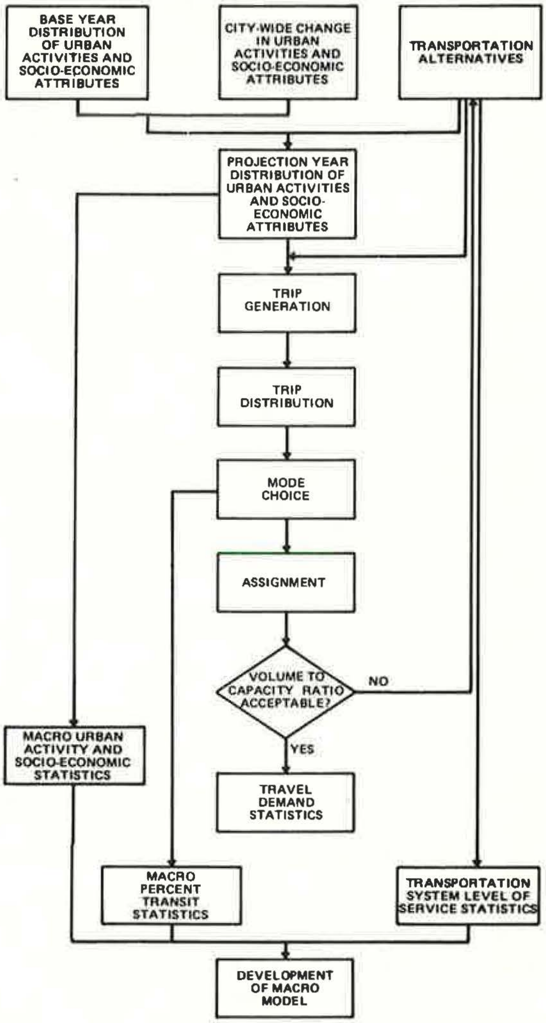


Figure 3. Three-dimensional surface of micromodel for modal choice.

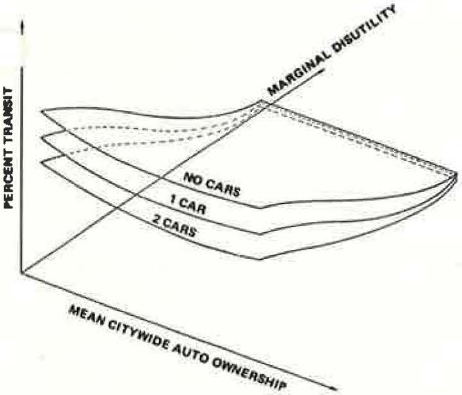


Table 1. City-wide variables generated by each simulation test for input to macromodel.

	System Characteristics		
Component	Highway	Transit	Urban Activity Characteristics
Demand[a]	Number of trips Vehicle-miles	Percentage of transit work trips Percentage of transit nonwork trips Percentage of transit total trips Passenger-miles	Population and population density (CBD, central city, and suburbs)
Level of service[a]	Mean run time Mean travel cost (running costs and parking costs) Mean speed Mean trip length	Mean walk time Mean wait time Mean run time Mean fare Mean speed Mean trip length	Manufacturing employment and manufacturing employment density (CBD, central city, and suburbs) Office employment and office employment density (CBD, central city, and suburbs)
Supply	Arterial miles Freeway miles Highway miles Arterial lane-miles Freeway lane-miles Highway lane-miles Capacity-miles	Fleet size[a] (number of buses and number of rail cars) Miles of private right-of-way Route-miles Number of rapid transit stations Seat-miles[a]	Population following employment and population following employment density (CBD, central city, suburbs) Mean income Mean automobile ownership
estment		Capital costs (guideway and rolling stock) Daily operating costs (labor and other)	

[a] Stratified by peak and off-peak period.

mean zonal income, automobile ownership, parking costs, and transit fares were established for both cities. [The distibution of mean zonal income was calculated from 1968 in-house data for Cleveland (6). Automobile-ownership rates were calculated from a relation among income, transit level of service, and automobile ownership developed for Memphis, Tennessee. Base-year zonal parking costs were developed from a procedure used in the Washington Metropolitan Area Transit Authority study (7), and zonal transit fares were developed from in-house data and library research material.]

3. The application of urban-activity, trip-generation, trip-distribution, modal-choice, and trip-assignment models to each simulation test developed inputs to the macromodel. Initially, distributions of population, employment by type, and income were allocated to the simulated city zones to produce the base-year patterns of concentrated and dispersed urban activity (which in each case had a base-year population of 2.5 million and a labor force of 900,000). At the same time, the 13 simulation tests shown in Figure 1 were devised, and the respective highway and transit systems were developed. Peak and off-peak highway and transit skim trees were then built for each test, and a composite skim tree was developed and weighted in terms of work and nonwork trips, peak and off-peak trips, and highway and transit trips. The projected city-wide increases in population and employment by type were then distributed to zones on this composite skim tree by using a simple activity-allocation model (8). [A projection period of 20 years from a 1970 base year was assumed in the study. Both simulated cities were assumed to have the same growth in population and employment, and growth was based on the mean projections for the 40 largest urbanized areas in the United States during the period 1970-1990 (8).]

It was also necessary to project the zonal increase in income and automobile ownership. In the case of income, however, a literature survey indicated that little information was available on changes in the spatial distribution of income over time. A simple model was, therefore, developed to allocate the total city-wide increase in income to zones. The mean city-wide increase in income was based on the Cleveland study projections (8), and this total was allocated to zones in proportion to each zone's population growth during the projection period relative to the city-wide population growth and to each zone's base-year mean income relative to the city-wide base-year mean income.

Automobile-ownership rates for the projection were developed by applying the 1990 zonal income projections to the base-year relation established among income, transit level of service, and automobile ownership.

The 1990 distributions of activity and socioeconomic characteristics were then used to forecast trip-generation rates [which were sensitive to the level of service provided (9)] for home-based work, home-based nonwork, and non-home-based trips. So that they would be consistent with the 1990 projected distributions of activity, socioeconomic characteristics, parking costs, and transit fares, the composite skim trees were updated. The 1990 trip productions and attractions, composite skim trees, and F-factor curves (10) were used, and the gravity model was run for each of the following trip purposes associated with each test: home-based work, home-based nonwork, and non-home-based.

Trip tables were then combined into the categories of work and nonwork trips and factored to provide peak and off-peak trip tables. The resulting 4 tables were split into automobile and transit trip tables by using the micromodel described previously and shown in Figure 3, and the trips were loaded onto the highway and transit networks.

The reasonableness of the assumptions made about the volume to capacity relations was tested by using a standard equilibrium approach to check the output highway speeds against the initially assumed highway speeds using capacity restraint. If the output speeds were significantly at variance with the initially assumed speeds, the output speeds were adopted, and the entire procedure was reiterated for that particular simulation test. If the initially assumed speeds were within an acceptable range of the output speeds, the highway assignments were adopted.

Transit trips were then assigned to the transit network; for those tests having more than 1 transit mode, the assignment to a specific mode was made on the basis of relative modal capacity.

4. The parameters describing each simulated city test were spatially aggregated, and the relations between those parameters and the city-wide percentage of transit usage were developed after the simulation tests were complete. Table 1 gives the parameters that were generated and considered as possible variables for the macromodel.

THE MACROMODEL FOR MODAL CHOICE

The variables from the simulation tests given in Table 1 were used to develop city-wide relations of transit supply and investment, level of service and supply, and demand and level of service. Alternative combinations of variables were tried for each of the 3 components of the macromodel in an attempt to develop a model that was both conceptually sound and statistically accurate. The final relations established for each of the components of the model are now presented.

Supply and Level of Investment

Figures 4 and 5 show the relations that were developed between transit supply and level of investment. Transit supply (measured in terms of rapid transit seat-miles) is highly correlated with both capital costs and operating costs on a per capita basis. The number of seats per vehicle was based on the characteristics of existing transit systems (11), and it was estimated that each bus had 50 seats, each rail car had 80 seats, and each train had 6 rail cars in the peak period and 3 in the off-peak period.

For each system test, capital costs were developed for both rolling stock and guideways by using 1968 Washington Metropolitan Area Transit Authority unit cost figures (12). Guideway costs were made up of the type of line construction (e.g., subway, arterial, or grade construction for rail systems) and of station costs, which included the cost of providing parking facilities. Rolling-stock costs were calculated per vehicle for buses and rail cars, and operating costs were calculated (from the same source as capital costs) for conventional buses, rapid transit buses, feeder buses, and rapid transit rail cars on a daily basis.

Level of Service and Supply

Because the demand component of the macromodel was developed to deal with both peak- and off-peak-period travel, it was necessary to develop peak and off-peak relations between the level of service and supply variables. (Peak and off-peak level-of-service measures were then input into the demand component of the model.)

Figure 6 shows the peak-period model developed to relate transit-travel time to rapid transit seat-miles per capita. As shown in the figure, there are distinct relations for bus-oriented systems and rail-oriented systems. The model indicates that the level of service provided by rapid transit bus-oriented systems is very sensitive to changes in the level of supply over a comparatively narrow range of supply. In contrast, the model shows that the changes that can be brought about in the level of service for rail-oriented systems by varying the level of supply are of a small magnitude, although the supply of rapid transit seat-miles is greater and covers a wider range. Figures 4 and 5 show that both capital and operating costs of rail-oriented systems tested are higher than those of bus-oriented systems. The implication is, therefore, that, for a given increase in investment, the bus-oriented systems are more cost effective in terms of providing an improved peak-period level of service than are rail-oriented systems. (This conclusion relates to the experiment described in this paper. It is not meant as a general conclusion for all metropolitan areas.)

For the off-peak period, little correlation was found to exist between off-peak level of service and supply variables in absolute terms. However, when a ratio relation was developed, a strong correlation was found to exist between the off-peak to peak ratio of transit-travel time and the off-peak to peak ratio of total system seat-miles per hour. This model, stratified by highway level of service (expressed in terms of capacity-miles of highway per capita), is shown in Figure 7. The stratification by highway level of service is significant. It implies that different levels of off-peak service can be provided for a given level of off-peak supply, depending on the quality of the highway system.

Figure 4. Rapid transit seat-miles per capita versus capital costs per capita.

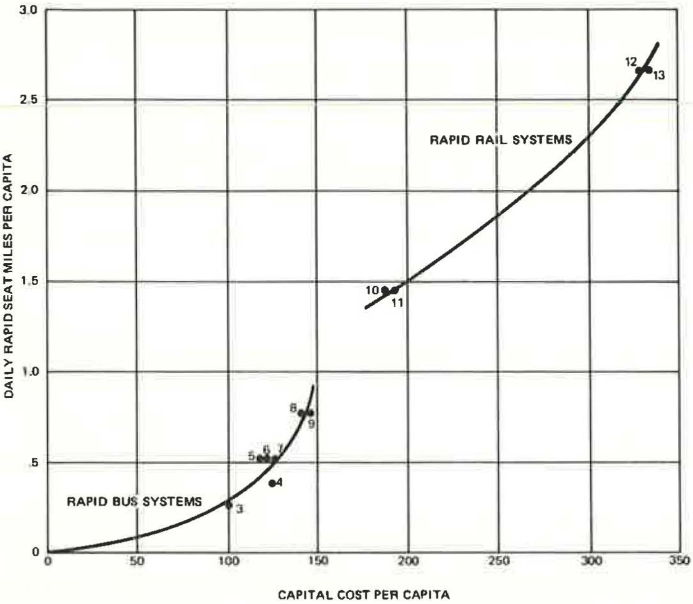


Figure 5. Rapid transit seat-miles per capita versus daily operating costs per capita.

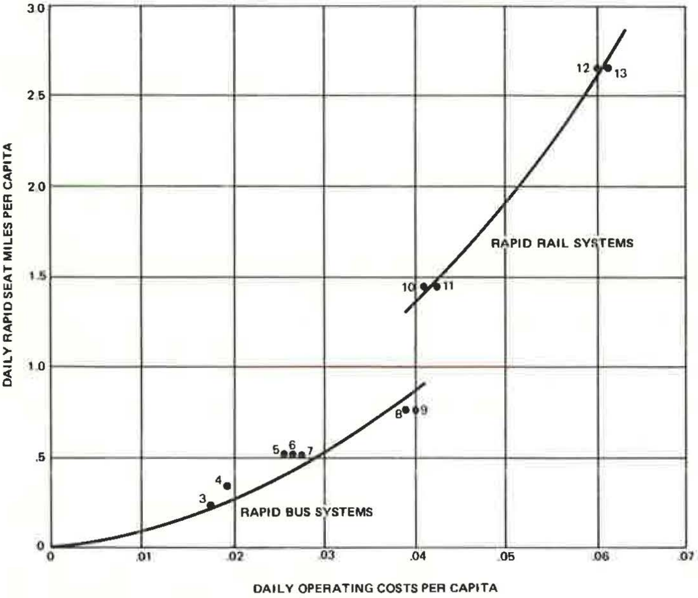


Figure 6. Peak-period transit travel time versus peak-period rapid transit seat-miles per capita.

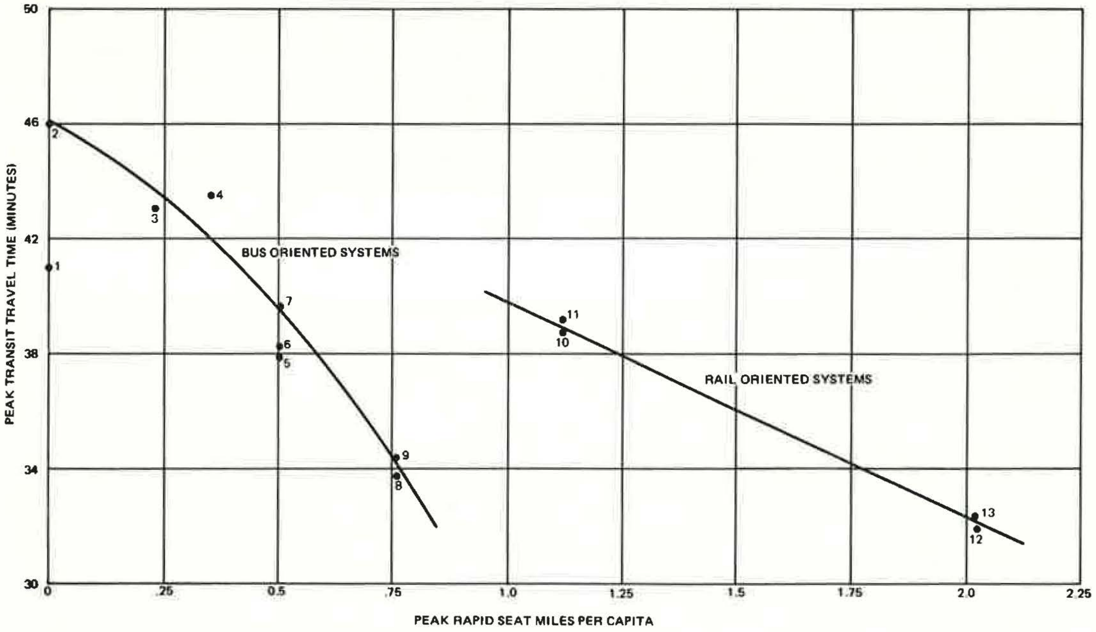
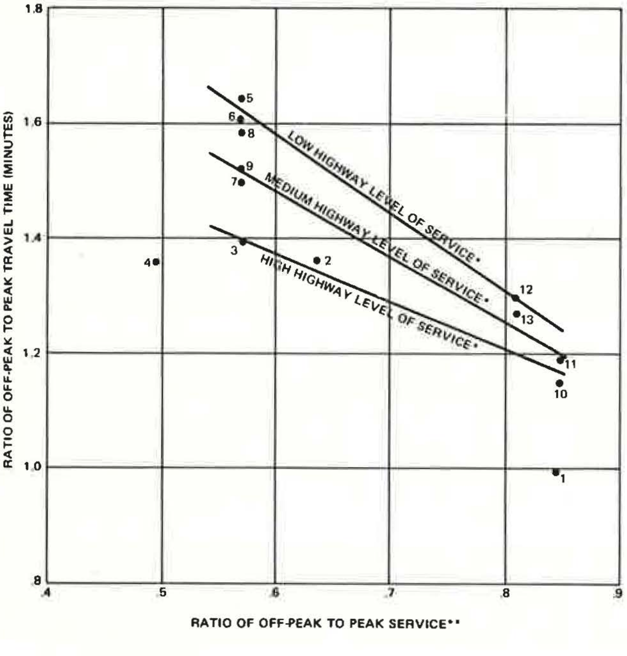


Figure 7. Ratio of off-peak to peak-period travel time versus ratio of off-peak to peak-period seat-miles per hour.



* Highway Capacity Miles Per Capita

LOW LEVEL OF SERVICE .415 CMS/PERSON
MEDIUM LEVEL OF SERVICE .622 CMS/PERSON
HIGH LEVEL OF SERVICE .864 CMS/PERSON

Capacity miles = $\sum_i$ (capacity of each link (i) x length of link in miles)

**Service measured in Seat Miles Per Hour

Perhaps more important, it indicates that, to provide the same level of off-peak service, cities with poorer highway systems require a higher level of supply than cities with good highway systems. Third, the relation also makes explicit the effect of different levels of highway investment (and hence supply) on the existing level of off-peak transit service.

Inasmuch as the demand component of the macromodel was based on a level-of-service difference variable (transit minus highway) for both peak and off-peak periods, relations were developed to provide the highway level of service inputs to the demand model. These relations are shown in Figures 8 and 9 for the peak and off-peak periods respectively. In each case, highway-travel time was found to be highly correlated with the supply of capacity-miles of highway per capita.

Demand and Level of Service

Of the alternative combinations of macroparameters that were used in the development of a relation between transit demand and level of service, it was found that travel-time and travel-cost differences (transit minus highway) best accounted for the variations in transit demand. Figures 10, 11, 12, and 13 show the 4 demand models developed to estimate transit usage for work and nonwork purposes in the peak and off-peak periods. The initial relations were developed at a 10-cent cost difference, and the sensitivity of demand was later tested against the range of cost differences shown in these 4 figures.

The relations developed here between demand and level of service are significant for several reasons. For a given travel-cost difference, the demand models make it possible to identify the effects that changes in transit-travel time (brought about by specifying different levels of investment and mixes of transit supply) have on demand. These effects can be stratified by time of day (peak and off-peak) and by trip purpose (work and nonwork). Similarly, the effects that changes in transit-travel cost have on demand can also be identified for a given travel-time difference. This is achieved by varying the transit-fare policy, and again the effects would be stratified by time of day and type of trip.

Alternatively, the demand models could indicate the effects that different levels of highway supply (measured in terms of changes in highway travel time) and highway cost (measured in terms of changes in parking cost policies) have on demand.

IMPLICATIONS

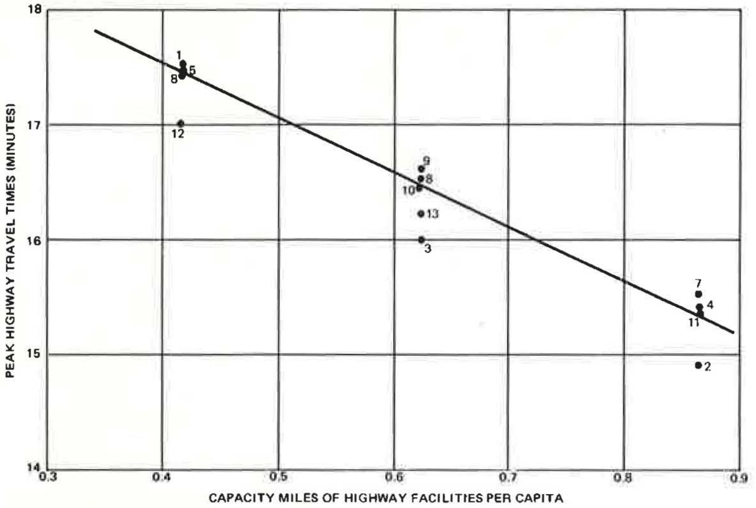
The macromodel for modal choice presented in this paper would have considerable application in multimodal and multiregional transportation-investment planning. It can be utilized to test alternative transit-investment policies across cities. For a given level of investment, a measure of supply in terms of rapid transit seat-miles per capita can be used to determine the level of transit service for that level of investment. From a prespecified highway supply, the highway level of service can be determined. Knowing both the highway and transit levels of service, one can estimate the percentage of transit trips.

The model can also be applied incrementally to predict the likely increase in transit patronage that would result from alternative investment policies for existing systems or to estimate the usage that would result in investing in new systems. It, therefore, provides a capability and analytical link that previously required a considerable elapsed time and computer cost to obtain. It is not a panacea, however, and should be considered as an initial rather than a final screening process when a wide range of transit-investment policies is examined.

The components of the model offer flexibility to the analyst. For a constant investment, for example, differing systems mixes of total, peak, and off-peak rail rapid and bus rapid transit supply can be specified, and the sensitivity of total transit demand can be ascertained. In addition, the model allows alternative fare and parking policies to be examined.

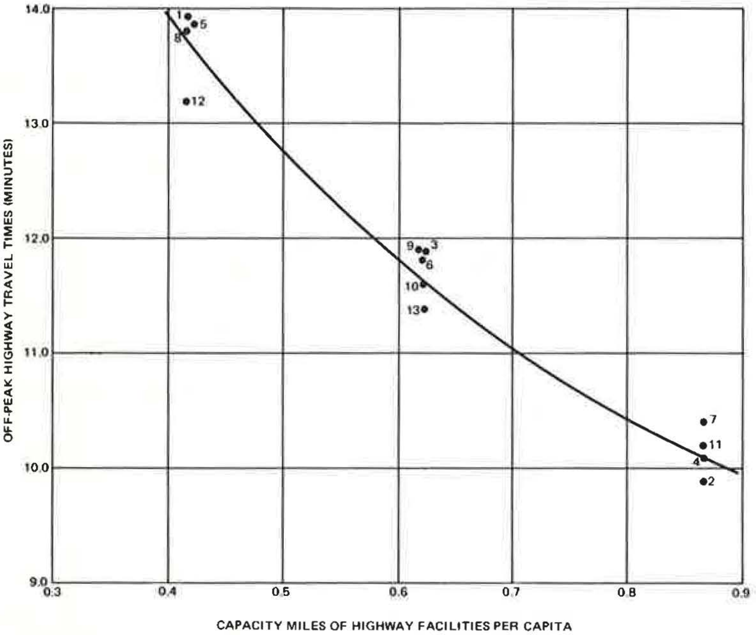
The user, however, should be careful that results are properly interpreted in analysis of transit-investment policies. The model is a guide for checking aggregate demand, supply, and level of service characteristics. It is not a substitute for microanalyses that are needed to determine the economic, social, and environmental effects of transit-investment policies for different groups of the regional community.

Figure 8. Peak-period highway travel time versus highway capacity-miles per capita.



*Capacity miles = $\sum_i$ (capacity of each link (i) x length of link in miles)

Figure 9. Off-peak-period highway travel time versus highway capacity-miles per capita.



*Capacity miles = $\sum_i$ (capacity of each link (i) x length of link in miles)

Figure 10. Peak-period work trips by transit (percent) versus peak-period travel time and cost difference.

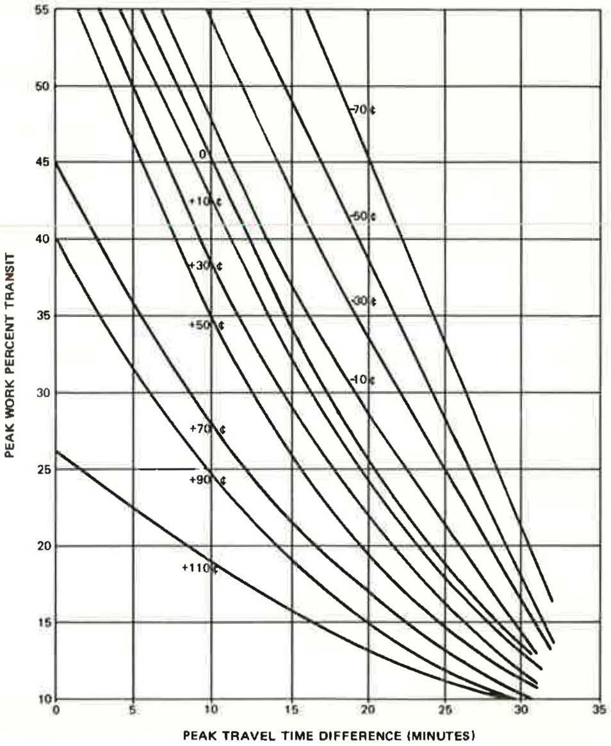


Figure 11. Peak-period nonwork trips by transit (percent) versus peak-period travel time and cost difference.

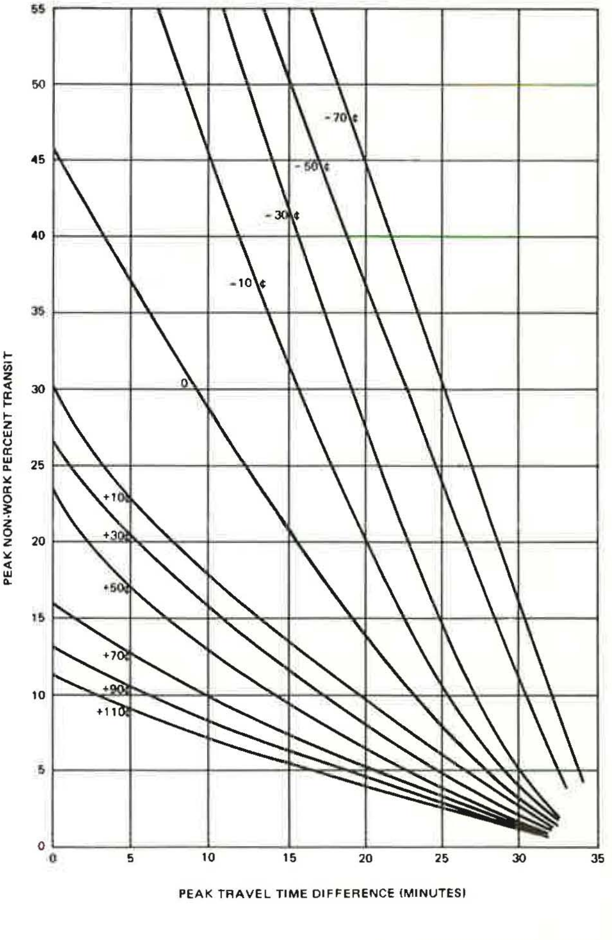


Figure 12. Off-peak-period work trips by transit (percent) versus off-peak-period travel time and cost difference.

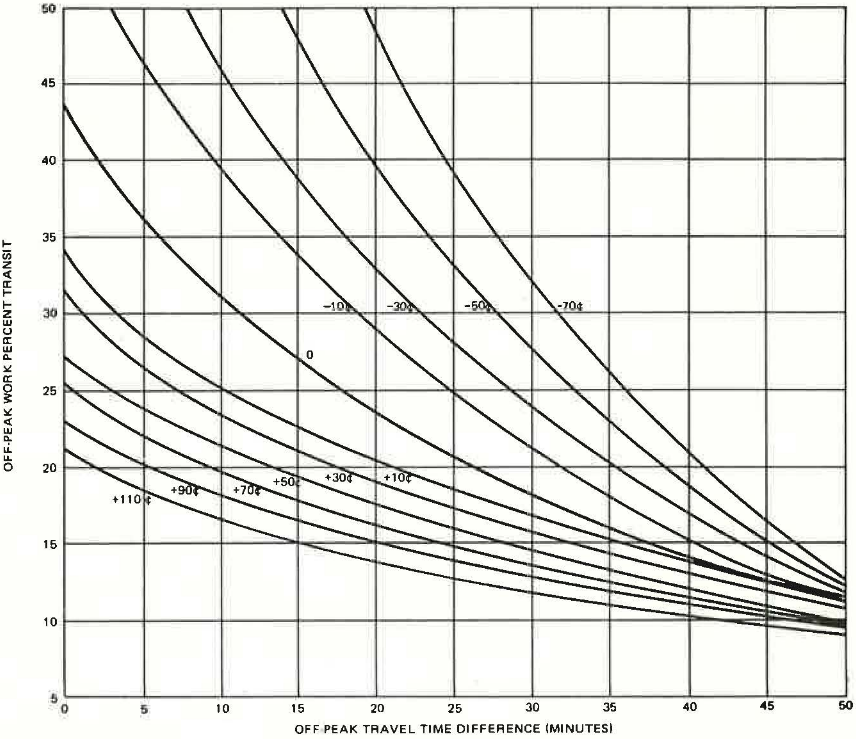
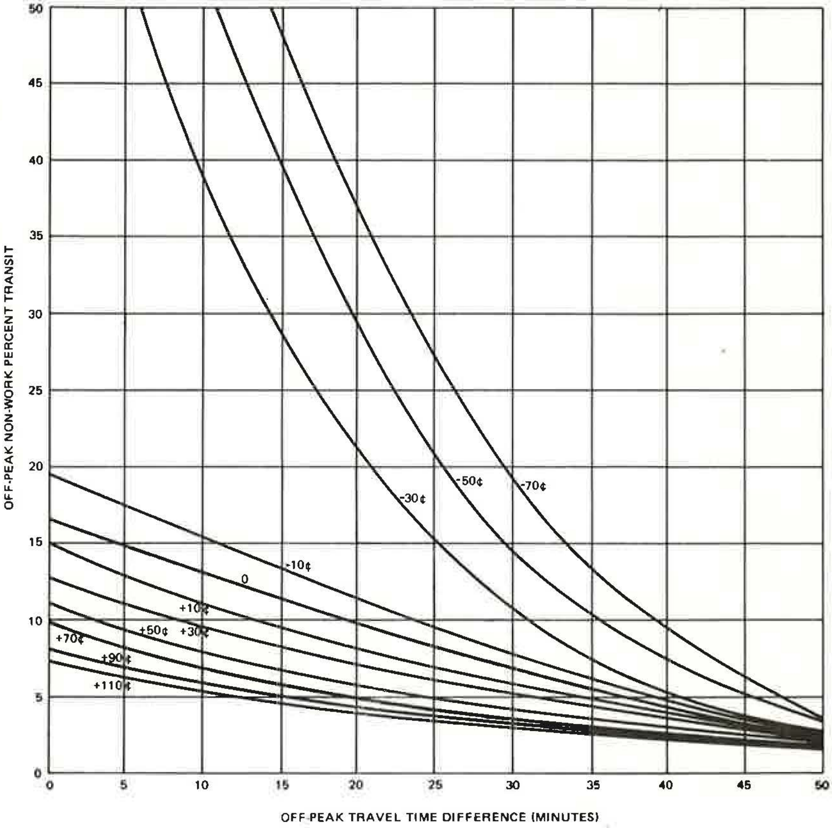


Figure 13. Off-peak-period non-work trips by transit (percent) versus off-peak-period travel time and cost difference.



REFERENCES

1. Kassoff, H., and Gendell, D. An Approach to Multiregional Urban Transportation Policy Planning. Highway Research Record 348, 1971, pp. 76-93.
2. Lesley, L. J. A Parametric Analysis of European Urban Public Transport. Traffic Engineering and Control, March 1971.
3. Shunk, G. A., and Bouchard, R. J. An Application of Marginal Utility to Travel Mode Choice. Highway Research Record 322, 1970, pp. 30-39.
4. Pratt, R. H. A Utilitarian Theory of Travel Mode Choice. Highway Research Record 322, 1970, pp. 40-53.
5. A System Sensitive Approach to Area-Wide Urban Travel Forecasting. Alan M. Voorhees and Associates, Inc., Tech. Memo. 4, May 1971.
6. Feasibility and Evaluation Study of Reserved Freeway Lanes for Buses and Car Pools. Alan M. Voorhees and Associates, Inc., Traffic Institute, and Edelman, Inc., Jan. 1971.
7. Net Income Analysis: Traffic, Revenue, and Operating Costs. Gilman and Company, Inc., and Alan M. Voorhees and Associates, Inc., Supp. Rept. 2, Feb. 1969.
8. Study in New Systems of Urban Transportation. Peat, Marwick, Livingston and Company, March 1968.
9. A System Sensitive Approach to Area-Wide Travel Forecasting. Alan M. Voorhees and Associates, Inc., Tech. Memo. 2, 1971.
10. Technical Procedures Used for Distribution, Mode Choice, Assignment and Priority Evaluation for the Recommended Highway Plan. Alan M. Voorhees and Associates, Inc., Jan. 1971.
11. Motor Bus Operations. American Transit Association, Bull. 930, July 1970.
12. Report on the Relationship of Fares, Services, and Costs for D.C. Transit System. Alan M. Voorhees and Associates, Inc., Nov. 1969.

CONSISTENCY IN TRANSPORTATION DEMAND AND EVALUATION MODELS

Dan G. Haney*, Peat, Marwick, Mitchell and Company

This paper is addressed to the evaluation of traveler benefits associated with transportation system alternatives. It is asserted that trip-generation, trip-distribution, modal-split, and traffic-assignment steps, which are carried out by individual mathematical models, in the transportation planning process are frequently not consistent with one another. The inconsistencies arise because, while each is intended to represent travelers' behavior, its mathematical form can imply inconsistent behavior. It is further asserted that a formal evaluation of traveler benefits, which is normally the last step in the modeling process, may provide further inconsistency. These inconsistencies may lead to erroneous conclusions regarding the relative desirability of one system over another. An earlier paper on this subject dealt with achieving modal-split model and evaluation-model consistency. The current paper covers 2 additional aspects of demand modeling: the use of a trip-distribution model and the use of a total-demand model. Procedures for making consistent traveler-benefit calculations with each model are suggested.

•THE PURPOSE of this paper is to attempt to provide a consistent and logical bridge between the methods used to conduct evaluations of alternative transportation systems and the methods used in other phases of the transportation planning process.

An earlier paper (1) dealt with specific problems and inconsistencies that result in the evaluation of multimodal alternatives when the modal-split procedure is not closely coordinated with the evaluation procedure. That paper compared conventional methods of calculation of traveler benefits with recommended methods designed from different modal-split models. It was found that a suitably adapted consumers' surplus approach to benefit calculations provided correct results, whereas the more conventional traveler-expenditure methods were erroneous. A procedure for structuring the analysis method and for making the calculations was presented.

The reader is referred to this paper for background and discussion of the theory behind the results derived here. The presentation in this paper assumes that the reader is familiar with the earlier paper.

The present paper offers suggestions for tying evaluation methods more consistently to the earlier demand-modeling phases in the planning process. Separate sections of the paper deal with recommended procedures applicable to the following 2 approaches to demand estimation:

1. The use of a trip-distribution model, in which different distributions of trips are estimated for different transportation system alternatives; and
2. The use of a total-demand model, in which trip generation, trip distribution, and modal split are accomplished in a single model.

Specific examples of each approach are described; however, the results are generally applicable to other formulations of demand.

---

*Mr. Haney was with the Stanford Research Institute when this paper was developed.

TRIP-DISTRIBUTION MODELS

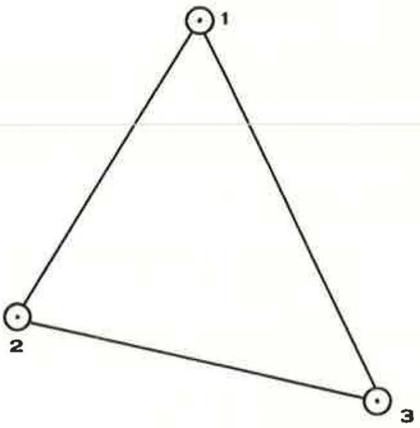


Figure 1. A simple network.

As different transportation system alternatives are postulated and analyzed in a planning process, the most common method of dealing with demand is to make a single estimate of total trip-interchange volumes (a trip table) and to consider that demand as fixed over all alternatives. This procedure has appeal; with demand fixed, a search can be undertaken to find the alternative that produces the given transportation service at minimum direct cost, and no thought need be given to whether different amounts of service are being provided.

However, such a procedure would seem to violate logical inferences as to travel behavior. A reasonable hypothesis of travel behavior is that, as transportation impedances are selectively altered in a study area, persons will change their travel patterns. As particular areas (zones) become more accessible, there will be greater demand for travel to those areas. The analyst may wish to attempt to incorporate such change in travel behavior in the planning process.

If the planning process has been designed along conventional lines, with trip generation, trip distribution, and modal split being accomplished in 3 separate steps, the analyst may adopt a procedure of running the trip-distribution model separately for each transportation system alternative. If this is done under a fixed set of trip ends, redistribution will simply rearrange the trip patterns. The estimate will show some person trips lengthened along paths whose travel times or costs are reduced by a new alternative. Others will be shortened because the model compensates for the farther trips.

As an example, consider a very simple network and zonal system consisting of 3 links and 3 centroids as shown in Figure 1. Assume that the trip ends are to be held constant, as follows: $O_1 = D_1 = 570$, $O_2 = D_2 = 540$, and $O_3 = D_3 = 420$. Assume that the interzonal travel costs of alternative 0 are as follows:

	Destinations		
Origins	Zone 1	Zone 2	Zone 3
Zone 1	6	12	15
Zone 2	12	5	11
Zone 3	15	11	8

If a simple gravity model with a cost exponent (cost being used in a general sense) of 2.0 is used, the resulting travel volumes are as follows:

	Destinations			
Origins	Zone 1	Zone 2	Zone 3	Total
Zone 1	420	82	68	570
Zone 2	82	357	101	540
Zone 3	68	101	251	420
Total	570	540	420	

Now, assume that the travel cost under alternative 1 between zone 1 and zone 3 is reduced by 2 units. The new cost matrix is as follows:

	Destinations		
Origins	Zone 1	Zone 2	Zone 3
Zone 1	6	12	13
Zone 2	12	5	11
Zone 3	13	11	8

The resulting trip distribution, using the same gravity model, is as follows:

	Destinations			
Origins	Zone 1	Zone 2	Zone 3	Total
Zone 1	402	81	87	570
Zone 2	81	365	94	540
Zone 3	87	94	239	420
Total	570	540	420	

Looking at the results in detail, we infer that the changes in behavior can be described in terms of 6 groups:

1. Eighteen persons previously traveling internal to zone 1 now choose to travel to zone 3,
2. One person previously traveling from zone 1 to zone 2 now chooses to travel to zone 3,
3. One person previously traveling from zone 2 to zone 1 now chooses to travel within zone 2,
4. Seven persons previously traveling from zone 2 to zone 3 now choose to travel within zone 2,
5. Seven persons previously traveling from zone 3 to zone 2 now choose to travel from zone 3 to zone 1, and
6. Twelve persons previously traveling within zone 3 now choose to travel to zone 1.

The changes in behavior must be carefully understood. While 2 groups (groups 3 and 4) are saving travel costs by their changes in behavior, the other 4 are incurring higher costs of travel.

Conventional Benefit Calculation

Unless that fact is recognized, a conventional approach of comparing travel costs for the 2 alternatives would simply multiply the travel cost times the travel volume for each zone pair and sum these products over all zone pairs. The most attractive alternative, in terms of travel cost, would be the one with the lower cost. Alternative 1 is preferred if

$$\sum_{ij} C^0_{ij} D^0_{ij} > \sum_{ij} C^1_{ij} D^1_{ij}$$

Alternative 0 is preferred otherwise. The variables given above signify that C^k_{ij} = travel cost between zones i and j for alternative k, and D^k_{ij} = travel volume between zones i and j for alternative k. In the example the total unit costs are 12,543 for alternative 0 and 12,423 for alternative 1. Thus, alternative 1 would appear to be favored by 120 cost units. This net difference is the result of both some cost decreases and some cost increases.

Improved Benefit Calculation

Another way of viewing the changed behavior is to hypothesize that persons choose their travel destination in relation to 2 factors: the value of being at a destination and the cost of getting there. Looking at all possible destinations that are available, each person assigns his particular value to each and assesses the travel cost. The destination that has the highest excess of value less cost will be chosen.

For persons in 4 of the groups given above, the following perceptions should explain their changes in behavior:

Group	Alternative 0	Alternative 1
1	$\text{value}_1 - \text{cost}_1 > \text{value}_3 - \text{cost}_3$	$\text{value}_1 - \text{cost}_1 < \text{value}_3 - \text{cost}_3$
2	$\text{value}_2 - \text{cost}_2 > \text{value}_3 - \text{cost}_3$	$\text{value}_2 - \text{cost}_2 < \text{value}_3 - \text{cost}_3$
5	$\text{value}_2 - \text{cost}_2 > \text{value}_1 - \text{cost}_1$	$\text{value}_2 - \text{cost}_2 < \text{value}_1 - \text{cost}_1$
6	$\text{value}_3 - \text{cost}_3 > \text{value}_1 - \text{cost}_1$	$\text{value}_3 - \text{cost}_3 < \text{value}_1 - \text{cost}_1$

where value_x is the value of being at zone x, and cost_x is the cost of getting to zone x. Because it can be argued that the value of an individual's being at a destination does not change, the benefits for persons in each of the 4 groups must be directly related to the change in travel cost to the new destinations. Some people in the group will have rather high preferences for their selected destinations, while others will feel less strongly. Therefore, depending on the improvement in travel, different amounts of travel will shift. The maximum individual traveler benefit will accrue to the traveler who was previously on the margin between the 2 destinations. As an example, the group 1 marginal traveler will perceive benefits as follows:

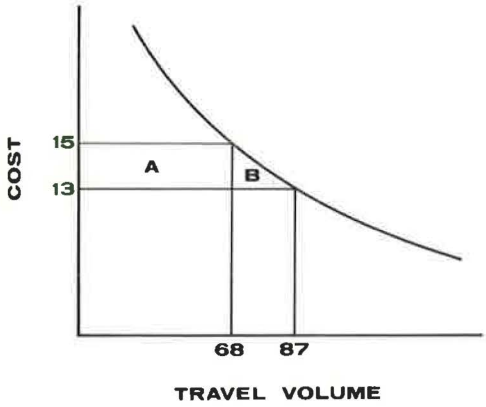
$$\begin{aligned}\text{Benefit} &= (\text{value}_1 - 6) - (\text{value}_3 - 15) - (\text{value}_1 - 6) - (\text{value}_3 - 13)\\ &= 2 \text{ cost units}\end{aligned}$$

At the other extreme, the minimum benefit will accrue to the traveler who is just barely induced to change destinations by the change in travel cost. His benefit is slightly greater than zero.

The consumers' surplus concept can again be used to assess benefits. In this case we consider the zone pairs that have experienced a reduction in travel cost. Using the conventional demand curve diagram as shown in Figure 2, we have the following situation for zone 1 to zone 3 travel: The benefits accruing to the travelers in groups 1 and 2 are shown in the triangle labeled B. In addition, the travelers who traveled between zones 1 and 3 under both alternatives will perceive a benefit, represented by the rectangle A. Using the consumers' surplus formula, we find that the benefits for the persons traveling from zone 1 are

$$\begin{aligned}\text{Consumers' surplus} &= \tfrac{1}{2}(C_0 - C_1)(V_0 + V_1)\\ &= \tfrac{1}{2}(2)(68 + 87)\\ &= 155 \text{ cost units}\end{aligned}$$

Figure 2. Travel demand between zone 1 and zone 3.



A similar argument and calculation can be made for persons traveling from zone 3. The 2 calculations will account for groups 1, 2, 5, and 6, plus the benefits to travelers who continue to travel between the same zones as before.

But what about groups 3 and 4? They have changed their travel behavior, but neither the values of being at the before-and-after locations nor the costs of getting to those locations have changed. In the normalizing process, the gravity model has forced them out of their preferred destinations to less preferred locations. We can, however, observe that, for group 3,

$$\text{Value}_1 - \text{cost}_1 > \text{value}_2 - \text{cost}_2$$

$$\text{Value}_1 - \text{value}_2 > \text{cost}_1 - \text{cost}_2$$

$$> 12 - 5$$

$$> 7$$

This group would be prepared to spend 7 additional cost units in order to travel to zone 1 rather than zone 2, but, by the formulation and operation of the gravity model, they are forced to travel to destinations in zone 2 and thus must be incurring a disbenefit of at least 7 cost units.

$$\text{Disbenefit} > \text{cost}_2 - \text{cost}_1$$

Not having information as to the actual values of being at the 2 destinations, we can understate the disbenefit by the calculation

$$\text{Disbenefit} = \text{cost}_2 - \text{cost}_1$$

$$\text{Disbenefit} = 5 - 12 = -7$$

and, for group 4,

$$\text{Disbenefit} = 5 - 11 = -6$$

Suggested Procedure

In summary, the total procedure for analyses of benefits and disbenefits is as follows:

1. Determine which zone pairs would have reduced travel cost when compared with the base case. In the example, zone 1 to 3 travel and zone 3 to 1 travel would be identified.
2. Identify the costs of travel and the travel volumes for those zone pairs. In the example, the following data would apply:

Item	Alternative 0	Alternative 1
Zone 1 to 3		
Cost	15	13
Volume	68	87
Zone 3 to 1		
Cost	15	13
Volume	68	87

3. Make the consumers' surplus calculation for each of the zone pairs given above. In the example, the following calculations would be made:

$$\text{Consumers' surplus for zone 1 to 3} = \tfrac{1}{2}(15 - 13)(68 + 87)$$
$$= 155$$
$$\text{Consumers' surplus for zone 3 to 1} = 155$$

4. Identify all other zone pairs that would have increased travel. In the example, zone 2 to zone 2 would have increased travel.
5. Determine the travel cost and volumes for those zone pairs for the alternative being studied and for the base case. In the example, for zone 2 to 2 cost = 5, and volume for alternative 0 = 357 and for alternative 1 = 365.
6. Determine the number of trips that are reduced from the origins of the zone pairs given above to each destination and the original costs of the trips. In the example, the cost for zone 2 to 1 is 12 and for zone 2 to 3 is 11. The volumes are as follows:

Zone	Alternative 0	Alternative 1	Difference
2 to 1	82	81	1
2 to 3	101	94	7

7. Calculate the estimated disbenefit for each change in volume.

$$\text{Disbenefit} = (\text{cost}_0 - \text{cost}_1)(\text{volume change})$$

or

$$\text{Benefit} = (\text{cost}_1 - \text{cost}_0)(\text{volume change})$$

In the example, the following calculation would be made:

Benefit for zone 2 to 1 = (1 - 12)(1) = -7

Benefit for zone 2 to 3 = (5 - 11)(7) = -42

8. Add the results of steps 3 and 8 to obtain the total benefit. In the example, the total benefit is

Consumers' surplus for zone 1 to 3	155
Consumers' surplus for zone 3 to 1	155
Benefit for zone 2 to 1	-7
Benefit for zone 2 to 3	-42
Net benefit	261

Such a benefit calculation produces an overestimate of benefits because the place-value losses of some travelers are not known. One can suppose that the overestimate is not large because the travelers who compete and win for the smaller number of trip destinations—in the example at zone destinations 3 and 1—probably have higher excess of values than those who compete and lose. It is the losers who are included in the disbenefit calculation.

It is significant to compare the benefits by the 2 methods of calculation:

Conventional method = 120 cost units
Suggested method = 261 cost units

If the theoretical approach of the suggested method is accepted, it appears to present significantly greater benefits than the conventional method.

Another comparison is also of interest. If the analyst chooses not to redistribute travel for each new alternative and thereby to deal with a fixed trip table, only the travel from zone 1 to zone 3 and from zone 3 to zone 1 would be included. The net benefits here would be 2 cost units per traveler × 68 travelers × 2 zone pairs = 272 cost units, which is much closer to the benefits calculated by the suggested method. However, it appears that not many significant observations can be drawn from the relative similarity of the values in this case. The methods are simply different from each other. However, the suggested method total of 261 cost units is composed of 272 units of benefit to those who do not change destinations at zone 1 and 3, 38 units of consumers' surplus benefits to those who do change destinations [which is shown by the triangle labeled B, in Figure 3: $(2)(\frac{1}{2})(15 - 13)(87 - 68)$], and 49 units of disbenefit for those travelers from zone 2.

Further Observations

Following the procedure described above may not correct all of the inconsistencies between trip distribution and evaluation models. Most frequently, trips are distributed by using a friction factor relation or some other function of one variable, travel time. On the other hand, the evaluation model may use travel time differences multiplied by a value of travel time to estimate the time benefits and may also include other cost variables. One approach that might be taken to avoid this inconsistency would be to accomplish the trip-distribution process by using a number of combinations of time and cost variables, selecting the one that produces the best fit to the observed data, and finding an equivalent price by manipulation of the equations. This price would then be used in a consumers' surplus formulation such as the one presented in the previous section.

Also, even if trip distribution and evaluation were conducted consistently, a multimodal planning problem may not be able to also resolve the modal-split and evaluation problem discussed in the earlier paper. Efforts to attain consistency in demand models are discussed next.

TOTAL-DEMAND MODELS

Within the past 5 years or so, a number of investigators, recognizing the inconsistencies in the various demand models and the fundamental commonality of travelers' decision-making, have attempted to develop overall demand models for transportation planning. Those models place trip generation, trip distribution, and modal split into a single estimation process. Although the successful implementation of such approaches must solve a myriad of problems, it appears that future research may produce promising results that will make the total-demand model more attractive than the currently popular sequence of models.

As an example of the total-demand model, consider the one developed in the Northeast Corridor Project (2, 3, 4). In the context of this paper, the model can be presented as

$$D = (K)\left(\alpha_t C_t^{\beta_t} T_t^{\gamma_t} + \alpha_a C_a^{\beta_a} T_a^{\gamma_a}\right)^{\delta}$$

where

D = total demand between 2 zones, in number of persons;
K = variable representing a combination of economic, demographic, and travel characteristic variables plus a constant term, all of which do not change with changes in the transportation system;
C_t = cost of travel by transit;
C_a = cost of travel by automobile;
T_t = time of travel by transit; and
T_a = time of travel by automobile.

The variables α, β, γ, and δ are constants determined in calibration.

The demand for the transit mode is computed as follows:

$$D_t = (D)\left[\alpha_t C_t^{\beta_t} T_t^{\gamma_t} / \left(\alpha_t C_t^{\beta_t} T_t^{\gamma_t} + \alpha_a C_a^{\beta_a} T_a^{\gamma_a}\right)\right]$$

and similarly for the automobile mode.

After demand is estimated for each transportation system alternative by using a model such as that shown above, the planner needs to compare alternatives and to provide information that can be used in selecting the one deemed most desirable. Among the comparisons usually made are comparisons of traveler benefits, in which travel cost, travel time, and other effects are assessed.

A method of evaluating traveler benefits that is consistent with the travel behavior implied by the total-demand model is illustrated by considering 2 alternative transportation systems having costs and times as follows:

Alternative 0	Alternative 1
C_{0t}	C_{1t}
C_{0a}	C_{1a}
T_{0t}	T_{1t}
T_{0a}	T_{1a}

then

$$D_0 = (K)\left(\alpha_t C_{0t}^{\beta_t} T_{0t}^{\gamma_t} + \alpha_a C_{0a}^{\beta_a} T_{0a}^{\gamma_a}\right)^{\delta}$$

and

$$D_1 = (K)\left(\alpha_t C_{1t}^{\beta_t} T_{1t}^{\gamma_t} + \alpha_a C_{1a}^{\beta_a} T_{1a}^{\gamma_a}\right)^{\delta}$$

The benefits that accrue to travelers between 2 alternative transportation systems can be found by using the concept of equivalent price (1). The equivalent price is, as defined in the earlier paper, the price that would have resulted in the same new demand if only the cost had changed. In other words, a demand D_1 for a new alternative might result from an improvement in both time and cost. The equivalent price would produce the same new demand but only with a cost (price) change. The equivalent price (if an improvement in the transit system is being analyzed) is found by solving the following equation for C_{1t}^{e}:

$$D_1 = (K)\left(\alpha_t C_{1t}^{e\,\beta_t} T_{0t}^{\gamma_t} + \alpha_a C_{0a}^{\beta_a} T_{0a}^{\gamma_a}\right)^{\delta}$$

If rearranged,

$$C_{1t}^{e} = \left\{\left[(D_1/K)^{1/\delta} - \left(\alpha_a C_{0a}^{\beta_a} T_{0a}^{\gamma_a}\right)\right]/\alpha_t T_{0t}^{\gamma_t}\right\}^{1/\beta_t}$$

In words, the demand on the new alternative is first found by using the demand equation and the values of cost and time for both modes under the new alternative; then, the equivalent price is found by using the values of the new demand and the old automobile time, automobile cost, and transit time.

An alternative method of finding the equivalent price would be to use the modal-split formula. Here, C_{1t}^{e} is as follows:

$$C_{1t}^{e} = \left(\left\{\left[(D_1/D_{1t}) - 1\right]\left(\alpha_a C_{0a}^{\beta_a} T_{0a}^{\gamma_t}\right)\right\}/\alpha_t T_{0t}^{\gamma_t}\right)^{1/\beta_t}$$

The choice among such alternatives will depend on the mathematics. In some cases, one method may not be reducible to an analytical expression, whereas another may be.

In the operation of the demand model, an improvement in the cost or time of 1 mode will cause not only a diversion of trips from the previous mode to the improved mode but also an increase in the total number of trips—which is referred to as induced traffic.

The total traffic on the improved transit mode results from a combination of the original traffic, the diverted traffic, and the induced traffic. The diverted and induced traffic are as follows:

$$D_{1t} - D_{0t} = D_{\text{diverted}} + D_{\text{induced}}$$

$$D_{\text{diverted}} = D_{0a} - D_{1a}$$

$$D_{\text{induced}} = (D_{1t} - D_{0t}) - (D_{0a} - D_{1a})$$

The traveler-benefit calculations can be separated into the 3 groups described in the following subsections.

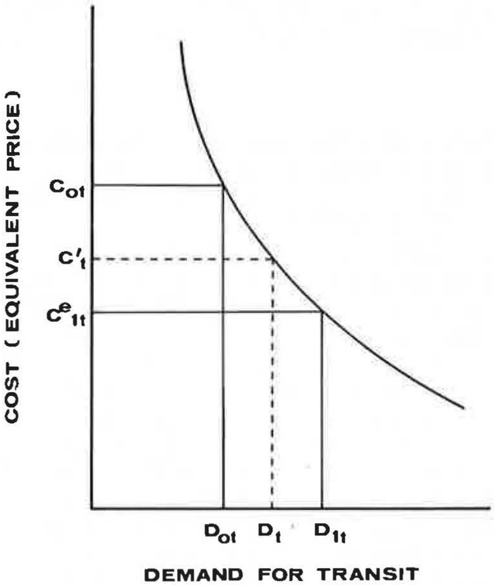
Original Traffic—For the previous transit travelers who experienced reduced costs, the net benefits are easily calculated.

$$NB_{\text{transit travelers}} = (C_{0t} - C_{1t}^{e})\,(D_{0t})$$

Induced Traffic—For the induced travel, arguments similar to those presented in the earlier paper can be made regarding the demand for travel, the willingness to pay, and the cost of travel. These arguments result in a consumers' surplus approach to demand estimation.

Figure 3 shows a typical demand-curve relation between price and demand. Because the following arguments assume a transit system improvement, a transit-demand curve is shown. (Similar arguments regarding highway improvements could be pursued by using an automobile-demand curve.)

Figure 3. Transit-demand curve.



Those persons having a willingness to pay greater than the cost C_{0t} are the transit travelers under alternative 0 and are represented by points on the demand curve to the left of D_{0t}. Those persons represented by points on the demand curve to the right of D_{0t} have a willingness to pay for transit less than C_{0t}. It is important to recognize that these persons really consist of 2 groups because some are traveling by automobile and some are not traveling at all under alternative 0. Thus, they are those that may be diverted to transit and those that may be induced to travel (by transit).

If improvement alternative 1 were installed, at an equivalent price of C^e_{1t}, the total demand for transit would increase to D_{1t}. The increase is made up of members of each of the 2 groups identified above. Even a very small reduction in cost would result in an increase in travel traceable to both groups. Therefore, members of both groups lie at all points along the demand curve. This is an important characteristic of the model.

The group of persons who would be induced to travel would, in theory, have considered both automobile and transit under alternative 0 and decided that they were not willing to pay either price. If transit improves and highways do not, such as in alternative 1, the crucial choice of this group is between transit and not traveling.

Suppose a tiny fraction of the latter group is at the margin under the original conditions. Because they are at the margin, they perceive no difference in benefit between using transit or not traveling. However, if the transit price is reduced, they would choose to travel by transit, and their benefit is indicated by the difference between their willingness to pay and the price of transit. For the assumed improvement, this difference is $C_{0t} - C^e_{1t}$. Another tiny fraction of the latter group will be at the margin between travel by transit and not traveling if the transit price is C'_t. They lie at the D'_t point on the curve, and if the improved transit system were installed, the difference between their willingness to pay and the price is $C'_t - C^e_{1t}$. Similar reasoning can be applied to each traveler regardless of where he is represented on the demand curve.

For all of the persons who would choose to use transit rather than not traveling, the total benefits can be derived. If the induced travelers are divided into small m-increments depending on their location on the demand curve, the net benefits could be estimated by summing the benefits that accrue to each increment as follows:

$$NB_i = \sum_{j=1}^{m} (\hat{C}^e_{itj} - C^e_{it})\, n_j$$

where

NB_i = net benefits from the induced travel for alternative i,
$\hat{C}^e_{itj}$ = average equivalent cost at a point on the demand curve representing the j th increment,
C^e_{1t} = equivalent price of alternative 1, and
n_j = number of induced travelers in the j th increment.

If the procedure described in the earlier paper is followed, the sizes of the increments could be reduced and the number of increments increased until their number approached infinity. Then, methods of calculus could be used to produce a more exact calculation. However, these net benefits can be approximated, following arguments similar to those presented in the earlier paper, by a consumer's surplus formulation.

$$NB_{\text{induced}} = \tfrac{1}{2}\left[C_{0t} - C^{e}_{1t}\right]\left[(D_{1t} - D_{0t}) - (D_{0a} - D_{1a})\right]$$

This formulation depends on the earlier recognition that members of both groups lie at all points along the demand curve.

Diverted Traffic—For the diverted travelers, similar arguments on the willingness to pay can be made; these travelers also consider automobile travel, transit travel, and no travel. Under the alternative 0 conditions, the members of the group lying along the demand curve between D_{0t} and D_{1t} choose to travel by automobile. As transit improves from C_{0t} to C^{e}_{1t}, their crucial comparison is between automobile travel and transit travel.

Under alternative 0, a tiny group of those travelers who lie at D_{0t} on the demand curve are indifferent between transit and automobile. They are willing to pay C_{0t} to use transit and no more. Similarly, they are willing to pay C_{0a} to use automobile and no more. However, if the transit price were reduced, they would choose to travel by transit. Their benefit would be indicated by the difference between their willingness to pay for transit and the cost of transit. For the assumed improvement, this difference is $C_{0t} - C^{e}_{1t}$. Another tiny fraction of previous automobile travelers will choose to use transit only if the transit price is less than C'_{t}. They lie at the D'_{t} point on the curve shown in Figure 3; and, if the alternative 1 transit systems were installed, the difference between their willingness to pay and the cost is their perceived benefit, $C'_{t} - C^{e}_{1t}$.

Thus, the total benefits can be derived for the group of persons who would choose to use transit under alternative 1, those being diverted from automobile. The argument is similar to that presented earlier, which displayed summation of benefits over a number of traveler increments for the induced travel. Now, the m-increments are increments of travelers in the diverted category.

The result is that the net benefits to the diverted traffic can also be approximated by a consumer's surplus formula.

$$NB_{\text{diverted}} = \tfrac{1}{2}(C_{0t} - C^{e}_{1t})(D_{0a} - D_{1a})$$

Total Benefits

The total net perceived traveler benefits for the improvement in the transit system is the sum of the net benefits to the 3 types of travelers.

$$\begin{aligned} NB &= NB_{\text{original transit traffic}} + NB_{\text{diverted traffic}} + NB_{\text{induced traffic}} \\ &= (C_{0t} - C^{e}_{1t})(D_{0t}) + \tfrac{1}{2}(C_{0t} - C^{e}_{1t})(D_{0a} - D_{1a}) \\ &\quad + \tfrac{1}{2}(C_{0t} - C^{e}_{1t})\left[(D_{1t} - D_{0t}) - (D_{0a} - D_{1a})\right] \end{aligned}$$

which simplified to the well-known consumers' surplus formula is

$$NB = \tfrac{1}{2}(C_{0t} - C^{e}_{1t})(D_{1t} + D_{0t})$$

Suggested Procedure

The procedure that has been followed in this example to calculate the traveler benefits can be applied to any transportation system in which a total-demand model is used. The following steps should be followed to compute net traveler benefits. The benefits should be computed for each zone pair.

1. Identify those zone pairs that, compared with the base case, have increased traffic under the alternative being studied.
2. Compute the equivalent price C^{e} for each zone pair.

3. Compute the consumer's surplus for each zone pair. The expression

$$NB = \frac{1}{2} (C_{0t} - C^e_{1t}) (D_{1t} + D_{0t})$$

would be used in the case of transit improvement.

4. Sum the net benefits over all improved zone pairs.

It has been recognized that one difficulty with the Northeast Corridor demand model is that it tends to reflect improvements in a given mode in much larger quantities of induced traffic and much smaller quantities of diverted traffic than are reasonable. Various steps have been taken to correct this problem.

Although this problem is very real and must be dealt with when the needed capacity on the various modes is considered, the resulting consumers' surplus formulation given above mitigates the problem somewhat. The formulation does not require separation of the 2 types of increase in travel—induced and diverted. In other words, the individual benefits to the travelers who elect to travel by the improved mode is the same regardless of whether they are induced or diverted travelers.

A NOTE ON THE INTERPRETATION OF THE ALPHA TERMS

In the total-demand and modal-split formulations presented in this and the earlier paper, the term α is used as a constant term in an expression describing either a difference in utility between modes or the utility of an individual mode.

Such models should normally be developed by experimentation with a number of formulations of the various independent variables in order to arrive at an expression that most nearly explains the observed behavior. Regardless of the amount of experimentation or the number of variables included, an unexplained difference invariably remains. This difference is represented by the α terms used in the formulations given in this and the earlier paper. Some investigators have referred to this unexplained difference as a comfort and convenience factor. The author would prefer to refer to it as the total unexplained difference, without theorizing any particular name or cause.

This difference is a difference between explicit modes of travel, between the modes against which travel decisions have been studied and the model developed. The explicit model should only be used to evaluate changes in the modes that were used in the calibration process.

We submit that there is no such thing as an "abstract modal model," in the sense that any technological mode can be studied in the context of a demand or modal-split model calibrated for 2 explicit modes. The new technological mode would produce a different α term that would stand for the preferences for that mode that are not explained by the independent variables. (Similarly, the significance of the independent variables under a new mode assumption may be different from that under the condition that existed for calibration of the model.)

This observation implies not only that considerable care should be used in applying demand and modal-split models to new modes but also that care should be used in conducting evaluations using these models.

REFERENCES

1. Haney, D. G. Consistency in Transportation Modal-Split and Evaluation Models. Highway Research Record 348, 1971, pp. 61-72.
2. McLynn, J., and Woronka, T. Passenger Demand and Modal-Split Models. Northeast Corridot Transportation Project, U. S. Department of Transportation, NECTP Rept. 230, Dec. 1969.
3. Donaldson, J., et al. National Bureau of Standards Modeling for the NECTP. Northeast Corridor Transportation Project, U. S. Department of Transportation, NECTP Rept. 213, Dec. 1969.
4. NECTP Rept. 209. Northeast Corridor Transportation Project, U. S. Department of Transportation, April 1970.

DISCUSSION

Ezra Hauer, Department of Civil Engineering, University of Toronto

Haney points out the fallacy of judging the efficiency of a transport system on the basis of aggregate cost incurred in travel. The minimization of total cost is a reasonable objective when different systems achieve the same utility. But, because alternative transport systems lead to different origin and destination linkages (and level of trip generation), the condition of equal utility obviously is not fulfilled. Specifically, the paper argues that changes in the transportation system induce some travelers to reexamine their selections of origins or destinations. For those travelers, before-and-after cost comparisons are meaningless because not only did the cost of the trip change but so did its utility.

The logic of the argument presented is sound. Yet, the reader is left with the uncomfortable sensation that its practical consequences (increased net benefits to be used in justification of transport investment) point in the wrong direction. A stone is being added to the benefit side of the scales while society's finger tries to push the pointer in the opposite direction. To repeat a familiar argument: Transport investment leads to ease of travel, which is conducive to a footloose selection of place of residence and work, which in turn generates an apparent dependence of the society on travel and makes investment in transport system improvements easy to justify, and so on ad infinitum. The outcome of this self-perpetuating process is the present urban structure; its diffuse activities render automobile dependence absolute and the concept of "choice through mobility" questionable. For some years now, several communities dared to question the wisdom and expertise of planners and opposed their recommendations for transport investment. The dilemma is obvious: How is it possible that investments, which are justified on the basis of values and preferences of all members of a community (as calculated by the planner), are frequently opposed by vigorous political action of the very same group?

The paradox may be easily explained in terms of vocal minorities, professional activists, irrationality and misinformation, uneven incidence of costs and benefits between groups, imperfections of the political process, or myopic decision-makers. The planners, however, must seriously consider whether the paradox stems, at least in part, from a professional bias. Specifically, has the planner not been systematically more diligent in searching for benefits than in scouting for "costs?" It is on this basis that Haney's paper may be found wanting.

Whether justification of public investment should proceed on the basis of the elusive "consumer surplus" while investment in the private sector can rarely do so is a moot question. But even if the legitimacy of incorporating the consumer's surplus into the benefits of investment is not questioned, the suggested evaluation scheme lacks in comprehensiveness.

It is well known that only a part of the cost associated with the performance of a trip is borne by the trip-maker himself. Some of the cost is imparted to his fellow travelers in the form of increased congestion, safety hazards, and the like. This component of the cost should not remain unaccounted for. Its neglect is particularly objectionable when, as a result of investment, new trips are being made. In this situation, it is the planner who is responsible for the incorporation of costs that the induced traveler is incapable of perceiving. The "congestion cost" is not merely academic hairsplitting relating to hypothetical situations. It may be seen at work in the common example of transport and other investment in the outer reaches of the cities, inducing new trip-makers into commuting to the downtown. Although the new travelers indicate by their decision to travel the receipt of a net benefit, the added plight of the original users remains anonymous and is not added to the accounting ledger. It is, however, present in a growing body of public sentiment that questions the desirability of "growth" because it usually means deterioration for those already in the system.

The second component of transport cost unaccounted for in Haney's accounting procedure is the cost imparted by trip-makers to nonusers of the transport system. Those are simply not present in the model. Yet, almost every major transportation investment in the past decade had to contend with relocation, air pollution, noise, visual obstruction, community disruption, and other impacts on nonusers.

In summary, Haney's method of calculating traveler benefits is certainly rational and consistent with currently used models. Lest consistency be construed to mean comprehensiveness, I find it necessary to point out that the benefits and costs accounted for in this procedure form but a part of the overall impact of transport investment. Under no circumstances can it be regarded as a "total procedure for analyses of benefits and disbenefits." And if used so, it will in all likelihood result in yet another confrontation of the public versus the "highway establishment."

AUTHOR'S CLOSURE

Hauer's principal concern is that transportation planners may become enamored with advancing the state of the art in the estimation of benefits from investments in transportation systems at the expense of advancing the art in estimating costs, as they are broadly defined. With this point I certainly agree, and I would hope that my earlier writings would testify to that philosophy.

The intent in the paper, perhaps more explicitly stated in the earlier paper (1), was to deal only with traveler benefits and costs. To treat isolated problems in methodology is appropriate within the format of professional papers, as I am sure Hauer will agree.

With regard to specific points in his comment, Hauer argues that a biased effort to find benefits can lead to diffused urban structure. Perhaps this is so when highway planners justify new highway projects. But what about transit planners? Do they search to find benefits so that they can counteract the urban-sprawl effect? I think not. I would hope that both the highway planner and the transit planner would search for 2 kinds of optimums: the optimal transportation system and the optimal effect of transportation in influencing land use. Neither of these is easy to define and measure, but I would hope that the procedures presented in this and the earlier paper would lead to improved solutions to the problem of finding optimal (balanced) transportation systems. At the present time, I would prefer not to argue as to the most attractive land use pattern, although a subsequent paper will address itself to a facet of that problem. Suffice it to say that the procedures described in the 2 papers could—or might—provide the planner with an improved way of assessing the potential of transit systems to condense patterns of land use as well as of highway systems to diffuse the patterns. Personally, I do not think that the effects of different technology cannot exclusively be labeled as producing, condensing, or diffusing land use changes.

Regarding Hauer's discussion of disbenefits to existing travelers, if both the demand and supply curves are specified and if the supply curve produces increased cost with increased usage, i.e., congestion cost, it is possible to estimate the magnitude of the increased cost to the original users. It can be estimated as the difference between the actual cost to the original users (as well as to the induced users) for the new system, less the cost to the original users for the new system had the induced travel not materialized. Thus, the disbenefits to original users can be added to the accounting ledger.

FORECASTING TRAVEL DEMAND FOR NEW URBAN TRANSPORTATION SYSTEMS

Anthony R. Tomazinis, University of Pennsylvania

The paper presents the experience gained and the conclusions reached by a University of Pennsylvania team who carried out the forecasting procedures for the minicar transit system as a part of a study by a larger group at the university. The technical problems in determining the system characteristics and the levels of service, in defining the travel needs of the study areas, and in developing and applying a method of travel-demand forecasting for a novice system are discussed in the light of experience gained in the minicar transit project. The new system was studied for the Philadelphia CBD and the low-income districts that surround the CBD on 3 sides. Major methodological dilemmas and limitations of market analysis tools are also discussed. Forecasts were made for manifest travel demand and latent travel needs.

•THIS PAPER presents the conclusions reached from a case study of the problem of forecasting demand estimates for new, untested urban transportation systems destined for use within an uncontrolled, competitive, and continuously changing urban environment. The system under consideration was one based on small vehicles available for rent by the trip, the hour, the day, or long periods of time by any qualified driver. The vehicles were called minicars, and the system was called minicar transit system or MTS (1, 2, 3, 4). The effort was carried out by researchers of the University of Pennsylvania as part of a larger project that included the development of the vehicle itself, the MTS, and the determination of the feasibility and desirability of the vehicle and the system as a whole. The project was financed by the Urban Mass Transportation Administration as part of its program in developing new systems.

THE PROBLEM

The technical problem in this aspect of the project was primarily based on estimating with satisfactory reliability the extent to which the new minicar system will be utilized by its prospective users. This simple definition of the problem is sufficient to indicate a number of particulars of the overall problem. The transportation planner was asked not only to forecast the potential demand for a new system but also to suggest what would be the location of the most desirable test and application of the new system. In the case of the minicar project, the efforts focused on testing the feasibility and desirability of such a new system in 2 parts of the Philadelphia region: the Philadelphia central business district and the low-income ring that surrounds the CBD on 3 sides. The 2 areas are to be taken together or separately.

A second major problem faced from the outset was the determination of both the technological system characteristics and the definition of the services that the system would provide for the study areas. It is of particular interest to note that, in developing a new transportation system and in testing the desirability and feasibility of such systems in the early stages of the effort, it is necessary for the transportation planner to undertake the responsibility of suggesting both the desirable technological characteristics of the new system and the type and pattern of services that the new system may be called on to offer. This was also true in the case of the minicar project.

These and several other problems form the core of any effort to forecast travel demand for new systems. The steps that ought to be followed were complicated and required a sequence that was supported by logic and was consistent with the availability of the data sets that were necessary for the completion of the project. Figure 1 shows the various steps and the sequence of actions that were developed and applied in the case of the minicar project in response to these methodological requirements.

What follows in this paper is a discussion of the various phases of analysis and the conclusions reached in performing the various tasks.

PROBLEMS OF AREA SELECTION AND LEVEL OF AREA ANALYSIS

Several problems of particular significance arise as soon as the area of the potential application of the new system is considered. One must immediately refer to the central goals of the entire effort and the initially postulated capabilities of the new system. For instance, in our minicar project the central goals of the effort were the alleviation of street traffic congestion and the provision of an additional means of transport to those urban residents who did not hiterto have a satisfactory choice of travel mode. Also, from the outset, the new mode of travel was defined as a means of serving short-to-median-length urban trips of those who can drive and of providing the service data cost somewhat below that of the private automobile and taxi and somewhat above that of the public transit system. In response to these 2 essential determinations—the central goals and the essential characteristics of the new system—the research team selected 2 areas for testing the initial system: the Philadelphia CBD where all types of urban traffic congestion occur in their worst form and the low-income districts that surround the Philadelphia CBD on 3 sides. The choice of the test areas or study areas was the start of an extensive analytical process to determine 4 essential aspects of each area, as follows.

1. Determination of the boundaries of each area proved to be critical in our effort in both areas. At least 2 major considerations enter here: One is the intrinsic characteristics of the area, and the other is the functional requirements of the system. With regard to the first consideration, it is frequently clear that neither does the CBD of a region have a demarcation line that separates it from the next block of activities nor do the low-income areas have a distinct demarcation point (physical or economic) that separates them from areas with slightly higher income levels. With regard to the second consideration, the requirements of a "systems testing" imposed 2 additional actions. First, a small "island" of high-income households in the midst of a low-income district or a park area in the midst of the CBD would have to be included in the overall outline of the study area. Second, the core study area would have to be related and functionally associated with the surrounding area with which the "study area" interchanges the vast majority of its trips. Thus, 2 study areas become apparent: the core study area (CSA) with which we were primarily concerned and the regional study area (RSA) within which the CSA is located and with which the CSA constantly interacts.
2. A second aspect of the problem was a conflict between our project objectives and the availability of data for each component of the study area. It soon became clear that restrictions on the outline, size, and even location of the study area ought to be accepted because of data requirements.
3. The third aspect of the problem that rapidly emerged was the appropriate level of analysis of the conditions prevailing in the study for the purposes of the project. As the analysis of the conditions prevailing in the study areas and of their travel needs was carried out, it became necessary to frequently impose severe restrictions on the individual analysis in order to avoid aimless analytical ventures that might be of interest but that had very little relation to the objectives of the project.
4. The determination of indexes of service deficiency was the fourth major aspect of the problem. This requirement is indeed of particular significance and is especially complex. It involves both the physical and the operational characteristics of the present and planned systems as well as an essential determination of "standards" and service levels that are either desirable or feasible or both. It seems that an incomplete, shortsighted, or utopian determination of standards or service levels may prejudge the outcome or even the evolution of the entire test of the new system. In our case we found particular difficulties in establishing widely acceptable and quantifiable concepts of latent travel demand for the low-income areas and in establishing satisfactory differentiations between public and private concepts of requirements for service in both core

Figure 1. Forecasting procedure of minicar project.

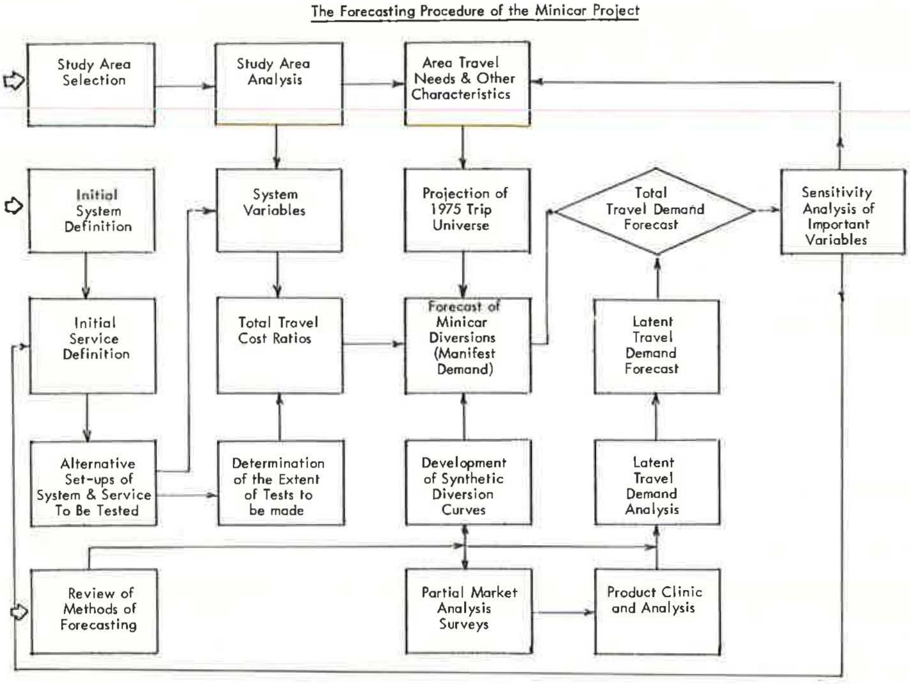


Table 1. Minicar system setups for low-income areas.

Setup	Minicar Fare (cents/mile)	Value of Time (dollars/hour)	Terminal Spacing (mile)	Automobile Parking Cost (cents)	Maximum Fee on Round Trip (dollars)
I	20	1, 2, and 3	1	—	2.00
II	20	1, 2, and 3	⅓	10	2.00
III	15	1, 2, and 3	⅔	10	1.50
IV	10	1, 2, and 3	⅓	25	1.00

study areas. Even the determination of the "travel needs" of the study areas proved to be particularly complicated when the introduction of a totally new system with smaller and different use limitations was taken into account.

SYSTEM DETERMINATION FOR TESTING

The system determination for testing involved 3 components. One was the outline and the "firming up" of the physical and technological characteristics of the system. In addition to the means of costing a trip, such characteristics in our case proved to be the vehicles' operating speed, capacity, maneuverability, driving requirements (i. e., licensed drivers required, physical availability in parking lots and garages (a minimum delay in picking up and delivering vehicles), and driving reliability. The second component was an early determination of the unit cost of travel in a manner that included all pertinent costs such as vehicle costs, costs of parking and storing, costs of accounting and billing, costs of repairs, accident costs, and all other overhead costs of the system. The third component of system determination for testing was the early areal distribution of the system associated with an early determination of the types of services (trips) the new system will be able, appropriate, and available to compete for.

All 3 components of system determination for testing proved difficult to define in our project. In general the more novice the system is and the more alternatives it makes possible, the greater is the difficulty in firming up a total system for testing purposes. In the case of anticipating adoption of a new system within the framework of a free urban society and in an open competition with all the other systems, the determination of the new system for testing purposes is as critical as any other aspect of the overall effort. The transportation planner can right there "kill" the project or, on the other extreme, delude himself into believing the superiority of the idea he is testing.

GEARING THE SYSTEM FOR TESTS

At the completion of the 2 first phases (system determination and study area selection and analysis), the stage was set for gearing the system for comparative simulation tests. The work in this phase could most reasonably start from a determination of what and how many combinations of systems and services will be tested. In our case we felt it necessary to test at least 2 alternative setups for the CBD and 4 setups for the low-income areas. Table 1 gives the setups for the low-income areas. Several elements can be varied at this early stage in the search for various definitions of optimum system. In the minicar system considerable variation was available in setting the fare levels, in determining the maximum cost of a trip, in assuming different values of time for different groups of users, in determining the frequency of system terminals and hence the level of availability of minicars, and in assuming parking cost for the private automobile. Additional items could be varied but the ones already included presented enough complications to restrict us from expanding the list.

The notions of optimum system were also quite variable and only to a limited extent subject to analytical methods of optimum determination. This is so because in only a few cases would the new systems be called on to simply maximize their revenue or to simply produce most efficient services. For most cases, the type of service offered, the group of beneficiaries, and the impact on other competitor or supplementary systems are of equal significance in current testings of new systems. In fact, concepts of this broader type have in the recent past cast unfavorable light on many new systems that otherwise would appear desirable and ready for development.

The determination of distinct combinations of services and facilities to be tested leads to the development of system variables that would, from then on, represent the system. This aspect of the undertaking produces a welcome clarification in the minds of the analysts with regard to what the new system can offer. The system variables express in essence the operating characteristics of the system and represent the dimensions along which the new system can be measured and compared both with expressed travel needs of the study area and with the capabilities of the competitive systems. For our minicar project, we found the need to carry investigations that measure out-of-pocket requirements, running time requirements, and excess time and cost requirements for the completion of the trip. Unfortunately, no measures could be carried out with regard to the system availability to all its potential customers or with regard to comfort, convenience, safety, and reliability of the new systems. Thus, we had to assume that these systems variables would not be any different from the ones characterizing the competitive systems and, therefore, would not produce any influence anyway.

The measurement of system variables involves both the new system and all the other systems currently in use or contemplated to be in use by the time the new system will be in place. The actual mechanics, however, involve both abstract or averaged measurements and specific measurement of interchanges for which more than one system will be available. Thus, it is an important step in the whole process to determine the extent to which detailed measurements will be carried out. There are 2 approaches at this point. The analyst can either produce a blanket measurement of system variables for all probable interchanges or select only a set of representative interchanges for which detailed measurements can be made. The choice depends on the availability of funds and time. The first approach is all inclusive and presents a plethora of measurement points. In our minicar project, we selected the second approach primarily because of limitation in money and time. However, we included enough representative measures to cover the spectrum of probable trip lengths and sets of circumstances that may influence measurements in the new system or the competitor system.

SELECTING FORECASTING PROCEDURES

The entire effort is brought to its most critical point at the moment the choice for a specific forecasting procedure is made. There are several approaches in forecasting travel demand for a totally new system. In general they can be classified into 2 groups: the manifest preference method and the expected preference method or alternatively the travel needs simulation method and the market analysis method. Both approaches include a number of variations and are subject to ad hoc combinations of methods. [The methods and techniques of the abstract-modes models can easily be seen as falling within the manifest demand classification because they clearly utilize a manifest data base (5, 6).] In general the market analysis approach would proceed with one of the many possible forms of questionnaires on the basis of which the potential consumer will be asked to express an a priori judgment and preference about what he would consider preferable and what he would do in case the new system becomes available to him. The exercise is accompanied with a brief description of the new system and with a request to express a choice, all other things being equal. One of the most sophisticated questionnaires in this group is the one that is usually called a discriminant questionnaire in which the potential customer is asked to express an incremental like and dislike for the various characteristics of the new system as part of a technique usually referred to as semantic differentials. Variations of this method are used by most major manufacturing companies (the automobile industry included) in testing what their new products will encounter in the market or how their new products must be improved (7).

The market analysis approach is based on market segmentation from the outset, and in many cases it is reinforced with a product clinic. The term "product clinic" is given to a special effort, usually carried on annually by the automobile companies, in which the new product is exhibited and explained to successive segments of the market. At the end of each routine, the potential consumers are usually asked to complete a questionnaire indicating their feelings, preferences, and choices on the basis of what they know, what they have seen, and what they have been told about the new product.

In contrast to these market analysis techniques, the analyst has available for his use considerable data that indicate what the potential consumers have already done in cases where a choice was involved. The approach is based on manifest behavior and preferences, consciously or unconsciously made. If, then, the analyst succeeds in relating the manifest choice with the circumstances within which the choice was made and with the relative characteristics of the competing systems, he can assume that in similar cases in the future the potential consumer will act similarly. This approach is the essence of modal-split analysis and projection carried out by most urban transportation studies. In doing so, the analyst has available several methods that he may employ such as correlation analysis, diversion curves, abstract-modes models, and stochastic choice models.

As the minicar system forecasts were carried out, the dilemma of the approach to be followed was felt in its full force. The new system was sufficiently novice in its physical outline and operating characteristics to warrant market analysis experimentations and product clinic findings. The new system was, however, sufficiently similar to present systems to render the findings of other modal-split studies very useful in forecasting usage. After all, cost variations and travel time variations have been in the heart of modal-split analyses of most major modal-split studies.

As a result of these realizations, the team proceeded by utilizing both approaches but placing major emphasis on the manifest preferences method. A number of market analysis studies were made for most segments of the potential market. In addition, a product clinic was held in a special room of the university for about a month to which more than 100 groups were invited in sequence. The market analysis studies were small in size and attempted to measure probable future preferences in specific cases. The product clinic participants were given the opportunity to inspect and experience the minicar itself, review the results of the first phase of the study, listen to additional explanations, and then carry on a conversation with key members of the research team. The invited groups varied from leadership groups such as councilmen and civic leaders to small street-corner groups and informal associations. At the end of each session each participant was asked to complete a questionnaire.

The main forecasting effort proceeded, however, by utilizing data and conclusions based on the 1960 home interview survey of the Penn-Jersey Transportation Study. The target date was 1975 for which a preliminary regional transportation plan and projection was made by the Delaware Valley Regional Planning Commission in the period 1963-66.

The application of a modal split approach in determining the minicar share of the manifest trips involves still many steps and assumptions that must be carried out before the process is completed. The first major commitment to be carried out was the exact method to be utilized in estimating the share of the minicar trips, among the methods available, the conclusion was in favor of developing a set of synthetic diversion curves. The correlation analysis was excluded because of both the need it has for empirical data in deriving the coefficients of the equations and the undue strictness of the relations that the equations convey. The stochastic approach was also rejected from the outset because of the undue articulation of the hypothesis and relations that the approach requires and that in no way could be met in this speculative and experimental study.

The set of synthetic diversion curves was developed by using the experience gained in the recent past in developing modal-split diversion curves in actual situations and by utilizing also a set of logical-commital statements that establish the 2 extreme points of each curve as well as its midpoint. Then by assuming the applicability of the economic principle of diminishing marginal returns, the team formulated a set of diversion curves approximating the shape of diversion curves of the recent past. Figure 2 shows the resultant curves for 2 sets of choices. These curves were then utilized in association with a composite system variable developed with the total travel cost of each interchange under consideration.

The total travel cost variable was formed as the key variable for the diversion curves, against which the share of minicar travel could be estimated. The variable represented the summary cost of all out-of-pocket costs, the time cost, and an approach time cost at the 2 ends of the trip. The minicar fare was explicitly counted on the basis of the assumptions advanced in each system setup for testing. The automobile costs were inclusive of operating costs and fixed costs of depreciation and insurance. Varying automobile parking costs were added. For transit and taxi trips, the fares of each trip were explicitly included. The total costs of each system were then made to form ratios that were further investigated with respect to trip length, value of time, level of parking costs, and amount of approach costs for each trip. Figure 3 shows some of the variations of the cost ratios with trip length and value of time for minicar versus automobile and minicar versus transit.

On the basis of this set of assumptions, presumptions, and partial calculations, the

Figure 2. Synthetic diversion curves for low-income areas.

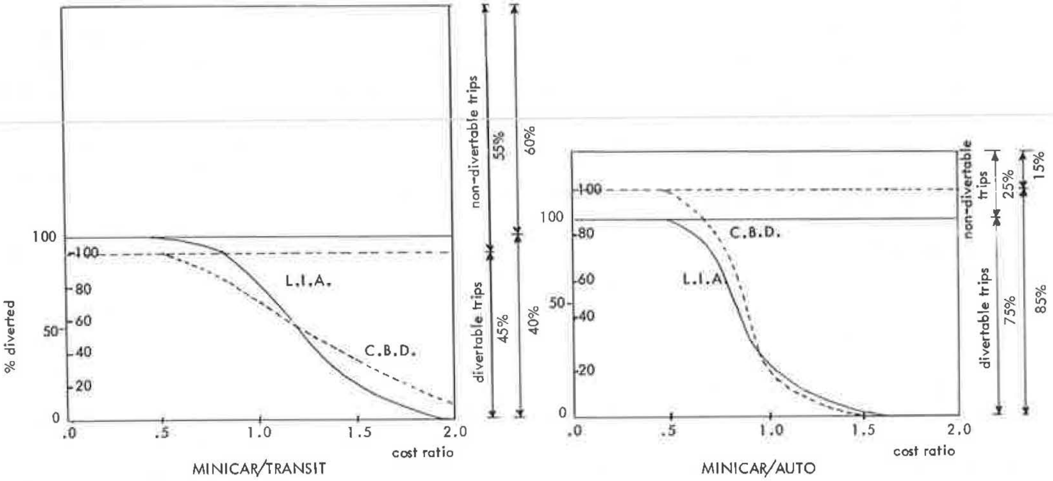
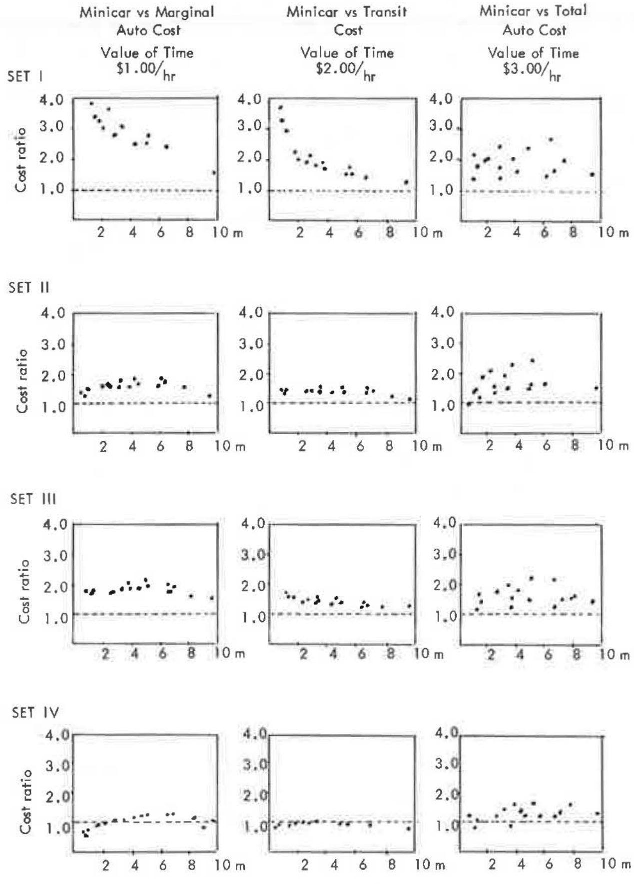


Figure 3. Minicar cost ratios.



forecasts for the share of minicar trips from the total manifest trip projection of the residents in the region were made for 1985. The essential question that the team was attempting to answer was, What would happen to the trip patterns by mode if, in addition to what exists and is planned in the mean time, a minicar transit system were in operation in the study areas in 1975? The first part of the answer was to be derived through these sequential and somewhat complicated steps. However, in addition to what one can expect to occur in shifting trips from one mode of travel to the other, there is a serious reason to expect that additional trips will be made in 1975 if the new system is in operation, because of the existence of the new system. What is introduced here is the possibility of realizable latent travel demand due to the characteristics of the new system. The essence of latent travel demand is that a real need and a desire to make a trip are surpressed because of the limitations imposed by the system on which the trip must be made. Thus, when a bus line stops operating after 10 p. m. transit trips that would be made at that time have to be postponed or altogether canceled if the trip-maker cannot make use of an automobile. The cost to the individual of the canceled or postponed trip cannot be easily conveyed but is nonetheless real. Environmental limitations or constraints are also equally frequent and real. Trips must be postponed, combined, or outright canceled if the trip-maker cannot bear the direct cost of the trip in money or time, regardless of any indirect benefits to the potential trip-maker. Thus, the need and desire for these trips constitute the latent travel demand for the given set of users and for the given set of circumstances. When the new system is introduced with its essentially different levels of availability, low cost, and high speeds, it is bound to make a proportion of the existing latent travel demand realizable and thus manifest for 1975. This is what other researchers some years ago called induced trips, i. e., trips that the new expressway systems were found to produce.

The methods of calculating latent travel demand are still in their infancy. They are usually based on measuring the elasticity of a demand function with respect to either the income of the consumer or the price of the commodity. A similar approach was followed in the minicar project (although the team avoided producing a mathematical solution by taking the partial derivative of the demand function with respect to the unit cost of the minicar trips). In general, the procedure followed was to "estimate" the elasticity of demand due to the price changes within the guidance we found in specifying the shift of trips from one mode of travel to the other due to price changes for manifest travel patterns. The forecast total travel demand was the summary of the 2 partial demand estimates, i. e., the manifest and latent travel demand of the study area.

SENSITIVITY OF DEMAND FORECASTS

The long sequence of steps required to produce a forecast for travel demand of a new urban system plus the extensive number of assumptions and presumptions that the analyst has to undertake in order to complete the missing pieces of information necessary to his work clearly indicate that at the end of the process he possesses an estimate with considerable limitation. This realization, together with the fact that a limited number of system setups tested represented only a small part of what can be devised and tried in the future, imposed a clear need to investigate and specify as clearly as possible the sensitivity of the forecast estimates to the several important system variables utilized in the process. This was exactly the concluding part of the minicar project forecasting effort.

Usually there are 2 types of variables: the variables that are endogenous to the new system (which are also under considerable design control) and the variables that are exogenous to the new system per se but endogenous to the study area. In the minicar project the endogenous variables were the minicar fare structure, the spacing of the terminals, and the conditions and delays of the vehicle pickup and delivery. Exogenous variables were the value of time, the automobile parking cost and availability, and the operating costs of automobiles and the fares of transit and taxis. The sensitivity analysis carried as part of this project imdicated that the forecast estimates were extremely sensitive to the endogenous and exogenous variables. Further, the type of trips attached to and generated by the new system proved also to be very sensitive to variations of many of the endogenous and exogenous variables. In fact, our analysis indicated that the flexible variables tended to be more important than the plain technological characteristics of the new system per se.

This sensitivity to system variables was considered to constitute both an important risk in introducing the new system and an important policy tool in influencing the success of the new system and directing the impact it may have on automobile trips, transit trips, and taxi services. In essence, extensive knowledge of what influences the new system would have, and of the way the various influences would be felt, was found to be an important element in managing such a system and in interweaving it within the fabric of other urban transportation systems and services.

CONCLUSIONS

The conclusions of this case study, above and beyond the specific findings of the minicar project, led the research team into a series of reappraisals of the entire effort. In many respects, the team realized time and again that the available methods and techniques in the field are indeed of limited ability in forecasting travel demands for new systems for use within modern urban areas. The need for relevant empirical findings was indeed pervasive. On occasion it seemed that nothing could make up for lack of data from direct experiment. Nonetheless, it was also felt in such early stages of research for a new urban transportation system even the most fundamental decisions (i. e., the most appropriate study area and the most pervasive technological characteristics of the system must be stressed) did not have sufficient grounds for a sound answer. Thus, the team concluded that the most productive approach would have been an interplay between office research and system simulation on the one hand and actual experimentation on the other.

Barring this interplay, the analyst would be advised to select an approach that utilized the most direct methods of analysis with the most feasible openness on the required assumptions and presumptions necessary to be made in the process of forecasting. The market analysis approach yielded limited results and helped primarily to increase the understanding of a few critical segments of the potential market. The product clinic was also helpful to a very limited extent for several reasons, among which the difficulties the team encountered in bringing in larger crowds and in communicating the whole purpose of the project were prominent. Concerns over "salesmanship," "profits," and "propriety" increased as the idea was transformed gradually from a pure speculative research notion into a physical, probable system. In overall terms, the transportation planning method that was based on past data from an origin and destination study and on findings of previous modal-split analysis proved to be the "saving method" that permitted the team to complete its efforts and to produce a set of forecasting estimates for limited use and some indication of their sensitivity to the important system variables. This was so because of the availability of several central concepts in this approach and the availability of a plethora of partially relevant data.

Among the numerous dilemmas that were faced in this effort, the ones related to the questions of Who will generate the system data? Who should make the critical assumptions? and What analytical tools should be selected? were the ones that permeated the effort from its beginning to its conclusion. It seems that these questions are bound to be in the center of concern of any similar effort. No hurried answers would be advisable to these questions because they are only partly technical. Their essential nature is social, political, managerial or financial or all of these. Thus, in most cases, the answers must come primarily from the sector that carried the responsibility of implementation or the group that would either be the potential users of the system or feel the impact of the system otherwise or both. The researcher or the planning analyst in these cases should be satisfied to take the back seat and not to make the numerous decisions that predetermine so much of the answer.

REFERENCES

1. Minicar Transit System. Univ. of Pennsylvania, Final Rept., 1970.
2. Minicar Transit System—Phase I: Feasibility Study. Univ. of Pennsylvania, 1968.

3. Tomazinis, A. R. Travel Demand Projection for a Minicar Fleet. Institute for Environmental Studies, Univ. of Pennsylvania, June 1968.
4. Tomazinis, A. R. A Minicar Transportation System in an Urban Low Income Area. Institute for Environmental Studies, Univ. of Pennsylvania, July 1970.
5. Quandt, R. E., and Baumol, W. J. The Demand for Abstract Transport Modes: Theory and Measurement. Jour. of Regional Science, No. 6, 1966, pp. 13-26.
6. Quandt, R. E., and Young, K. H. Cross Sectional Travel Demand Models: Estimates and Tests. Jour. of Regional Science, No. 9, 1969, pp. 201-214.
7. Wallace, J. P. Some Applications of Marketing Research Techniques to the New Mode Demand Forecasting Problem. In Methods and Concepts in Forecasting Travel Demand for Future Transportation Systems: Conference Proceedings, Transportation Studies Center, Univ. of Pennsylvania, 1972.

TRAFFIC ASSIGNMENT WITHOUT BIAS

Alvin H. Benesh, Federal Highway Administration, U.S. Department of Transportation, Pierre, South Dakota

Traffic assignment procedures based on shortest time path tend to overassign traffic to high-speed facilities. Because of this, 2 assignment formulas were developed that, taken as parameters in the assignment process, eliminate the overall bias. One formula, known as the time-distance factor, was described in an earlier publication. The formula introduced in this paper, called the exponential factor, is based on a combination of link distance and speed raised to an exponential power. Both were tested by using data from 9 different studies in 6 states. When assignments were based on time, the mean overassignment to the high-speed facility was 45 percent. At the optimum adjustment in the formulas this difference was eliminated, and the percent rms error was less than a third of the error in assignments by the shortest time path. A count of assignment errors in individual zone-to-zone movements revealed that in assignments by the shortest time path there were 18 times as many overassignment errors as underassignment errors. At the optimum assignments by the 2 formulas, the number of overassignment and underassignment errors was about in balance, and the total number of assignment errors was only about 60 percent of that based on assignments by shortest time. Also the traffic volumes assigned closely matched those actually counted.

•BASIC to any procedure for assigning traffic to a network by the shortest path is the development of a measure or parameter that will in general route a trip to the correct highway. If the measure used gives close to an optimum assignment, then various calibration and restraint procedures to improve the assignment further will involve only very minor changes in speed. On the other hand, if the measure of the spatial relation in itself has a strong bias, then the procedures will involve much larger adjustments to overcome the bias. The result will be a coarser assignment with regard to actual traffic and speeds.

In a previous paper (1) it was shown that the use of travel time for this measure results in too many trips being assigned to the high-speed facility. Conversely, the use of shortest distance gives too few trips to the high-speed facility.

One of the purposes of this paper is to document more thoroughly those statements by using a greatly expanded amount of data and new research techniques and then, by applying these statistical techniques, to ascertain the proportion of time and distance in a time-distance (TD) factor previously described (1) that will give the best assignment.

The principal purpose is the introduction of a new parameter, the exponential factor, for traffic assignment. It gives results quite similar to those obtained by use of the TD factor, but, because the formula involving the use of speed to an exponential power is completely different, it should prove of interest and eventually of considerable value to the transportation planning profession.

THE TD FACTOR

The formula for the link length for assignment purposes is based on a combination of time and distance, as follows:

$$\text{TD factor} = TP + D\,(1.00 - P)$$

where

T = time, min;
D = distance, miles; and
P = decimal fraction of weight assigned to time.

When $P = 1.0$, the assignment is based on travel time; when $P = 0$, the assignment is based on distance. Between these 2 extremes, any value of P can be tried until one is found that will give the best results as determined by the root mean square error for the entire network. Once a value of P is decided on, the same constant value of it is used for the entire network assignment.

THE EXPONENTIAL FACTOR

The formula for the link length for traffic assignment purposes is based on a combination of distance and speed raised to an exponential power, as follows:

$$\text{Exponential factor} = D(K/V)^N$$

where

D = link distance,
V = speed, and
K = arbitrarily selected constant.

It is believed that the value of the exponent N will never exceed the range of $N = 0$ to $N = 1$.

Suppose that V is expressed in miles per hour, D in miles, and $K = 60$. When $N = 1$, the expression becomes $60D/V$, which is the formula for travel time in minutes. When $N = 0$ the fraction $(K/V)^0$ becomes 1 and the formula ends with only the link distance.

For a sample computation of an exponential factor using an exponent of 0.5, assume that the times, distances, and resulting speeds between a pair of zones are as follows (1):

Facility	Time (min)	Distance (miles)	Speed (mph)
Freeway	12.5	8.3	39.8
Alternate	13.2	6.4	29.1

$$\text{Freeway exponential factor } (N = 0.5) = 8.3\ (60/39.8)^{0.5} = 10.20$$
$$\text{Alternate exponential factor } (N = 0.5) = 6.4\ (60/29.1)^{0.5} = 9.19$$

Because the alternate route has the smallest assignment factor, it will be assigned the trips.

The exponential factors for various values of N are shown in Figure 1. The data are plotted on full logarithmic scale so that the curves will be straight lines. Figure 2 shows a comparison of the exponential factor and the TD factor. The exponential factor with $N = \frac{1}{2}$ agrees quite closely with the TD factor using $\frac{1}{3}$ T and $\frac{2}{3}$ D. In the exponential factor curves, the constant K had been set at 60 in order to give an assignment factor of 1 at 60 mph. (With $N = 1$, this also gives the curve for travel time.) As assignments are based on the relative link lengths, K could just as well be 1. If we make $K = 1$, the exponential factor becomes simplified to the following:

$$\text{Simplified exponential factor} = D/V_N$$

SOURCES OF DATA

To test the formulas required the following for each zone-to-zone movement, which represents a group of trips, all originating in one zone and terminating in another: total number of trips, trips using freeway, travel time and distance via freeway, and travel time and distance via alternate route. Published information was available from

Figure 1. Comparison of exponential factors for values of N.

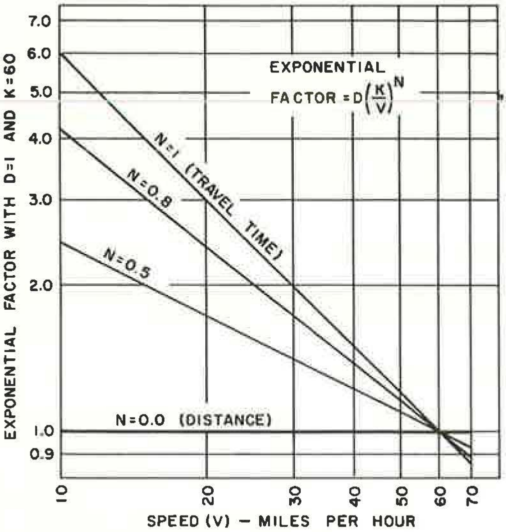
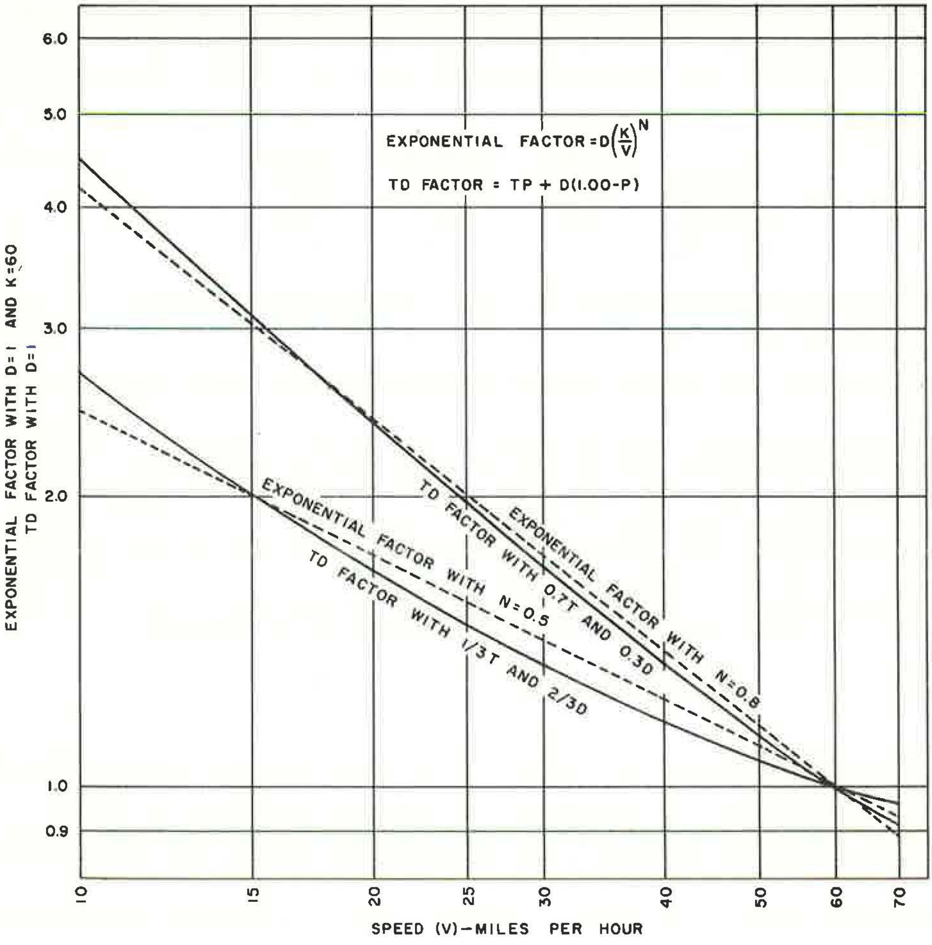


Table 1. Trips between zones and on facilities.

Facility	Trips Between Zones	Trips on Facility	Zone-to-Zone Movements
Urban freeway			
Gulf	22,556	8,413	19
Alvarado	92,278	23,856	154
Cabrillo	61,140	27,060	101
Oceanside	6,904	2,823	53
Shirley	19,756	8,152	88
Small city bypass			
Kokomo	5,526	2,053	69
Lebanon	6,110	4,464	50
Rural Freeway			
Kansas	6,054	3,967	46
South Dakota	7,388	4,388	88

Figure 2. Comparison of exponential and TD factors.



8 separate studies. In addition, an unpublished report on the Gulf Freeway in Houston was obtained from the Texas Highway Department. This made a total of 9 sources of data from 6 states, as follows:

- Urban Freeways
  - Gulf Freeway, Houston (2)
  - Alvarado Expressway, San Diego (3)
  - Cabrillo Freeway, San Diego (3)
  - Oceanside-Carlsbad Freeway, California (3)
  - Shirley Highway, Arlington and Fairfax Counties, Virginia (4)
- Small City Bypasses
  - Kokomo, Indiana (5)
  - Lebanon, Indiana (5)
- Rural Freeways
  - Interstate 70, Kansas (6)
  - Interstate 29, South Dakota (7)

METHOD OF ANALYSIS

Data from these studies were used to write a computer program that provided a tabulation of the following data for each study: Number of zone-to-zone movements, number of trips actually using freeway, number of trips assigned to freeway, percentage of zone-to-zone trips assigned to freeway, number of trips assigned to freeway as a percentage of actual trips using freeway, number of zone-to-zone movements erroneously overassigned to freeway, number of zone-to-zone movements erroneously underassigned to freeway, and total number of zone-to-zone movements erroneously assigned. The actual and assigned trips are self-explanatory. The erroneously assigned trips can be described as follows.

In an assignment by shortest path method, there are only 2 assignments possible. Either all zone-to-zone trips are assigned to the freeway, or no trips are assigned to it. For example, suppose that 49 percent of the trips represented in a zone-to-zone movement actually use the freeway. Then the best possible assignment to the freeway by this procedure would be 0 trips, and such an assignment would be considered correct. However, if all of the trips were assigned to the freeway, the assignment would be incorrect and the computer would classify the assignment as a plus error. Examples of the assignment classifications that are possible for the percentages of zone-to-zone trips actually using the freeway and assigned to the freeway are as follows:

Actual	Assigned	Classification
49 and under	0	Correct
49 and under	100	Plus error
51 and over	0	Minus error
51 and over	100	Correct

The assignment data were computed and tabulated in increments of 0.1 for the value of P in the TD factor and of N in the exponential factor. This gave 11 assignments for each factor for each study. In addition, in the range from 0.46 to 0.24, assignments were made in increments of 0.02, which improved accuracy in the optimum range.

ANALYSIS RESULTS

Total Traffic Assigned

As indicated in the earlier report on the TD factor, all assignments to the higher type of facility were too high by the shortest time path and too low by the shortest distance. The percentages for each of the 9 highways are given in Tables 2 and 3. The percentages for each highway at the optimum exponential factor and TD factor for the group are also given. Additional detail on number of trips assigned is given in Table 1. The tables give assignments at intervals of 0.2 in values of N and P, plus the

Figure 3. Percentage of actual trips assigned by exponential factor for values of N in intervals of 0.1.

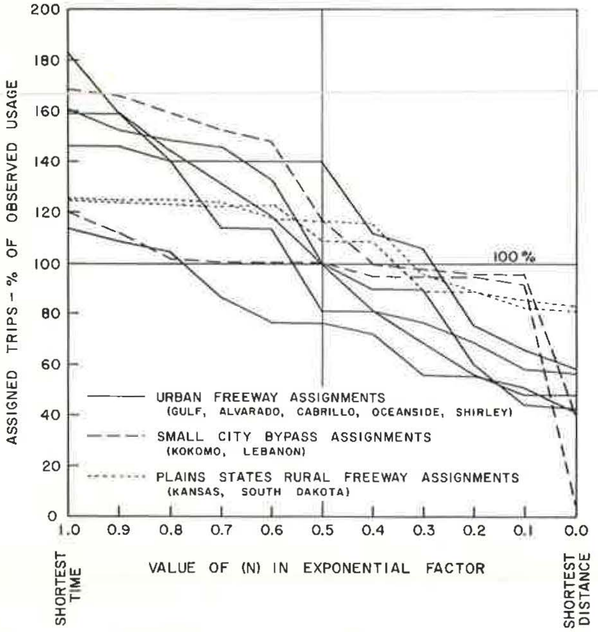


Figure 4. Percentage of actual trips assigned by TD factor for values of P in intervals of 0.1.

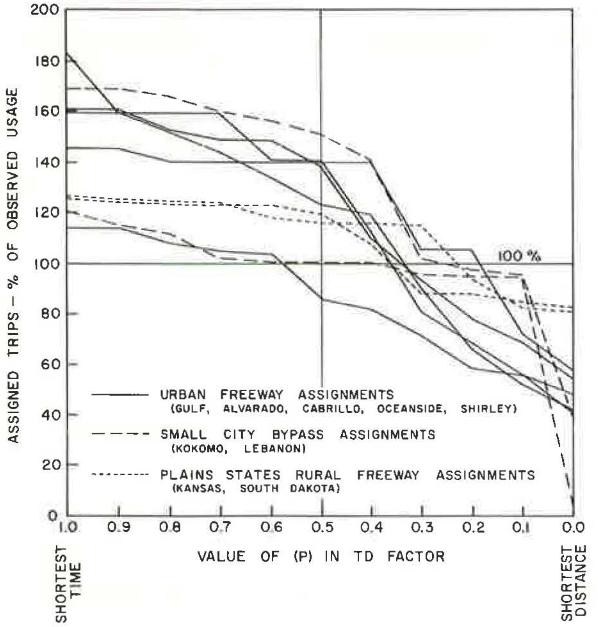


Table 2. Percentage of actual trips assigned by exponential factor for values of N in intervals of 0.2.

Facility	1.0[a]	0.8	0.6	0.46[b]	0.4	0.2	0.0[c]
Urban freeway							
Gulf	159	141	114	81	81	56	41
Alvarado	183	144	119	101	90	60	42
Cabrillo	161	149	133	94	80	69	54
Oceanside	146	140	140	115	113	75	58
Shirley	114	105	75	72	71	55	48
Small city bypass							
Kokomo	169	159	148	104	100	96	39
Lebanon	120	102	101	96	95	95	4
Rural freeway							
Kansas	125	124	123	108	108	89	82
South Dakota	126	125	118	117	116	90	81
Mean	145	132	119	99	95	76	50

[a] Shortest time. [b] Optimum. [c] Shortest distance.

Table 3. Percentage of actual trips assigned by T factor for values of P in intervals of 0.2.

Facility	1.0[a]	0.8	0.6	0.4	0.32	0.2	0.0[b]
Urban freeway							
Gulf	159	159	141	114	81	69	41
Alvarado	183	152	134	119	90	66	42
Cabrillo	161	153	148	111	97	77	54
Oceanside	146	140	140	140	113	106	58
Shirley	114	109	104	82	72	58	48
Small city bypass							
Kokomo	169	166	156	141	115	98	39
Lebanon	120	112	101	101	101	95	4
Rural freeway							
Kansas	125	124	123	108	107	89	82
South Dakota	126	125	118	117	115	94	81
Mean	145	138	129	115	95	84	50

[a] Shortest time. [b] Shortest distance.

assignments at the optimum values of N and P. Computations, however, were made at intervals of 0.1, and plots of the assignment data for the exponential factor and the TD factor on this basis are shown in Figures 3 and 4. The number of lines, one for each study, tends to slightly obscure the basic trend; therefore, Figures 5 and 6 show the mean values with boundary lines at 1 standard deviation from the mean. Assignments of 99 percent of the observed number of trips are obtained at a value of N = 0.46 in the exponential factor and P = 0.32 in the TD factor. As N and P were computed at intervals of 0.02 in the optimum range, the 99 percent figure is the closest one based on an actual assignment. At N and P values that give an average assignment close to the actual volumes, the dispersion about the mean, as measured by the standard deviation, is about a third less than the dispersion when assignments are by the shortest time path.

Root Mean Square Error

The formula for the computation of the rms error is as follows:

$$\text{rms} = \sqrt{[\Sigma(X_{GC} - X_A)^2]/(N - 1)}$$

where

X_{GC} = total vehicles counted on a given route,
X_A = assigned volume to the route, and
N = number of routes.

It is possible to interpret the results of the computation given above better in this case if the reader can review a sample computation. The following one is for shortest time path.

Facility	Facility Type	Actual Trips (percent)	Assigned Trips (percent)	Difference	Difference^2
Gulf	Freeway	37	59	-22	484
	Alternate	63	41	+22	484
Alvarado	Freeway	26	47	-21	441
	Alternate	74	53	+21	441
Cabrillo	Freeway	44	71	-27	729
	Alternate	56	29	+27	729
Oceanside	Freeway	41	60	-19	361
	Alternate	59	40	+19	361
Shirley	Freeway	41	47	-6	36
	Alternate	59	53	+6	36
Kokomo	Bypass	37	63	-26	676
	Alternate	63	37	+26	676
Lebanon	Bypass	73	87	-14	196
	Alternate	27	13	+14	196
Kansas	Freeway	66	82	-16	256
	Alternate	34	18	+16	256
South Dakota	Freeway	59	75	-16	256
	Alternate	41	25	+16	256
	Total	900	900	0	6,870

Total alternates and freeways = 18
Average trips on each, percent = 900/18 = 50
rms error $= \sqrt{6{,}870/17} = 20.1$
Percent rms error = (20.1 x 100)/50 = 40.2

Figure 5. Mean percentage of actual trips assigned by exponential factor for values of N in intervals of 0.1.

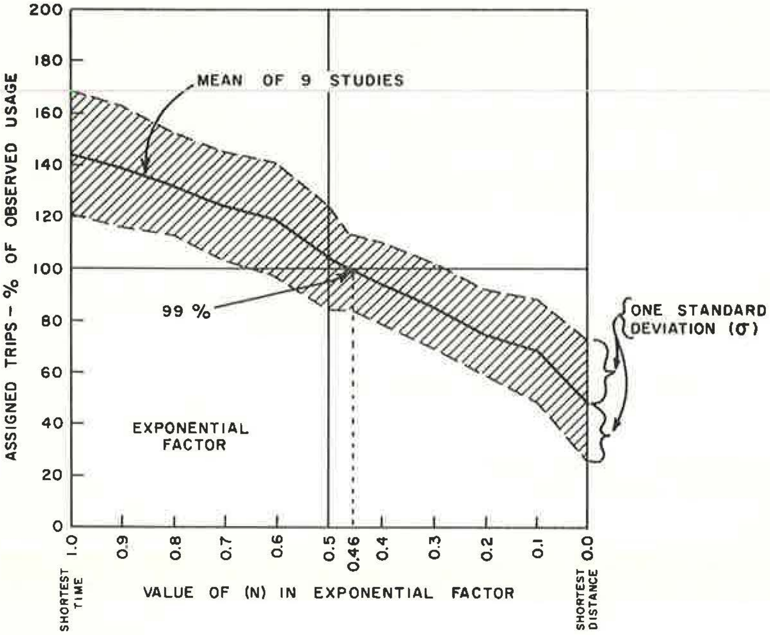


Figure 6. Mean percentage of actual trips assigned by TD factor for values of P in intervals of 0.1.

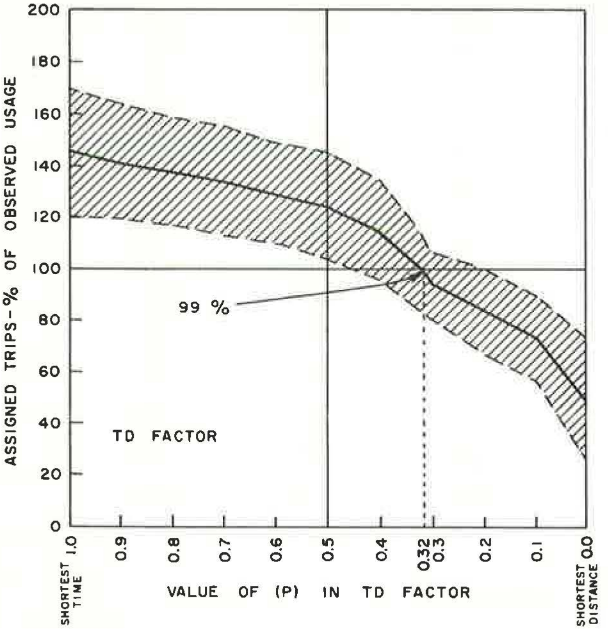


Figure 7 shows rms errors for the exponential factor and TD factor. In the optimum range for each method, the percent rms error is less than a third of the error in assignments by the shortest time path.

Movements Incorrectly Assigned

This analysis is probably the most significant part of the report. Results based on route totals alone do not take into account the variation in accuracy that is possible from one end of the route to the other. A count of the number of zone-to-zone movements incorrectly assigned, classified according to the number of plus errors and minus errors, gives a better insight into the workings of an assignment procedure. Of course, in an all-or-nothing assignment there are only 2 possibilities: either 0 percent or 100 percent of the trips represented in a zone-to-zone movement are assigned to the facility. In this analysis, therefore, the assignment that comes closest to the actual is considered the correct one, even though numerically it may be considerably different from the actual. Thus, if 37 percent of trips represented in a zone-to-zone movement actually uses a freeway, an assignment of 0 percent would be correct, and an assignment of 100 percent would represent a plus error. If 80 percent of the trips in a zone-to-zone movement actually uses a freeway, an assignment of 0 percent would represent a minus error, and an assignment of 100 percent would be considered correct.

Tables 4 and 5 give the percentage of plus and minus errors for different values of N and P for each of the 9 facilities in this study. In assignments by shortest time nearly all of the errors are plus errors, and in assignments by shortest distance most of the errors are minus errors except on 1 facility. Stated in another way, in assignments by shortest time there are 18 times as many plus errors as minus errors, and in assignments by shortest distance there are 8 times as many minus errors as plus errors. These facts are considered by the author to be of great statistical significance. At some point between the 2 extremes, where the number of plus and minus errors is about in balance, the total number of errors averages only about 60 percent of that based on assignment by shortest time and the traffic volumes assigned closely match the actual traffic counts. The trend in percentage of assignment errors for various values of N and P is shown in Figures 9 and 10. Figure 11 shows a comparison of assignments by shortest time and shortest distance and the optimum exponential and TD factors. For the latter two, the volumes assigned were each 99 percent of the actual number of trips.

APPLICATION TO NETWORK ASSIGNMENT

Most computer programs are written for assignments by shortest time path. All that would be necessary for these procedures to be applied would be to substitute the formula for exponential factor or TD factor in the program for the computation for travel time in determining the link spacing for traffic assignment. The program should be written so that the value of N or P could be easily changed between assignment runs to get the best assignment.

There is in existence at the present time a program for the TD factor. It was prepared in 1970 and is now a part of the standard FHWA urban transportation program system battery for the IBM 360 computer. The TD factor, in principle, will be found in the instructions for the weight card (WEIGHT) in the program deck BUILDVN, although the TD factor as such is not mentioned. It can be developed by assigning a weight to time and to distance in accordance with the instructions for using the card.

The computation of network root mean square error is now available in assignment programs. This measure would appear to be a logical one in determining the optimum value of N or P for the best assignment. After the optimum assignment by this procedure is reached, some further improvement in the calibration is possible by an iterative process. However, because assigned traffic volumes will probably already be close to the actual, it would appear that the size of subsequent speed adjustments per iteration should be set somewhat lower than usual when assignments are made on the basis of travel time.

Figure 7. Root mean square error for exponential and TD factors for values of N and P.

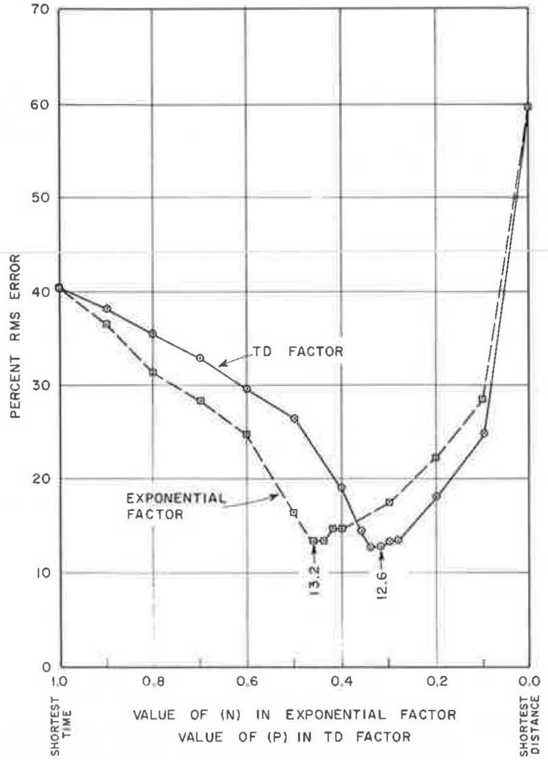


Figure 8. Percentage of zone-to-zone movements incorrectly assigned by exponential factor for values of N.

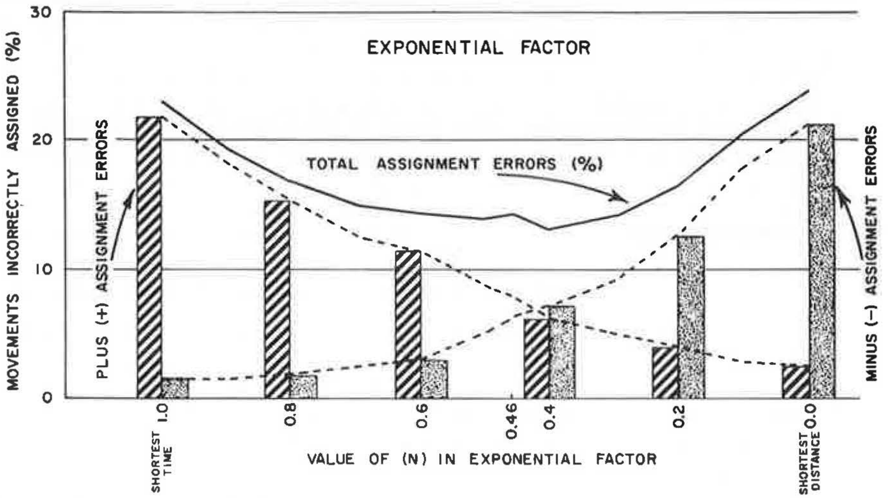


Table 4. Percentage of zone-to-zone movements incorrectly assigned by exponential factor for values of N.

	1.0[a]		0.8		0.6		0.46[b]		0.4		0.2		0.0[c]	
Facility	+	-	+	-	+	-	+	-	+	-	+	-	+	-
Urban freeway														
Gulf	16	0	11	0	5	0	0	5	0	5	0	21	0	37
Alvarado	28	1	20	2	13	2	12	4	7	5	3	10	1	17
Cabrillo	24	1	20	1	18	4	13	14	7	14	7	20	5	29
Oceanside	15	2	11	2	11	2	5	4	6	5	0	13	0	28
Shirley	9	1	7	2	3	9	3	11	3	12	2	21	2	25
Small city bypass														
Kokomo	36	0	29	0	22	0	13	1	11	1	9	3	10	9
Lebanon	38	4	12	6	8	6	8	10	4	10	4	10	0	22
Rural freeway														
Kansas	9	0	7	0	7	2	4	7	4	7	2	9	0	15
South Dakota	21	2	20	2	15	2	14	2	13	5	9	7	7	11
Mean	21.8	1.2	15.2	1.7	11.3	3.0	8.0	6.4	6.1	7.1	4.0	12.7	2.8	21.4
Total	23.0		16.9		14.3		14.4		13.2		16.7		24.2	

[a] Shortest time. [b] Optimum. [c] Shortest distance.

Figure 9. Percentage of zone-to-zone movements incorrectly assigned by TD factor for values of P.

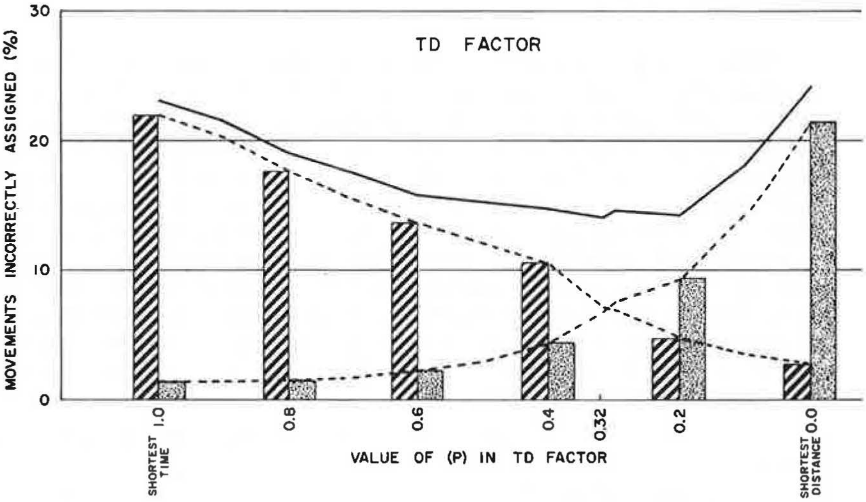
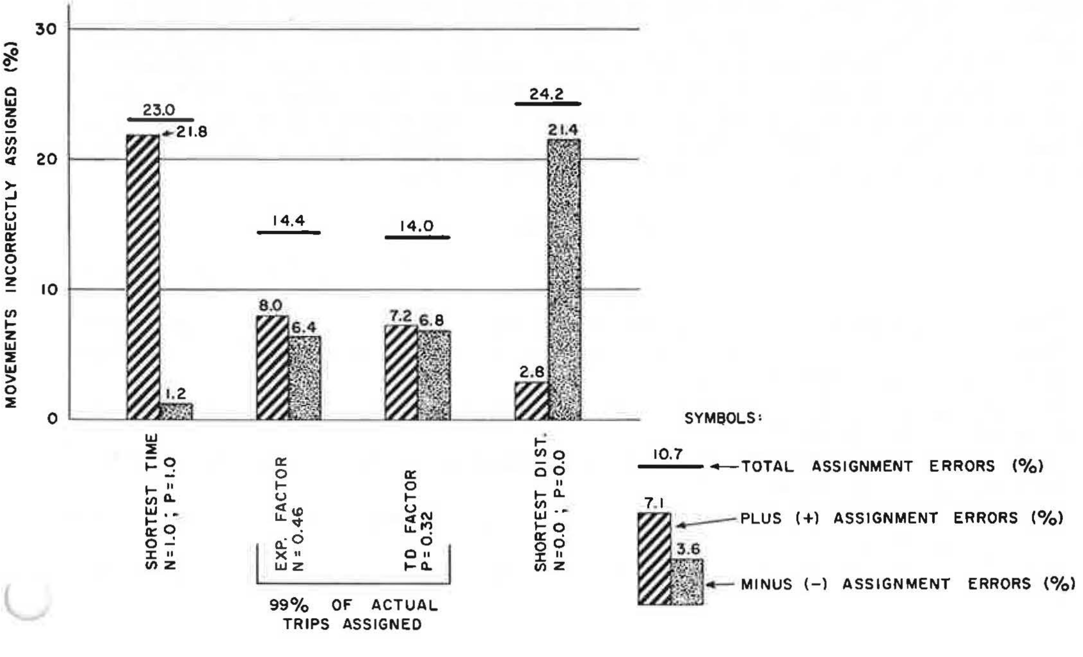


Table 5. Percentage of zone-to-zone movements incorrectly assigned by TD factor for values of P.

	1.0[a]		0.8		0.6		0.4		0.32[b]		0.2		0.0[c]	
Facility	+	-	+	-	+	-	+	-	+	-	+	-	+	-
Urban freeway														
Gulf	16	0	16	0	11	0	5	0	0	5	0	11	0	37
Alvarado	28	0	23	2	13	2	12	4	7	5	3	10	1	17
Cabrillo	24	1	22	1	19	1	17	10	13	13	7	16	5	29
Oceanside	15	2	11	2	11	2	11	2	6	6	6	7	0	28
Shirley	9	1	7	1	7	4	3	8	3	11	2	16	2	25
Small city bypass														
Kokomo	36	0	35	0	26	0	21	1	13	1	9	1	10	9
Lebanon	38	4	18	4	8	6	8	6	8	6	4	10	0	22
Rural freeway														
Kansas	9	0	7	0	7	2	4	7	4	9	2	9	0	15
South Dakota	21	2	20	2	15	2	14	3	12	5	9	6	7	11
Mean	21.8	1.2	17.7	1.3	13.7	2.1	10.6	4.3	7.2	6.8	4.9	9.3	2.8	21.4
Total	23.0		19.0		15.8		14.9		14.0		14.2		24.2	

[a] Shortest time. [b] Optimum. [c] Shortest distance.

Figure 10. Percentage of zone-to-zone movements incorrectly assigned for selected values of N and P.



Before adopting these procedures, the planner may wish to run some test assignments using the exponential factor or TD factor. In doing so, he will probably select for the tests some readily available input data from a network previously assigned. In this event, he is cautioned to be sure that the speeds used are objectively measured speeds that have not been revised or adjusted prior to or as a part of the previous assignment runs. A number of planners realize that when assignments are by shortest time, the measured speeds give too high an assignment to high-speed facilities. As a result some of them may have either subconsciously or systematically reduced input speeds of the high-speed facilities relative to those of low speed. This may even have been done by changing the definition of the speed measurement for each type of facility. If such an adjustment in input has been made, the exponential factor or TD factor will not give satisfactory results. Some typical changes that may have been made that the planner should look for are

1. Selective manual speed reductions on some high-speed facilities, usually freeways;
2. Use of peak-hour speeds on freeways and off-peak speeds on other facilities;
3. Manual upward adjustment of arterial speeds in the CBD; and
4. Use of average speeds on rural freeways and 85 percentile speeds on other rural arterials.

The discussion given above has related strictly to assignment of trips, without regard to how the trips are distributed. The distribution could be the origin-destination data from the original survey, or it could be based on a gravity or other model. If the distribution is by a gravity model, some error may be introduced by using travel time as a basis. Logically, better results would be possible by using the optimum TD or exponential factor in the model instead of travel time.

When the exponential factor is compared with the TD factor, it is apparent that although both give strikingly better assignments than that given by the shortest time path the difference between the optimum assignment by each is not great enough to be conclusive. Either can be used with confidence, and perhaps after several full-scale assignments on complete networks a preference will gradually develop for one or the other.

ACKNOWLEDGMENT

The opinions, findings, and conclusions expressed in this paper are those of the author and not necessarily those of the Federal Highway Administration. Sincere appreciation is extended to the Research and Planning Division of the South Dakota Department of Highways for preparing the charts and for excellent cooperation in providing computer services through the Computer Services Division. Also I wish to thank Charles W. Chappell, Jr., of the Federal Highway Administration for his unusually conscientious work in writing the computer program and following through on the task to the final computer tabulations. Acknowledgment is also made to the Texas Highway Department for furnishing the Gulf Freeway data.

REFERENCES

1. Benesh, A. H. Traffic Assignment by the Shortest Path Method Using the TD Factor. Traffic Quarterly, Oct. 1967, pp. 553-567.
2. Traffic Assignment to the Gulf Freeway. Texas Highway Department, Dec. 1954.
3. Moskowitz, K. California Method of Assigning Diverted Traffic to Proposed Freeways. HRB Bull. 130, 1956, pp. 1-26.
4. Trueblood, D. L. Effect of Travel Time and Distance on Freeway Usage. HRB Bull. 61, 1952, pp. 18-37.
5. May, A. D., Jr., and Michael, H. L. Allocation of Traffic to Bypasses. HRB Bull. 61, 1952, pp. 38-58.
6. Kansas Interstate 70 Usage Study. State Highway Commission of Kansas, Summary Rept., 1962-1965.
7. South Dakota Interstate Highway 29 Usage Study. South Dakota Department of Highways, 1970.

SEGMENTED, MULTIMODAL, INTERCITY PASSENGER DEMAND MODEL

John W. Billheimer, Stanford Research Institute, Menlo Park, California

This report documents the specification, calibration, and performance of a segmented mathematical model developed to predict intercity travel by mode within and around the state of Michigan. The performance of different existing demand models is studied, and a model is formulated that uses income data and cost, time, and frequency of modal service to predict the travel volumes linking a wide range of city sizes. Model parameters are estimated by using a constrained-search calibration technique. Model performance is documented, and the sensitivity of this performance to changes in input data and calibration parameters is discussed.

•MATHEMATICAL models developed to predict intercity passenger demand have typically focused on travel between densely populated urban areas. Models developed and calibrated in this fashion generally fare poorly in predicting travel demand between sparsely populated areas and between a large city and a smaller urban area. Yet, statewide transportation planning entails the forecasting of traffic between cities of any size. This report describes the development of a segmented mathematical model designed to represent the demand for travel between cities of any size in Michigan.

MODEL SPECIFICATION

Range of Performance

Michigan covers an area of 58,216 square miles. Its populated areas range in size and type from isolated rural hamlets to the heavily industrialized Detroit area. An intercity passenger demand model designed for use in the state must be capable of predicting traffic by mode between cities of widely varying population densities separated by distances ranging from 50 to 600 miles. To assist in the formulation and calibration of a model having this capability, a set of 15 sample origin-destination pairs was selected to reflect the range of population-distance combinations existing in and around Michigan. These sample pairs are given in Table 1. The designations given in Table 1 are as follows:

Item	Designation
Population	
20,000 to 50,000	Small
200,000 to 1,000,000	Medium
2,000,000 to 5,000,000	Large
Distance, miles	
50 to 100	Short
150 to 250	Medium
400 to 600	Long
City	
Alpena	ALP
Sault Sainte Marie	SSM
Houghton	HOU
Flint	FLI
Caro	CAR

Item	Designation
City	
Detroit	DET
Kalamazoo	KAL
Columbus (Ohio)	COL
Des Moines	DES
Chicago	CHI
Philadelphia	PHL

Performance of Existing Models

As a first step toward selecting a model for use in predicting Michigan intercity traffic, the 15 sample origin-destination pairs were used to test several existing unimodal and multimodal demand models. A list of the models is given in Table 2. Details regarding the structure of the models may be found in the indicated references. All of the models represent traffic between zones as a function of trip generation characteristics and some measure of interzonal impedance. As such, all of the models are descendants of the gravity model.

Table 3 gives the results of using each of the demand models to predict traffic between the city pairs. This table also gives actual measurements of intercity automobile traffic as compiled by the Michigan Department of State Highways in studies conducted between 1964 and 1968. A comparison of predicted and observed values shows that existing models generally cannot cope with the range of city sizes and distances to be found in Michigan. As might be expected, model 4, developed and calibrated specifically for use in the state, does the best job of reproducing actual traffic measurements, although it appears to overstate traffic between cities separated by short distances. Unfortunately, this model is limited to the prediction of automobile traffic.

The performance of the remaining models hardly can be termed promising. Models 1 and 2, both classic gravity models, perform poorly when uprooted from their places of calibration and applied to the range of city pairs existing in Michigan. Model 3, which woefully understates traffic, might profit slightly from a change of coefficients. Model 5, calibrated on the large cities of the Northeast Corridor, performs poorly in estimating traffic between the smaller Michigan cities.

The discrepancies between predicted and actual values given in Table 3 highlight the difficulty of predicting intercity passenger demand for a wide range of distances and city sizes. This difficulty is multiplied by the problem of designating modal preferences of passengers. No gravity model exists that can be pulled off the shelf and used with confidence to predict travel patterns in any arbitrary area. In this sense, the term gravity model, implying as it does an immutable law, is a misnomer. Isaac Newton himself might have had second thoughts about the validity of his gravity model had it been necessary to reformulate it for different masses and recalibrate it for different points on the earth's surface.

Model Selection

The ability to reproduce observed travel data with a reasonable degree of fidelity for the range of city sizes and separations encountered in Michigan was but one of the criteria considered in selecting an intercity demand model. In addition to this important consideration, it was desired that the model have the following attributes: simplicity, sound theoretical structure, and ability to reflect the intermodal consequences of system changes. Because each of the tested models failed to reproduce modal preferences for the range of sample city sizes, the selection process centered on these additional attributes. Once a model having these attributes was identified, an attempt was made to extend its range of applicability to include the city sizes of interest in Michigan.

A review of existing intercity demand models led to the selection of the basic model developed by McLynn (5), modified by the National Bureau of Standards (6), and summarized by the Northeast Corridor Transportation Project (7). The variables con-

Table 1. Sample origin-destination city pairs.

	Distance		
City Size	Short	Medium	Long
Small to small	ALP-SSM	SSM-HOU	
Small to medium	ALP-FLI	SSM-FLI	
Small to large	CAR-DET	SSM-DET	HOU-DET
Medium to medium	FLI-KAL	FLI-COL	FLI-DES
Medium to large	FLI-DET	FLI-CHI	FLI-PHL
Large to large		DET-CHI	DET-PHL

Table 2. Models used to test city pairs.

Model	Developer	Calibration Area	Modal-Split Capability	Reference
1	Unknown	Detroit	Single	1
2	Stanford Research Institute	California	Multiple	2
3	Wilbur Smith	Illinois	Single	3
4	Arthur D. Little	Michigan	Single	4
5	Office of High Speed Ground Transportation, U. S. Department of Transportation	Northeast Corridor	Multiple	5

Table 3. Demand model predictions of daily 1-way person trips by automobile.

City Pair	Model 1	Model 2	Model 3	Model 4	Model 5	Actual Traffic
ALP-SSM	1	2	1	22	0	19
SSM-HOU	0	1	0	7	0	11
ALP-FLI	4	17	2	29	1	27
SSM-FLI	1	6	0	14	1	51
CAR-DET	280	1,837	138	910	51	660
SSM-DET	8	46	0	62	4	274
HOU-DET	3	15	0	60	1	62
FLI-KAL	56	327	31	78	29	58
FLI-COL	37	231	1	n. a.	33	16
FLI-DES	2	8	0	n. a.	2	3
FLI-DET	7,021	59,378	1,302	24,859	2,877	14,600
FLI-CHI	262	2,032	1	127	148	77
FLI-PHL	22	151	0	n. a.	14	5
DET-CHI	2,635	25,732	2	597	1,391	775
DET-PHL	279	2,461	0	n. a.	161	74

Note: 1960 population data were used.

sidered by the model in determining the traffic by a mode m between origin-destination pair (i, j) are identified as follows:

t_m = total (i to j) travel time for the m th mode, hours;
c_m = total (i to j) out-of-pocket per capita cost, dollars;
f_m = frequency of (i to j) service, trips per day; and
F = number of families with annual incomes exceeding \$10,000 (families $\times 10^{-5}$) in the SMSA or county of the origin or destination city.

These variables can be used to define the modified demand model by the following relations:

$$w_m = \begin{cases} a_m t_m^{\alpha(1)} \; c_m^{\alpha(2)} \; [1 - \exp(-Kf_m)]^{\alpha(3)} & \text{for } m \neq \text{automobile} \\ t_m^{\alpha(4)} \; (c_m/1.7)^{\alpha(5)} & \text{for } m = \text{automobile} \end{cases} \tag{1}$$

$$W = \sum_m w_m \tag{2}$$

$$D = \begin{cases} \beta(0)(F_i F_j)^{\beta(1)} \; W^{\beta(2)} & \text{for } F_i F_j > G \\ \beta'(0)(F_i F_j)^{\beta'(1)} \; W^{\beta(2)} & \text{for } F_i F_j \leq G \end{cases} \tag{3}$$

$$D_m = Dw_m/W \tag{4}$$

The terms w_m and W may be regarded as modal conductance and total (i to j) conductance respectively. D_m and D are daily one-directional modal (i to j) demand and total (i to j) demand respectively (measured in persons).

So that the model could be adapted to the wide range of city sizes of interest in Michigan transportation studies, the demand model was segmented as indicated in Eq. 3. Thus, city pairs having population products $F_i F_j$ below a specified value G received a treatment different from that received by city pairs having larger population products.

Data Development

In the calibration of the demand model, numerical values were assigned to each of the model's parameters, and the effect of each assigned value on the model's ability to reproduce actual travel data was observed. The basic demand data used in this calibration process consisted of observed 1-way travel volumes by air, rail, bus, and automobile between each of 20 origin-destination pairs for the base year 1967. The 20 origin-destination pairs were the 15 city pairs given in Table 1 and the following 5 additional pairs: Detroit-Cleveland (CLE), Detroit-Pittsburgh (PIT), Detroit-Milwaukee (MIL), Flint-Cleveland, and Flint-Milwaukee. These pairs were added to broaden the data base and to place additional emphasis on travel between larger cities.

The cost, time, and frequency of common carrier service between each pair of cities were obtained from published schedules, and access times and costs were computed for each city. In the calculation of automobile costs and times, operating costs of 4 cents/mile were assumed and average speeds of 60, 30, and 15 mph were associated with freeways, arterials, and local streets. An average automobile occupancy of 1.7 persons/vehicle was assumed. Census data from 1960 were extrapolated to 1967 in the estimation of the number of families in each origin or destination zone having a real income exceeding \$10,000/year.

Calibration Technique

Attempts to use a series of log-linear regression analyses to calibrate the model formulated in Eqs. 1 through 4 proved unsuccessful. Part of the explanation for this lack of success may be traced to the failure of the log-linear regression format to deal adequately with the range of city sizes under consideration.

In lieu of regression analysis, the Michigan intercity passenger demand model was calibrated by means of a constrained-search technique. Through a combination of past experience and a knowledge of the model's structure, upper and lower bounds were set on acceptable values of each model parameter. A limited search was undertaken within those constraints for the combination of parameters that minimized a series of error functions describing model performance. The parameter bounds and error functions used in this constrained-search calibration process are described below.

Parameter Bounds—The following logical bounds were imposed on the model parameters in advance of the calibration process:

$$0 \leq \beta'(0) \leq \beta(0)$$

$$0 \leq \beta'(1) \leq \beta(1) \leq 1.1$$

$$0 \leq \beta(2) \leq 1$$

$$-5 \leq \alpha(j) \leq 0;\ j = 1,\ 2,\ 4,\ 5$$

$$\alpha(3) = 0.3247$$

$$K = 0.12$$

$$0 \leq a_m \leq 5$$

The model's consistency of behavior was ensured by imposing a positive or a negative constraint on each parameter. In addition, the positively constrained parameters $\beta'(1)$, $\beta(1)$, and $\beta(2)$ each had logical upper bounds. Experience with gravity models has shown that the exponent $\beta(1)$ associated with the population product rarely exceeds 1.1. Were this exponent to be higher, population increases would have a disproportionate effect on predicted travel demand. Furthermore, the exponent $\beta'(1)$ associated with small-city pairs cannot exceed the large-city exponent $\beta(1)$. This relation is indicated by empirical data relating intercity travel to population product for the sample city pairs.

Consideration of the conductance exponent $\beta(2)$ shows that the value of this exponent cannot exceed unity. Otherwise, a decrease in the time or cost of travel by 1 mode could cause corresponding increases in travel by competing modes. This can be shown by considering that, for small changes in time or cost, demand changes may be expressed as a function of the partial derivative of demand with respect to the changing variable. If the cost c_m of travel by mode m between 2 cities were to be changed, the effect on a competing mode n can be represented as follows:

$$\begin{aligned} \Delta D_n &= (\partial D_n/\partial c_m)\Delta c_m \\ \Delta D_n &= [\alpha(2)/c_m]\,[\beta(2) - 1]\,D_n(w_m/W)\Delta c_m \end{aligned} \qquad (5)$$

where ΔD_n represents a small change in demand D_n for a competing mode n, and Δc_m represents a small change in cost c_m. Thus, the intermodal effects predicted by the proposed demand model will remain consistent only as long as $\beta(2) \leq 1$.

Model consistency also demands that the modal conductance exponents $\alpha(1)$, $\alpha(2)$, $\alpha(4)$, and $\alpha(5)$, associated with time and cost, be negative. If these exponents are allowed to become too large, however, small changes of time or cost will have a disproportionate effect on demand. If $\alpha(2) \leq -5$, for example, sensitivity analysis shows that a 10 percent decrease in the cost of mode m could cause more than a 50 percent increase in the demand for that mode. Accordingly, a lower limit of -5 was placed on exponents $\alpha(1)$, $\alpha(2)$, $\alpha(3)$, and $\alpha(4)$ to forestall such unlikely results.

In calibrating his basic demand model, McLynn (5) empirically set K = 0.12. This value was used in the segmented model, as was the McLynn-calibrated value $\alpha(3) = 0.3247$. An upper limit of 5 was placed on the common carrier conductance multiplier a_m, because it was felt that larger values of a_m would create unrealistic imbalances between common carrier traffic and automobile traffic.

Error Functions—As the parameter values were varied within the established bounds, different error functions were computed and monitored to determine the overall effect of each parameter on the demand model's ability to reproduce observed travel data. These error functions reflected (a) the square root of the sum of the squares of the differences between calculated and observed modal travel values; (b) the sum of the absolute values of the difference between calculated and observed modal travel values; and (c) the number of calculated travel values that fell outside a predetermined range surrounding the observed value. Range settings within 10, 25, and 50 percent of the observed demand were monitored in the calibration process.

Calibration Procedures—In the calibration of the segmented demand model, attention was first directed toward the determination of the parameters $\beta(0)$ and $\beta(1)$, which were associated with larger-city pairs. Once these parameters were fixed, the search for $\beta'(0)$ and $\beta'(1)$ was undertaken. In the case of larger-city pairs, the constrained-search calibration procedure followed the steps outlined below.

1. Set $\beta(0) = a_m = 1$;
2. Select values for $\beta(1)$ and $\beta(2)$;
3. Select values for $\alpha(1)$, $\alpha(2)$, $\alpha(4)$, and $\alpha(5)$;
4. Compute D_{mj} for each city pair;
5. Compute error functions;
6. Adjust a_m to approximate modal-split proportions;
7. Adjust $\beta(0)$ to minimize error functions; and
8. Return to step 3 and try another combination of $\alpha(i)$, repeat until no further improvement in the error functions appears possible for the combination of $\beta(1)$ and $\beta(2)$ selected in step 2, and then try another combination of $\beta(1)$ and $\beta(2)$.

In the actual calibration process, $\beta(1)$ and $\beta(2)$ were varied in increments of 0.1 until a combination was found that appeared to fit the observed travel data associated with large-city pairs. At this point, $\beta(0)$, $\beta(1)$, and $\beta(2)$ were fixed, and a search was undertaken for appropriate values of $\beta'(0)$ and $\beta'(1)$.

Calibration Results

The calibration process given above resulted in the identification of the following parameter values:

$$a_m = \begin{cases} 1.5 & \text{for } m = \text{air} \\ 0.75 & \text{for } m = \text{bus, rail} \end{cases}$$

$$\alpha(1) = \alpha(2) = -1.5$$

$$\alpha(3) = 0.3247;\ K = 0.12$$

$$\alpha(4) = \alpha(5) = -1.8$$

$$\beta(0) = 25{,}000;\ \beta'(0) = 2{,}500$$

$$\beta(1) = 1.0;\ \beta'(1) = 0.1$$

$$\beta(2) = 0.9$$

$$G = 0.075$$

Table 4 gives a comparison of the demand calculated through the use of the parameters and the observed travel between each of the 20 city pairs. Although the overall agreement between calculated and observed values is satisfactory, the demand model severely understates travel between a city pair consisting of 1 small city and 1 large city. The reason for this understatement is shown clearly in Figure 1, where normalized demand is plotted as a function of population product. Normalized demand is defined as follows:

$$D_{normalized} = D/W^{\beta(2)}$$

This normalization process removes the effect of travel impedance from the demand term so that the resulting normalized demand should be a piece-wise log-linear function of the trip attraction measure, the income product F_iF_j. Figure 1 shows that the normalized demand between all city pairs except Sault Sainte Marie-Detroit, Houghton-Detroit, Sault Sainte Marie-Flint, and Caro-Detroit clusters closely about the log-linear form defined by the calibration process. It would appear to be impossible to use the chosen model effectively to represent travel between each of these 4 city pairs without destroying the model's ability to reproduce the remainder of Michigan's intercity traffic. There seems to be nothing within the framework of the mathematical model to explain, for instance, why automobile traffic between Detroit and Sault Sainte Marie should be nearly double the combined traffic between Detroit and the larger, closer cities of Pittsburgh and Milwaukee.

In addition to highlighting data inconsistencies, Figure 1 clearly shows the need for segmenting the Michigan intercity demand model. The data points plotted in this figure make it plain that a single log-linear function cannot reflect travel demand between city pairs of all sizes.

MODEL SENSITIVITY

Effect of Variable Changes

One test of the soundness of a demand model is its ability to behave logically in the face of changes in input variables. Because the Michigan intercity passenger demand model is a closed-form mathematical expression, its sensitivity to variable changes may be determined analytically. The first partial derivative of demand with respect to each input variable, $\partial D/\partial V$, provides a measure of this sensitivity and, by inference, also provides a measure of the impact of each variable on intercity demand.

The value of $\partial D/\partial V$ associated with each model input variable was computed and used to assess the effects of small (10 percent) changes in each variable on model demand and total intercity travel. Table 5 gives the results of this assessment. For large-city pairs, a 10 percent increase in the number of families in 1 city earning more than $10,000/year will increase travel demand by 10 percent across all modes. For small cities, an equivalent percentage increase will result in only a 1 percent increase in total travel. Although these differences in the modeled effect of population changes may be valid for extremely large cities and extremely small cities, it is illogical to expect such dichotomous behavior in the case of medium-sized cities. The abrupt transition from a 1 percent to a 10 percent increase in travel experienced when the income product F_iF_j exceeds $G = 0.075$ might be smoothed by replacing the segmented demand model with a continous function.

The effects of small changes in the model input variables, time, cost, and frequency, vary with the importance of the individual mode in intercity travel. If mode m dominates intercity travel (i.e., if w_m/W is nearly unity for mode m), the effects of modal changes on total intercity demand are maximized. Conversely, small changes in infrequently used modes (modes for which w_m/W is vanishingly small) will have slight effect on total intercity demand.

Data given in Table 5 show that a 10 percent increase in the cost of travel by common carrier between 2 cities might cause a decrease of 13.5 percent in the total travel demand between those cities if common carrier is the prevalent mode of intercity travel.

Table 4. Calculated and observed values.

City Pair	Air		Rail		Bus		Automobile		Total	
	Cal.	Obs.	Cal.	Obs.	Cal.	Obs.	Cal.	Obs.	Cal.	Obs.
ALP-SSM	0	0	0	0	0	1	14	19	14	20
SSM-HOU	0	0	0	0	0	0	3	11	3	11
ALP-FLI	0	0	0	0	2	2	13	27	15	29
SSM-FLI	0	0	0	0	1	5	3	51	4	56
CAR-DET	3	0	0	0	4	20	70	660	76	680
SSM-DET	1	5	0	0	1	10	3	274	4	289
HOU-DET	1	7	0	0	0	0	1	62	1	69
FLI-KAL	0	0	3	3	14	25	60	58	77	86
FLI-COL	4	0	1	2	1	2	13	16	19	20
FLI-DES	1	1	0	0	0	1	0	3	2	5
FLI-DET	51	9	55	30	269	250	4,096	4,618	4,470	4,907
FLI-CHI	29	31	4	5	9	20	59	77	101	133
FLI-PHL	14	4	0	0	1	2	2	5	18	11
DET-CHI	660	631	97	80	149	150	802	775	1,708	1,636
DET-PHL	155	251	5	5	19	20	40	74	220	350
DET-CLE	332	137	41	3	94	25	650	572	1,117	737
DET-PIT	188	139	34	2	62	10	127	103	411	254
DET-MIL	115	134	16	2	28	10	42	41	202	187
FLI-CLE	25	9	0	1	4	4	26	22	55	36
FLI-MIL	3	3	1	0	2	1	4	4	10	8

Figure 1. Normalized demand versus income product.

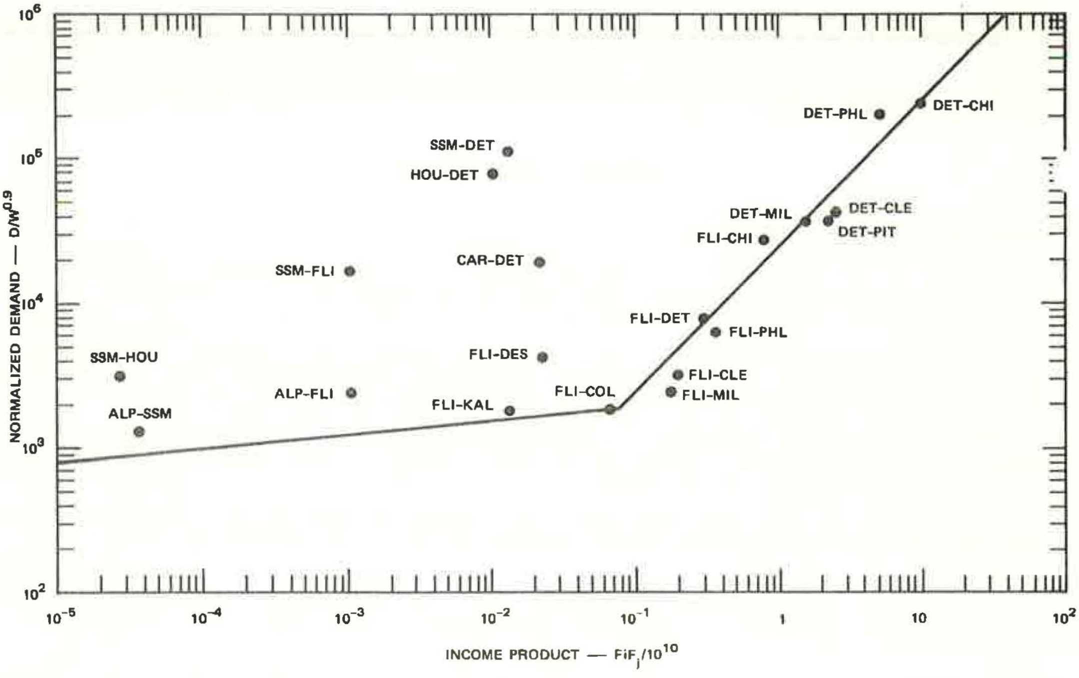


Table 5. Model sensitivity to 10 percent increase in input variables.

Variable	Case	Change in Modal Demand (percent)		Change in Total Demand (percent)	
		Minimum	Maximum	Minimum	Maximum
F	$F_iF_j > 0.075$	10.0	10.0	10.0	10.0
	$F_iF_j \leq 0.075$	1.0	1.0	1.0	1.0
t_m	m ≠ automobile	-13.5	-15.0	0.0	-13.5
	m = automobile	-16.2	-18.0	0.0	-16.2
c_m	m ≠ automobile	-13.5	-15.0	0.0	-13.5
	m = automobile	-16.2	-18.0	0.0	-16.2
f_m	m ≠ automobile	0.0	3.9	0.0	2.6

If automobile is the prevalent mode of intercity travel, the effect of such a cost increase on total intercity travel would be negligible. The effect of the fare increase on travel via the affected mode would be a loss of between 13.5 and 15 percent of the mode's pre-increase travel volume. Similar ranges would be expected in the event of a 10 percent increase in travel time. These ranges are given in Table 5 along with the corresponding ranges for changes in the time and cost of automobile travel. The magnitude of these small changes does not appear to be unreasonable, and, thanks to the constraints imposed in the calibration procedure, the direction of change is proper.

Service frequency is the least effective of the input variables in terms of its ability to influence sizable demand changes. A 10 percent increase in common carrier service frequency can effect no more than a 3.9 percent in modal patronage and no more than a 2.6 percent increase in total intercity travel.

Effect of Parameter Changes

Just as the effect of variable changes on predicted demand gives a measure of model reasonability, so the effect of parameter variations on demand gives a measure of model stability. If a small parameter change can drastically alter model output, the calibration procedure is complicated, and model validity may be suspect.

The effect of small parameter changes on total demand is quite complex and may depend on the relative impact of a mode on intercity travel; on the existing population product; on current levels of time, cost, and frequency; or on all of these factors. An evaluation of $\partial D/\partial P$, the first derivative of demand with respect to each model parameter, shows that the parameters whose changes have the greatest potential impact on demand are the time and cost components $\alpha(1)$, $\alpha(2)$, $\alpha(4)$, and $\alpha(5)$ and the conductance exponent $\beta(2)$.

The segmenting of the demand model buffers the effect of changes in the income exponent $\beta(1)$. Were it not for this segmentation, a 10 percent change in the parameter $\beta(1)$ could effect a 90 percent change in the demand calculated between a small-city pair. This buffering effect suggests that model stability and performance might be improved by similarly segmenting the model with respect to the conductance exponent $\beta(2)$. Such a segmentation would buffer the potentially pronounced effect of changes in the modal time and cost exponents.

INDUCED AND DIVERTED DEMAND

When improvements in a single mode cause an incremental increase in the number of travelers using that mode, these travelers can be assumed to come from 1 of 2 sources: (a) other modes (diverted demand) or (b) the pool of potential travelers who currently are not included in the total intercity demand (induced demand). Thus, total modal increases are made up of travelers diverted from other modes and travelers induced to make the intercity journey for the first time (or more often). Although the calibrated demand model behaves logically in reproducing the overall impact of variable changes, numerical results of a number of model runs revealed that the model clearly overstates induced demand at the expense of diverted demand.

The reason for this overstatement becomes clear if the sources of incremental demand increases are investigated. Equation 5, repeated here for the sake of convenience, expresses the effect of an incremental cost change in mode m on a competing mode n.

$$\Delta D_n = [\alpha(2)/c_m]\,[\beta(2) - 1]\,(w_m/W)\,D_n \Delta c_m \quad \text{for } n \neq m$$

The effect of the cost change on the demand for service via mode m is as follows:

$$\Delta D_m = [\alpha(2)/c_m]\,D_m\,\{1 + (w_m/W)\,[\beta(2) - 1]\}\,\Delta c_m \tag{6}$$

Summing the expression given above across all modes gives the total intercity demand increment.

$$\Delta D = \Delta D_m + \sum_{n \neq m} \Delta D_n \tag{7}$$

$$\Delta D = [\alpha(2)/c_m]\, \beta(2) D_m \Delta c_m \tag{8}$$

In the case of a cost decrease, the constrained calibration procedure forces ΔD_m to be positive and ΔD_n to be negative. Hence, the total demand increment ΔD will represent the total induced demand. The ratio of induced demand to the incremental demand increase via mode m may be found as follows:

$$\Delta D/\Delta D_m = [\beta(2)] / \{ 1 + (w_m/W)\, [\beta(2) - 1] \} \tag{9}$$

For the calibrated value of $\beta(2) = 0.9$, this ratio will vary from 0.9 to 1.0 as the ratio w_m/W varies from 0 to 1. Thus, the induced demand component of traffic increases predicted by the intercity demand model will range between 90 and 100 percent. This is not a realistic state of affairs. The model's realism may be improved, however, by defining arbitrarily a more reasonable limit on induced demand and redistributing demand forecasts in accordance with this limit. A simple means of accomplishing this redistribution is to let

$$D_m = (D_0 + \gamma \Delta D)\, (w_m/W) \tag{10}$$

where D_0 represents original intercity demand and γ represents an arbitrary scaling factor ($0 \leq \gamma \leq 1$).

FUTURE WORK

Model Improvements

Future work to improve the accuracy and plausibility of the Michigan intercity demand model might profitably explore the following subjects: segmentation over distance; formulation of a continuous model; development of an induced demand correction factor; and investigation of the variation of parameter values over time.

Distance Segmentation—The possibility of segmenting the demand model as a function of distance by associating different values of $\beta(2)$ with different conductances has been noted already. The intercity highway traffic model designed for Michigan (4) was segmented in this fashion with good results. Such a segmentation would correct for the tendency of the current model to understate long-distance trips (more than 600 miles).

Continuous Model—Certain inconsistencies in model performance might be overcome by developing a continous demand model having the features of the segmented model. A continuous model having these features is shown below.

$$D = \beta(0)^{(1-S)}\, (F_i F_j)^{\{\beta(1) + s[\beta'(1) - \beta(1)]\}}\, \beta'(0)^{s} W^{\beta(2)} \tag{11}$$

where $S = \exp[-\mu(F_i F_j) + \tau]$. The variables μ and τ are calibration constants, and the remaining model variables have the definitions stated in Eqs. 1, 2, and 4. SRI has achieved some success in calibrating the model of Eq. 11, but more experimentation is necessary before this model can replace the current segmented formulation. A similar continuous formulation could be employed to vary the parameter $\beta(2)$ for different intercity distances.

Induced Demand Correction Factor—Historical data regarding induced demand should be gathered in an effort to estimate the value of the parameter γ used in Eq. 10 to correct for the model's tendency to overstate induced demand.

Time-Varying Parameters—If the functional form of the demand model is correct, it seems likely that parameter values will change with time. This supposition should be checked by calibrating the model at different points in time and attempting to explain and quantify any differences in the calibration parameters.

Model Application

The true utility of the developed demand model is best tested by applying the model in the investigation of intercity transportation problems. In the course of SRI's Michigan studies, the model has been applied to the task of predicting potential air traffic from a proposed regional airport (8) and evaluating alternative high-speed rail routes between Detroit and Chicago (9). The model performed creditably in these tests. More such tests are needed to substantiate the model's current capability and to point the way toward future improvements.

REFERENCES

1. Lynch, J. T., et al. Panel Discussion on Inter-Area Travel Formulas. HRB Bull. 253, 1960, pp. 128-138.
2. Metzger, W., and Ross, H. R. An Analysis of Intercity Passenger Traffic Movement Within the California Corridor Through 1980. Stanford Research Institute, Menlo Park, Calif., Proj. SEU-5589, April 1966.
3. Whiteside, R. E., Cothran, C. L., and Kean, W. M. Intercity Traffic Projections. Highway Research Record 205, 1967, pp. 110-135.
4. Transportation Prediction Procedures—Highway Travel. Arthur D. Little, Inc., and Michigan Department of State Highways, Tech. Rept. 9B, Dec. 1966.
5. McLynn, J., and Woronka, T. Passenger Demand and Modal Split Models. Northeast Corridor Transportation Project, U. S. Department of Transportation, NECTP Rept. 230, Dec. 1969.
6. Donaldson, J., et al. National Bureau of Standards Modeling for the NECTP. Northeast Corridor Transportation Project, U. S. Department of Transportation, NECTP Rept. 213, Dec. 1969.
7. Northeast Corridor Transportation Project, U. S. Department of Transportation, NECTP Rept. 209, April 1970.
8. Billheimer, J. W. Regional Airport Passenger Projections. Stanford Research Institute, Menlo Park, Calif., Final Rept. 3, May 1971.
9. Billheimer, J. W. Analysis of Alternative Rail Passenger Routings in the Detroit-Chicago Corridor. Stanford Research Institute, Menlo Park, Calif., Final Rept. 4, July 1971.

EFFECT OF ZONE SIZE ON TRAFFIC ASSIGNMENT AND TRIP DISTRIBUTION

Bruno R. Wildermuth, Wilbur Smith and Associates;
D. J. Delaney, Transport Planning Division, Commonwealth Bureau of Roads, Melbourne, Australia; and
K. E. Thompson, Bureau of Transport Economics, Canberra, Australia

The analysis in this paper of the effect of zone size on traffic assignment and trip distribution was based on the results of a research study initiated by the Australian Commonwealth Bureau of Roads. This investigation was concerned with identifying criteria for zone size selection, based on the results of applying standard trip distribution and traffic assignment procedures to an actual road network and vehicle trip matrix. Vehicle trips were aggregated to 5 test systems, ranging from 40 to 263 zones, and assignments were made to the basic 1964 road network. The effect of different zone systems on traffic assignment was evaluated statistically by comparing assignments with those for the study area's 607-zone trip matrix and with 1964 traffic counts at 16 check lines located throughout the study area. The effect of zonal aggregation on the simulation of trip distributions was also tested, both for a 1-purpose and a 4-purpose distribution. Distribution errors attributable to the aggregation of zones and to the grouping of trip purposes were evaluated for 3 zone plans by comparing assignments of the resulting trip matrices with the 607-zone system assignment and with the 1964 check-line volumes. It is concluded that for many aspects of transport planning, zone plans with an average of 30,000 trip ends per zone will yield traffic assignments sufficiently accurate for predictions of traffic growth within transportation corridors.

• TRAFFIC assignments for many aspects of transportation planning are not required to be so accurate as those required for design purposes. Rather, emphasis may be placed on the reliable prediction of traffic growth for travel corridors or for subareas of metropolitan areas. Such predictions may be desirable for evaluating a number of alternative urban development schemes for a range of time periods. Substantial savings in efforts and computer costs could be expected if estimates could be made with fewer analysis zones. The Commonwealth Bureau of Roads, therefore, initiated a research study to identify and measure the likely traffic assignment errors that would result from the use of fewer zones than normally used in transportation studies.

The basic approach in the investigation was to devise a number of alternative zone plans and to assign the resulting rearranged trip matrix to a common road network. The original origin-destination survey trip matrix for all vehicle trips was used for the purpose. Traffic assignments were performed by utilizing 2 noniterative techniques, the all-or-nothing method and a multiple-route method. (The multiple-route method is based on the assumption that the user of a network does not know the actual link times he is going to encounter but that he associates with each link in the network a probable link time based on past travel experience. In assignments, the probable link time is drawn at random from a normal distribution of times for that link every time it is considered in path building. The mean of the normal distribution is equal to the initial estimate of link time, and the standard deviation of the distribution is assumed to be 18 percent of the value of the mean.)

An analysis was made of those trips that became intrazonal trips under different zone plans. Trip matrices containing only those trips were prepared and assigned to the basic network to determine average vehicle-miles and vehicle-hours of travel for intrazonal trips.

Standard gravity or interactance trip distribution models were calibrated for 3 alternative zone plans. Both a 4-purpose model and a 1-purpose model were developed, and the resulting trip matrices were assigned to the basic road network. A unique feature of the trip distribution model was the use of separate distribution functions for central business district trips, external trips, and the remaining internal trips.

Statistical tests were made to compare assigned link volumes with traffic counts at 16 check lines located throughout the study area. In addition, vehicle-miles and vehicle-hours of travel were summarized for 56 districts, identical to one of the test zone systems. Comparisons were made with the assignment that had the smallest differences from the traffic counts on all links crossing the 16 check lines. Finally, the frequency distribution of the assigned volumes on the links of each zone system were determined. Particular attention was given to the links that were not assigned any trips under alternative zone plans.

STUDY DATA

The source of data used for the research was the 1964 Melbourne Metropolitan Transportation Study, which used 607 internal traffic zones and 32 external cordon stations. A home interview survey of about 30,000 households (5 percent sample) provided most of the origin-destination data. Also surveyed were drivers of 9,000 trucks (10 percent sample) and nearly 600 taxis (25 percent sample). Roadside interviews at 32 external cordon stations provided travel data on approximately 90,000 vehicle trips. Some trips had both ends inside. Table 1 gives the internal and external trips by 4 trip-purpose categories.

The 1964 road network coded for study purposes comprised 4,936 one-way links representing 1,085 miles of arterial and collector roads. The network contained 2,440 centroid connector links and 2,519 nodes, including the 639 centroids.

Travel time studies made in 1964 provided estimates of average daily running speeds for the network. Subsequently, some of the link speeds were modified to achieve better correspondence between actual volumes and volumes obtained from the all-or-nothing assignment.

ALTERNATIVE ZONE SYSTEMS

Various criteria, sometimes conflicting, were used in the selection of the 607 zones (hereafter referred to as system 607). These criteria included uniformity of land use, presence of physical barriers to travel, existing area subdivisions, and a requirement that the diameter of the zones should generally have a travel time of 3 to 6 min. For the purpose of statistical comparison, zone systems were defined as combinations of the original 607 zones. All of these criteria were observed, to the extent practicable, in defining the various test zone plans.

The following ranges of internal zones were initially defined for investigation: 30 to 50, 100 to 150, and 250 to 300 zones. The zones would have average areas of 12 to 20, 4 to 6, and 2 to 2.5 square miles and average widths of 5, 2.5, and 1.5 miles respectively. These average values were used only as a general guide in defining zones. Each of the zonal systems had smaller zones near the center and larger zones in the outlying parts of the study area.

The central business district and several suburban centers have trip-end densities far higher than those found elsewhere in the study area. These centers were retained as separate zones in systems 136, 144, and 263 but necessarily were combined with immediately adjacent areas in the test plans with the largest zones (systems 040 and 056).

In addition to the reduction in the number of internal zones, external stations were also combined, where possible, on the basis of physical proximity. Traffic entering from different external stations had to use the same general travel corridors to justify the use of a combined external centroid.

The criteria selected were consistent with those generally adopted in transport planning; therefore, the conclusions of this study would be applicable to other cities. In accordance with these criteria, test systems 040, 136, and 263 were initially established.

After the analysis of assignment results from the initial test systems, it was concluded that a limit for the maximum number of trip ends per zone would substantially improve assignment accuracy without materially increasing the number of zones. Consequently, test systems 056 and 144 were established as modifications of the initial systems 040 and 136. Accordingly, assignment errors were investigated for the range of test plans given in Table 2.

ANALYSIS OF INTRAZONAL TRIPS

The amount and percentage of intrazonal trips depend primarily on the size and type of land use within zones. Because intrazonal trips were not assigned to the network by the normal traffic assignment process, the magnitude of travel caused by these trips must be considered in connection with changing zone sizes. Figure 1 shows the relation between the percentage of intrazonal trips and the average number of trip ends per zone for the zone plans tested.

To determine the magnitude of intrazonal travel in terms of vehicle-miles and vehicle-hours, the intrazonal trips associated with each test system were assigned to the system 607 network. In this way, it was possible to isolate the link loadings resulting from intrazonal trips in addition to the 220,124 trips that were intrazonal for system 607. The all-or-nothing method was used for this purpose. (This method was selected to eliminate effects of the stochastic process associated with the multiple-route method. However, because such trips would generally travel relatively short distances, few alternative paths would be identified in a multiple-route assignment.) Table 3 gives the resulting vehicle travel from these assignments for the major arterial and collector road links (i.e., all links except centroid connectors). Table 4 gives the average and incremental vehicle-miles and vehicle-hours for intrazonal trips. Incremental value are computed from travel generated by intrazonal trips that were not intrazonal in the system with which the comparison is drawn.

An investigation was also made of the concentrations of intrazonal trips. Table 5 gives frequency distributions of 1-way links by volume prepared from the assignments of the intrazonal trips to the system 607 network. Depending on the number of zones, between 20 and 70 percent of all links will not carry any intrazonal trips. Only for the very coarse zone plans of systems 040 and 056 are more than 3,000 intrazonal trips concentrated on some links.

It was concluded that the volumes missed in the assignment process, because of intrazonal trips of systems 263, 144, and 136, were not significant enough to affect capacity evaluations. The higher volumes of intrazonal trips assigned to some links in systems 056 and 040 could affect design, but even so the effects throughout a system planned for capacity continuity would not be great.

COMPARISON OF ASSIGNMENTS

The analysis and comparison of the traffic assignments resulting from alternative zone plans covered 2 primary areas of investigation. The first test evaluated the assignments with respect to actual traffic counts. Two-way link volumes for 16 check lines were available for this purpose. Three of these 16 ckeck lines represented the major screen lines used for the verification of the original origin-destination survey data, and the others were located throughout the study area. The second test involved the comparison of vehicle-miles and vehicle-hours of travel on a district basis. The results from the system 607 multiple-route assignment were used as a basis for these comparisons. The system 607 multiple-route assignment was selected because it showed the lowest root mean square (rms) error and chi-square error for the ckeck-line comparisons.

In addition to these major tests, comparisons were made of the frequency of links by volume range. Some line-printer plots were also produced for visual comparison of assignments from the different zone systems on a link-by-link basis.

Table 1. 1964 internal and external vehicle trips.

Category	Internal	External	Total
Automobile drivers			
Non-home-based	341,693	8,824	350,517
Home-based, work	518,578	19,443	538,021
Home-based, other	539,579	33,922	573,501
Commercial vehicles	631,666	25,407	657,073
Total	2,031,516	87,596	2,119,112

Table 2. Zonal characteristics of alternative test systems.

System	Internal Zones	External Stations	Total Centroids	Maximum Number of Trip Ends per Zone
607	607	32	639	35,894
263	263	23	286	124,422
144	144	17	161	79,176
136	136	23	159	217,355
056	56	17	73	124,422
040	40	23	63	342,813

Figure 1. Intrazonal trips in relation to average number of trip ends per zone.

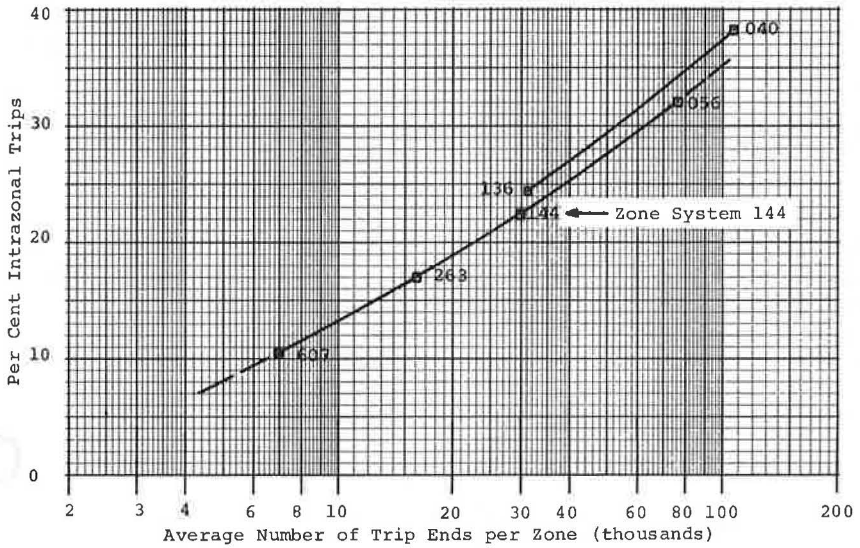


Table 3. Vehicle-miles and vehicle-hours of intrazonal trips for alternative systems.

System	Additional Intrazonal Trips[a]	Vehicle-Miles Number	Vehicle-Miles Percent of System 607[b]	Vehicle-Hours Number	Vehicle-Hours Percent of System 607[b]
263	139,373	104,584	1.17	4,194	1.19
144	254,742	234,929	2.63	9,009	2.56
136	297,504	277,471	3.11	11,479	3.27
056	454,840	540,992	6.05	20,242	5.76
040	594,044	793,704	8.88	31,120	8.85

[a] Intrazonal trips in addition to those already existing for system 607.

[b] System 607 vehicle-miles and vehicle-hours resulting from the minimum time path all-or-nothing assignment method.

Table 4. Average and incremental vehicle-miles and vehicle-hours per intrazonal trip for alternative systems.

System	Average[a] Vehicle-Miles	Average[a] Vehicle-Hours	Incremental[b] Comparison System	Incremental[b] Vehicle-Miles	Incremental[b] Vehicle-Hours
263	0.75	0.030	607	0.75	0.030
144	0.92	0.035	263	1.13	0.042
136	0.93	0.039	263	1.09	0.046
056	1.19	0.045	144	1.53	0.056
40	1.34	0.052	136	1.74	0.066
			056	1.82	0.078

[a] Computed from Table 3, excluding intrazonal trips from system 607.

[b] Computed for intrazonal trips that were interzonal in the comparison system cited.

Network Preparation

The original 1964 road network of the Melbourne Metropolitan Transportation Study was used for each assignment. This was accomplished by building a link data card image tape on which centroid connectors only were changed, or eliminated, as necessary.

For each zone system, one of the original system 607 zone centroids was designated to represent each new group of zones. All other centroids within the group were eliminated. The analysis of the initial assignments for systems 263, 136, and 040 indicated, however, that this method was not completely satisfactory because many of the original zones were connected with only 1 or 2 centroid connectors. Although this method is quite satisfactory for a large number of small zones, it proved to be a disadvantage for the zone plans with fewer large zones. The lack of an adequate number of alternative centroid connectors caused very large volumes of trips to be assigned to links in the immediate vicinity of the connectors and thereby caused gross overloadings on certain links. Increasing the number of connectors distributes interzonal trips more evenly onto the network, a detail that assumes increasing importance with decreasing numbers of zones. Consequently, additional centroid connecting links were coded for access to each zone in the preparation of networks for systems 144 and 056.

Analysis of Check-Line Crossings

Sixteen check lines ranging in volume from 21,500 to more than 500,000 vehicles/day were used in this analysis. The number of 2-way network links involved ranged from 3 for check lines 8 and 13 to 24 for check line 1, the major north-south screen line. The total for all check lines was 194 links.

The rms error, a measure of the average error on individual links, and the chi-square error, a cumulative measure of the error terms with emphasis on the large errors, were selected for the analysis.

$$\text{rms error} = \sqrt{(1/n) \sum_{i=1}^{n} (G_i - A_i)^2}$$

$$\text{chi-square error} = \sum_{i=1}^{n} [(G_i - A_i)^2]/G_i$$

where

G_i = ground count volume for link i,
A_i = assigned volume for link i, and
n = number of links.

Table 6 gives the assignment errors, measured on links crossing the 16 check lines, for each alternative zone system and for the 2 assignment methods. These comparisons show somewhat better results for the multiple-route assignment method than for the all-or-nothing method or the networks that were coded to allow an adequate range of alternative paths to be selected. The results illustrate that overall assignment errors at check lines increased relatively slowly as the number of zones was decreased. Systems 263 and 144, in which the original number of zones was reduced to less than one-half and one-fourth respectively, produced highly acceptable assignment results at all check lines. Chi-square errors for these 2 systems were 7 and 9 percent greater than for the lowest errors measured, which were for the multiple-route assignment to system 607. The error for system 144 was 6 percent greater than the lowest error, and the increase in error for system 263 was negligible. The rms errors measured for all 16 check-line crossings were nearly identical for systems 607 and 263, and the corresponding chi-square errors increased between 7 and 10 percent. This indicates that errors for some individual links were increased by the coarser zoning but that the average error remained unchanged.

Table 5. Frequency of 1-way links by volume for intrazonal trips assigned to system 607 network.

Link-Volume Range	System 263	System 144	System 136	System 056	System 040
0	3,514	2,371	2,242	1,379	1,038
1-1,000	1,408	2,529	2,546	3,207	3,059
1,001-2,000	14	36	142	310	683
2,001-3,000	0	0	6	32	134
3,001-4,000	0	0	0	6	20
4,001-5,000	0	0	0	2	2
Total	4,936	4,936	4,936	4,936	4,936

Table 6. rms and chi-square assignment errors for links crossing 16 check lines.

	Multiple-Route Assignment		All-or-Nothing Assignment	
System	rms Error	Chi-Square Error	rms Error	Chi-Square Error
607	7,022	591,928	7,596	648,321
263	7,047	635,564	7,494	714,428
144	7,454	647,452	8,940	850,789
136	8,722	800,039	8,635	723,685
056	9,618	1,189,126	10,553	1,327,603
040	12,510	2,105,617	12,255	1,895,987

Table 7. Vehicle-miles and vehicle-hours measured by 2 assignment methods.

	Multiple-Route Assignment		All-or-Nothing Assignment	
System	Vehicle-Miles	Vehicle-Hours	Vehicle-Miles	Vehicle-Hours
607	9,145,191	360,992	8,934,996	351,487
263	9,127,628	361,396	8,912,493	351,564
144	8,710,338	343,250	8,460,635	330,922
136	8,965,009	357,042	8,765,627	348,187
056	8,532,841	329,620	8,313,325	327,089
040	8,895,391	352,889	8,701,668	342,589

Note: Travel on major arterial and collector roads only, i.e., excluding centroid connectors.

Table 8. rms errors for district comparisons by multiple-route assignment.

	rms Error		Coefficient of Variation[a]	
System	Vehicle-Miles	Vehicle-Hours	Vehicle-Miles	Vehicle-Hours
263	9,826	411	6.0	6.4
144	16,244	637	10.0	9.9
136	15,818	778	9.7	12.1
056	22,160	879	13.6	13.6
040	38,307	1,594	23.5	24.6

[a] rms error expressed as the percentage of the mean vehicle-miles (163,307) and the mean vehicle-hours (6,446) for the system 607 multiple-route assignment.

Table 9. Cumulative difference by district of vehicle-miles by multiple-route assignment for system 607.

Percentage Difference Range	System 263	System 144	System 136	System 056	System 040
± 0.00- 2.49	19	11	14	12	7
± 2.50- 4.99	37	24	23	21	12
± 5.00- 7.49	44	31	28	26	20
± 7.50- 9.99	51	40	37	35	22
±10.00-12.49	54	43	43	38	25
±12.50-14.99	54	48	47	41	32
±15.00-19.99	55	50	50	48	35
±20.00-24.99	56	54	53	48	41
±25.00-29.99	56	55	53	52	44
±30.00-34.99	56	55	56	54	49
35.00-39.99	56	55	56	54	54
40.00 and more	56	56	56	56	56

Analysis of Total Vehicle Travel

From the analysis of intrazonal trips, it would be expected that reductions in the number of zones would be accompanied by a reduction in the assigned vehicle-miles and vehicle-hours of travel. Table 7 gives these travel measures for each alternative zone system for each of the assignment methods.

The reduction in vehicle-miles for system 263 from those for system 607 was less than 0.2 percent; the vehicle-hours showed a small increase. The largest decrease in vehicle-miles, registered for system 056, was almost 7 percent of the total travel mileage for system 607. The consistently higher amounts of travel for the multiple-route assignments would be expected because of the selection of alternative paths, which would be longer than the minimum paths.

The loss of vehicle travel on the network generally was less than the travel lost in intrazonal trips for the networks that did not have sufficient centroid connectors. Systems 056 and 144, for which additional connectors were coded, had travel losses only slightly larger than losses due to intrazonal trips. However, even for system 056, which had less than one-tenth original number of zones, the loss in vehicle-miles and vehicle-hours of assigned traffic was only 7 percent of the system 607 travel. For many aspects of transportation planning, errors of such small magnitude would be acceptable.

Further analysis of the assignments was concerned primarily with identifying the range of assignment errors in sections of the metropolitan area. System 056 was used as a basis for this analysis, and each link on the arterial and collector road system was identified by the district in which it was located. Vehicle-miles and vehicle-hours of travel from the multiple-route assignments for each test system were summarized, by each of the 56 districts, and compared with the travel summaries from the multiple-route assignment for system 607. The resulting rms errors are given in Table 8.

Table 9 gives the cumulative frequency of districts by percentage difference of vehicle-miles for the alternative zoning system. Vehicle-miles for system 263, for instance, differed from those of system 607 by 10 percent or more in only 5 districts. In only 2 of these 5 districts was the difference more than 12 percent. Thus, in 95 percent of the districts, assigned travel for system 263 would be in error by no more than 12 percent. For systems 144 and 136, however, an error larger than 12 percent occurs in 13 districts, or 23 percent of all districts.

A typical spatial distribution of percentage differences in vehicle-miles of travel by district is shown in Figure 2. Visual inspections of such distributions for all systems indicated that the spatial pattern of errors did not seem to be systematic from one test system to the next. Only 6 districts (2, 8, 9, 30, 41, and 46) were found to have an error greater than 5 pecent in all test systems. District 33 was the only district that had less than 5 percent error in each test plan.

Comparisons of Assignments by Link Volumes

Analyses were also made of frequency distribution of link volumes. Of particular interest was the number of links that were assigned either no volumes or very low volumes. The number of such links would be expected to increase as the number of zones decreased. Analysis of the frequency tabulations indicated that with the multiple-path assignment fewer links were unused. In fact, the number of links with 0 volumes was generally half that for a minimum time path assignment. Because only the major arterial and collector roads were represented by the network, the multiple-path assignment was more realistic in this aspect.

Figure 3 shows that a consistent relation existed among the number of low-volume links (0 to 2,000) the number of zones used, and the assignment technique applied.

Links with excessively high volumes could also constitute a problem in zone aggregation. However, most of the extremely high-volume links were found to be located immediately in the vicinity of centroid connections. With the limitation of the maximum number of trip ends and the provision of adequate alternative connections to centroids, as used for systems 056 and 144, these extremes did not occur.

Figure 2. Differences in vehicle-miles by multiple-route assignment for systems 263 and 607.

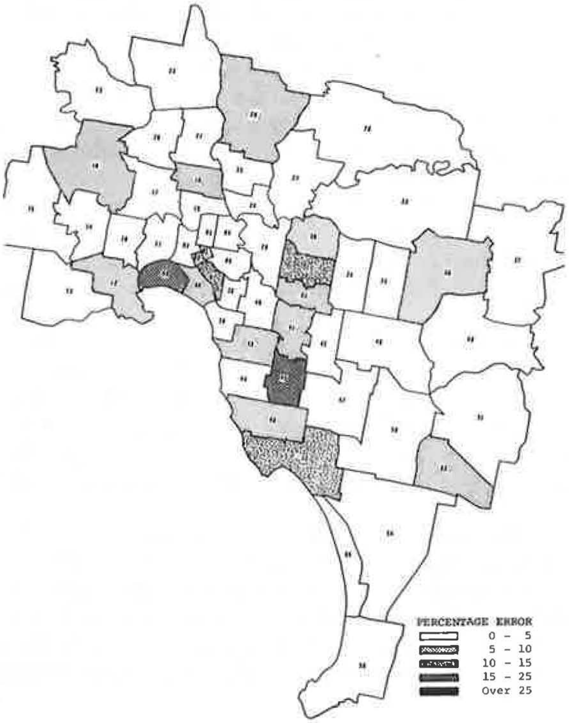
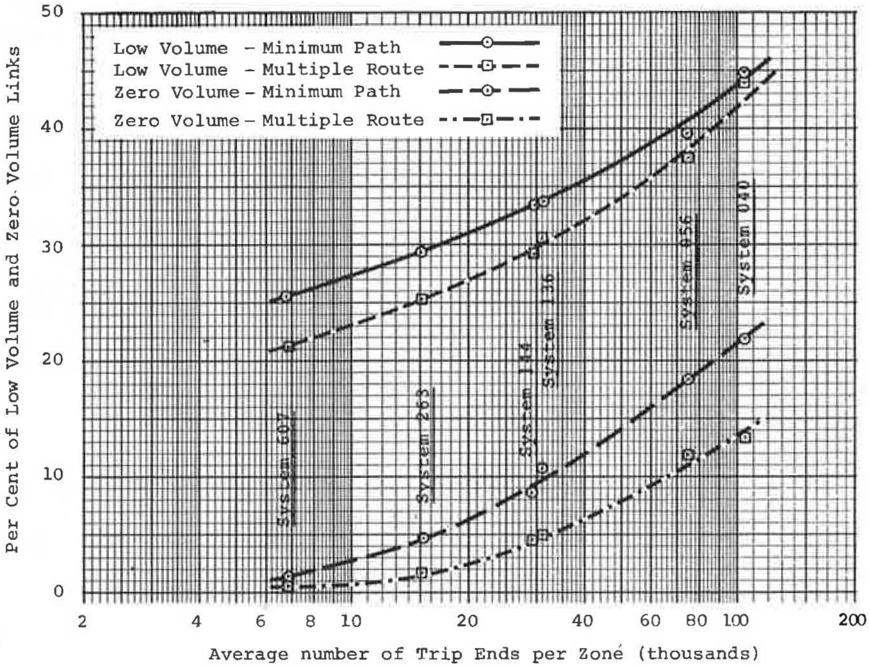


Figure 3. Low-volume lines and 0-volume lines for 2 assignment methods related to average number of trip ends per zone.



ASSIGNMENT TEST FINDINGS

The investigation established measures of traffic assignment errors resulting from tests of widely varying zone plans. Also analyzed were the percentage of intrazonal trips and the corresponding magnitude of travel lost in zone aggregation.

The amount of travel lost because of intrazonal trips was relatively small, even though these trips may constitute a sizable percentage of total trips. Intrazonal trips lost ranged from 17.0 percent for system 263 to 38.6 percent for system 040. However, these trips accounted for only 1.2 and 8.9 percent respectively of assigned vehicle-miles of travel. Such losses could easily be tolerated for many planning purposes, especially because the study found that these trips are not concentrated to any significant degree.

Comparisons of assigned volumes with traffic counts at 16 check lines indicated that the number of zones could be reduced to less than half of the original number before increases in rms errors occurred. The multiple-route assignment produced lower errors than the all-or-nothing assignment.

Within sections of metropolitan Melbourne, the comparison of alternative assignments could be made only against the system 607 multiple-route assignment that, of course, was affected by some errors of its own. rms errors for 56 districts for multiple-route assignment to systems 263, 144, and 136 were within 10 percent of the average vehicle-miles and 12 percent of the average vehicle-hours for system 607. All-or-nothing assignment increased these errors by between 1 and 2 percent for vehicle-miles and up to 3 percent for vehicle-hours.

The analyses undertaken in this study indicated that the number of zones commonly selected for metropolitan transportation studies may be reduced to a third or even a fourth and still produce traffic assignments with adequate reliability for many planning purposes. It was further shown that additional traffic assignment errors can be minimized by the use of multiple-route assignment, the provision of adequate centroid connections, and reasonable limitations on the maximum number of trip ends concentrated within any one analysis zone.

ANALYSIS OF TRIPS BY PURPOSE

In the transportation planning process, trips are generally categorized by a number of different purposes, and trips made by different vehicles, such as passenger cars, trucks, and taxis are usually treated separately. There are valid reasons for this separation, including differences in trip length, peak-hour concentration, and spatial distribution. However, each additional category increases the cost of data preparation and computer processing. Thus, substantial economies would result if categories could be combined for the purpose of trip distribution.

The study investigated the effect of reducing the number of purposes normally used in a transportation study and the effect of zone aggregation on trip distributions. A 4-purpose model and a 1-purpose trip distribution model were considered, each being calibrated for 3 different zone plans: systems 263, 144, and 056.

The 1-purpose model was calibrated against the all-vehicle trip matrix used for the assignment test phase of the study. The 4-purpose model consisted of 3 categories of automobile driver trips (non-home-based, home-based work, and home-based other) and 1 category of commercial vehicle trips. Home-based trips were arranged so that trip production was always from the same zone as the home of the driver. Origins and destinations of non-home-based and commercial vehicle trips were assumed to correspond to production and attraction zones respectively.

Trip Length Distributions

Trips of different purposes and vehicle types are generally found to have significant variations in trip length characteristics. Average trip lengths for the categories of trips used in the study are given in Table 10. Average trip lengths were affected to some degree by reductions in the number of zones because of different locations of zone centroids and changes in the proportion of intrazonal trips.

Stratification of Trips by Area

Early in the calibration process it was found that, despite excellent overall correspondence between observed and estimated trip-length frequency distributions, large differences in trip-length characteristics existed for trips associated with the CBD and with external stations. Accordingly, the trip matrices were partitioned for separate analyses of each of these 2 classes of trips and for all other trips.

It was found that the percentage of CBD trips was highest (15.4 percent) for non-home-based trips and lowest for home-based other trips (4.9 percent). For external trips, the order was reversed with 5.6 percent for home-based other trips and 2.3 percent for non-home-based trips. For all trips, the distribution included 9.1 percent CBD trips and 4.0 percent external trips. Of all CBD trips, 37 percent were made by commercial vehicles.

Average trip lengths for trips of the selected spatial classes are given in Table 11. For all categories of trips, CBD and external trips were substantially longer than internal non-CBD trips. Thus, although these 2 trip classes constituted only a small percentage of all trips, they accounted for a substantially larger portion of the total vehicle-hours of travel, ranging from 17 percent for home-based work trips to 25 percent for non-home-based and commercial vehicle trips. If all trips are processed in 1 group, the trip distribution does not adequately reflect the length and special orientation of these trips. However, with the calibration of separate distribution functions for each of the 3 spatial classes, as was done in the study, a satisfactory trip distribution was achieved.

Intrazonal Trips

The number of intrazonal trips increased substantially as the number of zones was reduced. Figure 4 shows the variation in the percentage of intrazonal trips of different trip categories plotted against average trip densities as obtained from the analyses of systems 263, 144, and 056. The proportion of intrazonal trips is best for home-based work trips, ranging from 3.3 percent for system 607 to 10.3 percent for system 056. Home-based other trips and commercial vehicle trips have the largest proportion of intrazonal trips, ranging from 13.3 and 14.0 percent to 30.5 and 26.1 percent respectively.

TRIP DISTRIBUTION MODEL CALIBRATIONS

Trip distribution models, using the 3 area stratifications, were calibrated for systems 263, 144, and 056 for 5 categories of trip. Four calibration cycles were applied in each case to determine the proper impedance function. No particular difficulties were experienced with the calibration of any of the distribution models. The average estimated trip length for each completed calibration differed from the observed trip length by no more than 0.1 min, the largest discrepancy between values of the cumulative trip-length frequencies being approximately 1 percent.

The differences between the impedance functions for trips associated with the spatial classes are shown in Figure 5 for the all-vehicle category and system 144. For long trips, impedance to travel tends to be greater for internal trips than for CBD or external trips. This same relation was observed for each of the categories considered in the study.

Variations among impedance functions for each of the 3 test systems are shown in Figure 6 for the all-vehicle category of trips. The curves were markedly similar, especially for the middle range of travel times into which most of the trips fall: Fewer than 2 percent of the trips were longer than 50 min, the value beyond which the curves diverge. At the other extreme, the variation among impedance values for the very short trips would be expected because of the differing numbers of intrazonal trips. However, for most of the range, the curves followed one another closely and demonstrated that the impedance function described trip-maker characteristics that, of course, were independent, for this range, of the number of zones into which the study area was divided.

Table 10. Average trip length in min by category.

Category	System 263	System 144	System 056
Automobile drivers			
Non-home-based	15.4	14.9	14.6
Home-based, work	19.6	19.0	19.0
Home-based, other	13.5	13.0	13.4
Commercial vehicles	15.0	14.7	14.4
Total	15.8	15.4	15.4

Table 11. Average trip length in min by category and spatial class for system 263.

Category	CBD	Internal	External	Total
Automobile drivers				
Non-home-based	21.2	14.0	24.9	14.3
Home-based, work	29.3	18.5	23.9	19.6
Home-based, other	29.5	11.8	27.1	13.5
Commerical vehicles	18.7	14.0	26.8	15.0
Total	23.4	14.5	26.1	15.8

Figure 4. Intrazonal trips by category related to average number of trip ends per zone.

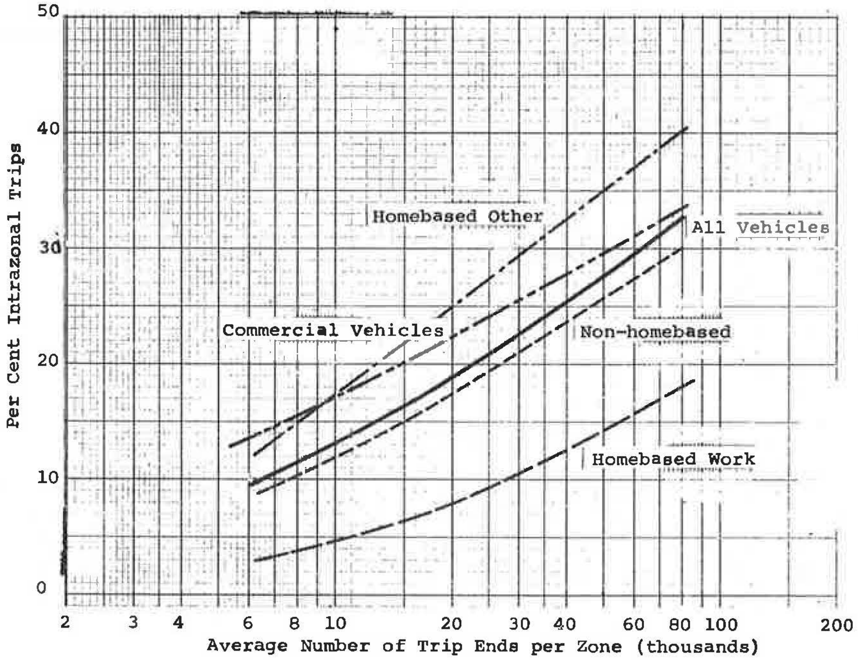


Figure 5. Impedance functions by spatial class for all vehicles and system 144.

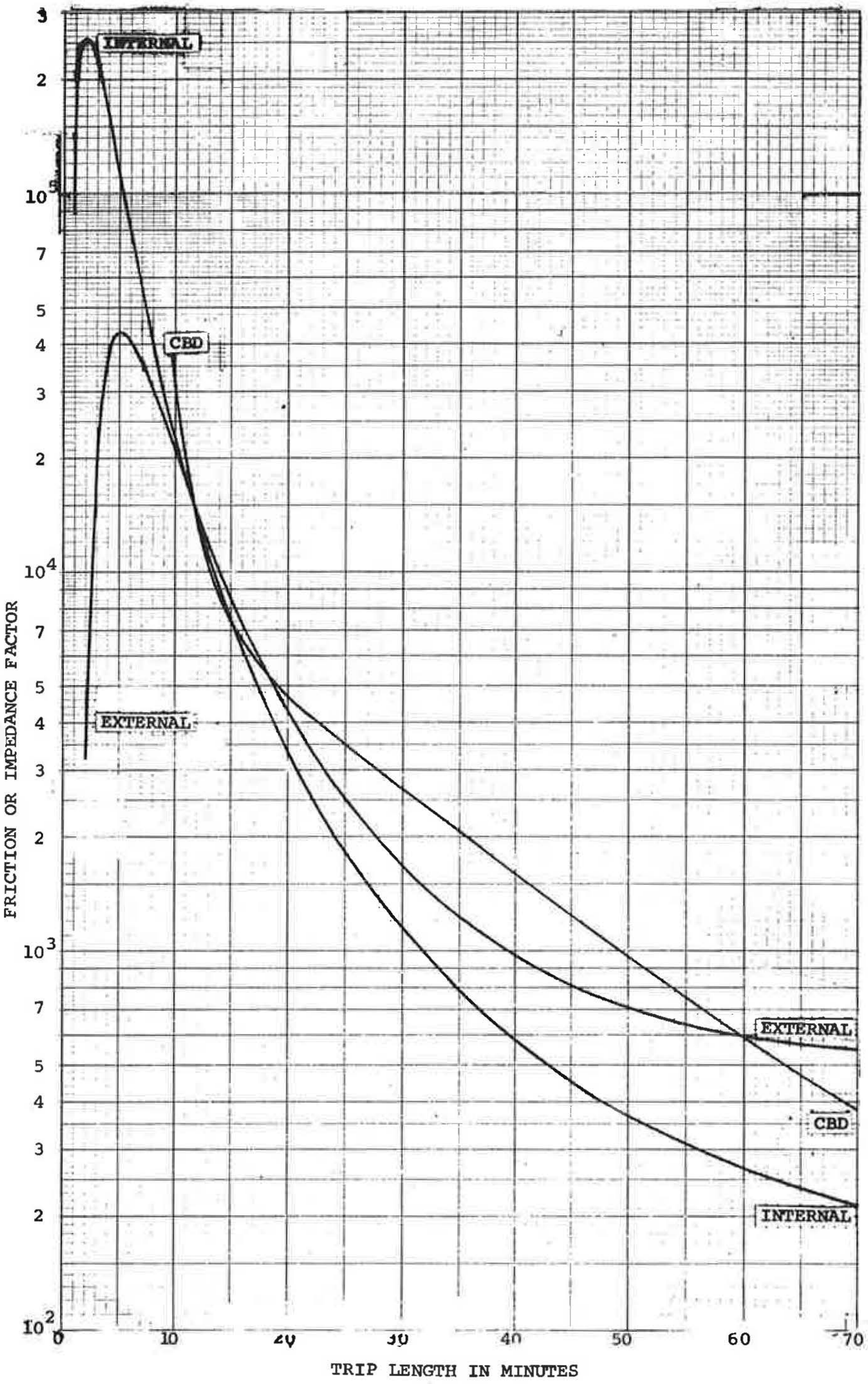
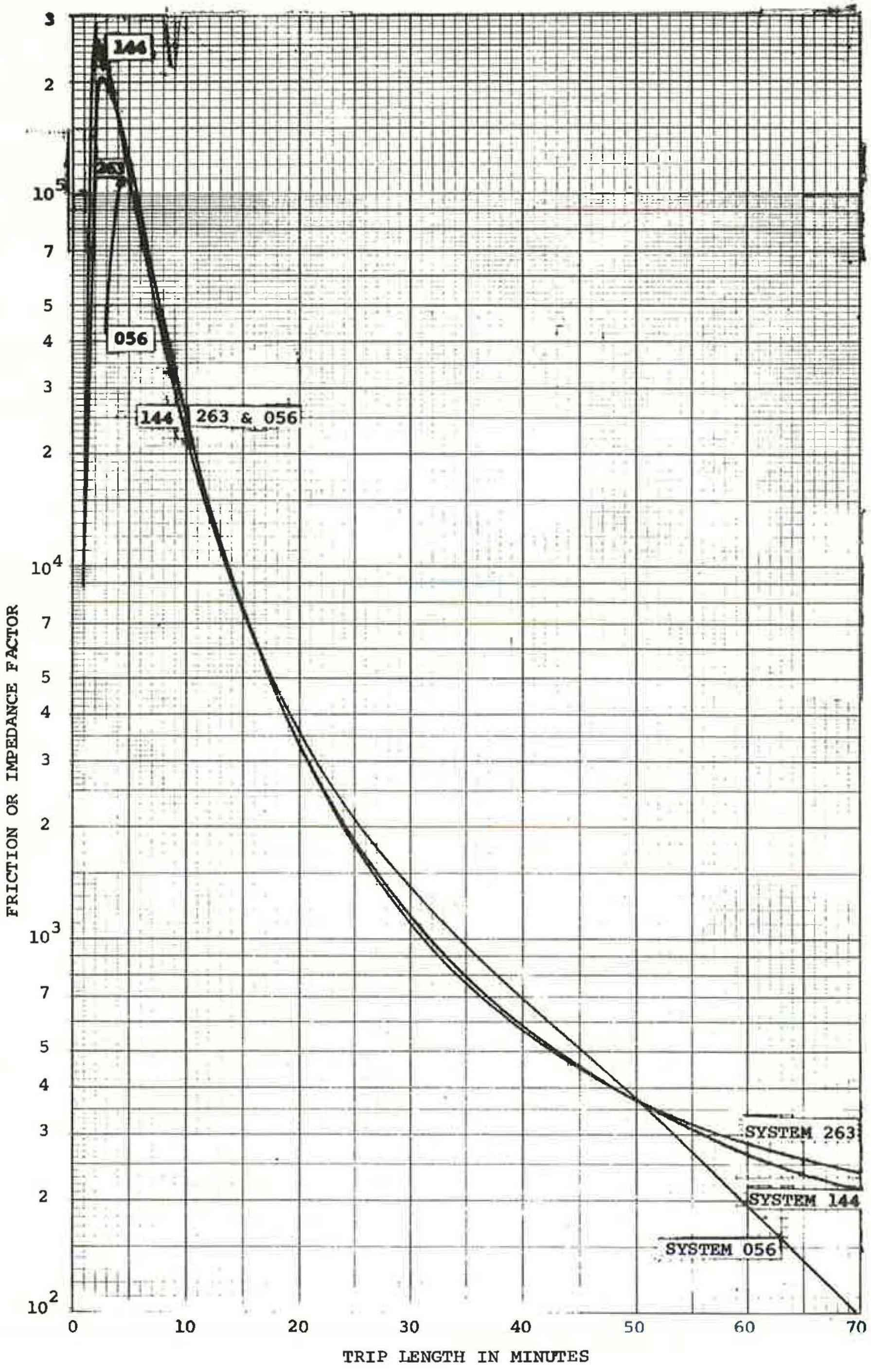


Figure 6. Impedance functions by system for all vehicles and internal trips.



Of the 3 impedance functions shown in Figure 6, the curves for systems 144 and 263 showed the most striking resemblance. Although the system 056 curve is similar, it seems that the drastic reduction in number of zones caused some loss of sensitivity. This tends to suggest that system 056 partitioned the study area into too few zones.

ANALYSIS OF TRIP DISTRIBUTION ASSIGNMENTS

The statistical evaluation and comparison of the 1-purpose and 4-purpose distribution models for the 3 test systems were carried out in the same way as the investigation of the effects of zone aggregation on assignment. Two assignments were prepared for each of the 3 zone systems: one with the trip matrix from the 1-purpose model and one with the all-vehicle trip matrix resulting from the sum of the 4-purpose trip tables. All assignments were made by using the multiple-route technique. The following analyses were made:

1. Two-way link volumes were compared with traffic counts for the 16 check lines;
2. Vehicle-miles and vehicle-hours of travel were accumulated for 56 districts and compared with the basic results from the system 607 multiple-route assignment of the O-D survey trip matrix; and
3. The frequencies of links by volume ranges were compared.

The results obtained for both the 1-purpose and the 4-purpose trip distributions were compared with the corresponding zone plan assignments of the original O-D survey trip matrix. This was necessary to isolate the errors of the trip distribution process from those inherently associated with the zone plan and the O-D survey data used for calibration.

Check-Line Comparison

rms errors and chi-square errors resulting from the 1-purpose and the 4-purpose trip distribution assignments are given for each test plan in Table 12. The results of both trip distribution assignments showed very close agreement with the O-D assignment for all zone systems tested. Both rms and chi-square errors were only slightly smaller for the 4-purpose trip distribution model than for the 1-purpose model. Increases in errors relative to the corresponding O-D assignment error decreased with reductions in the number of zones.

The check-line comparison indicated that reductions in the number of zones tended to reduce errors resulting from the trip distribution process taken alone. Furthermore, the elimination of separate-purpose and vehicle-type categories did not significantly affect the assignment results at any level of zone aggregation.

District Comparison

The reliability of assigned travel throughout the network was assessed by accumulating traffic volumes from the assignments of 1-purpose and 4-purpose trip distributions for each of the 56 districts and by comparing them with the volumes from the multiple-route assignment for system 607. Table 13 gives the total vehicle-miles and vehicle-hours for the 3 test systems. Again, only marginal differences occurred between the 1-purpose model and the 4-purpose model. Both models produced total travel estimates within 3 percent of the estimates obtained from the assignment of the O-D trip matrix. Table 14 gives the comparisons for rms errors of vehicle-miles and vehicle-hours respectively. The errors were greater when comparison was made by district. Table 15 gives the cumulative frequency of districts by percentage difference of vehicle-miles. These data suggest that the 4-purpose distribution provided more added accuracy than was indicated by the check-line comparison.

Link Volume Distribution

The analysis of individual link volumes for 1-purpose and 4-purpose distributions indicated only small differences for most volume ranges. Trip distribution assignments produced fewer low-volume links (0 to 2,000 volume range) than the O-D survey assign-

Table 12. rms and chi-square assignment errors for links crossing 16 check lines.

System	O-D Survey Data rms Error	O-D Survey Data Chi-Square Error	4-Purpose Distribution rms Error	4-Purpose Distribution Chi-Square Error	1-Purpose Distribution rms Error	1-Purpose Distribution Chi-Square Error
263	7,047	635,564	7,631 (8.3)	774,738 (21.9)	7,780 (10.4)	785,465 (23.5)
144	7,454	647,452	7,896 (5.9)	741,073 (14.5)	7,967 (6.9)	748,603 (15.6)
056	9,618	1,189,126	9,801 (1.9)	1,228,486 (3.3)	9,983 (3.8)	1,263,340 (6.2)

Note: Figures in parentheses indicate percentage increase in errors over O-D assignment errors.

Table 13. Vehicle-miles and vehicle-hours by distribution assignment for 3 test systems.

	Vehicle-Miles (thousands)			Vehicle-Hours (thousands)		
System	O-D	4-Purpose	1-Purpose	O-D	4-Purpose	1-Purpose
263	9,128	9,357	9,362	361	367	365
144	8,710	8,955	8,896	343	349	346
056	8,533	8,787	8,745	330	345	342

Table 14. rms errors for district comparison of vehicle-miles and vehicle-hours by distribution assignment for 3 test systems.

	rms Error			Coefficient of Variation		
System	O-D	4-Purpose	1-Purpose	O-D	4-Purpose	1-Purpose
Vehicle-Miles						
263	9,826	18,141	22,179	6.0	11.1	13.6
144	16,244	19,348	22,412	10.0	11.8	13.7
056	22,160	27,551	31,601	13.6	16.9	19.4
Vehicle-Hours						
263	411	692	839	6.4	10.8	13.0
144	637	779	903	9.9	12.1	14.0
056	879	1,021	1,181	13.6	15.8	18.3

Table 15. Cumulative difference by district of vehicle-miles by distribution assignment for 3 test systems.

Percentage Difference Range	System 263		System 144		System 056	
	1-Purpose	4-Purpose	1-Purpose	4-Purpose	1-Purpose	4-Purpose
± 0.00- 2.49	11	12	9	11	5	8
± 2.50- 4.99	19	23	15	21	14	17
± 5.00- 7.49	29	31	26	28	20	22
± 7.50- 9.99	37	39	33	33	26	29
±10.00-12.49	39	46	35	38	32	37
±12.50-14.99	43	49	40	45	36	39
±15.00-19.99	50	53	47	50	44	46
±20.00-24.99	53	55	52	53	47	49
±25.00-29.99	54	55	53	54	50	50
±30.00-34.99	55	55	54	54	51	51
±35.00-39.99	55	56	54	54	52	53
±40.00 and more	56	56	56	56	56	56

ment. This was to be expected because the trip distribution process generated trips for many cells of the matrix where the trip frequency is too small to get complete coverage in an O-D sample survey. There were also improvements relative to the O-D assignment for links with volumes of more than 30,000.

Figures 7 and 8 show assignments resulting from the 1-purpose and the 4-purpose distributions for systems 263 and 144. The line-printer graphs shown in these figures illustrate the high degree of correspondence between the 2 distributions for all link-volume ranges.

TRIP DISTRIBUTION TEST FINDINGS

The study established that trip distribution models can readily be calibrated for analysis systems with widely varying numbers of zones. One-purpose and 4-purpose models were developed that reproduced the observed trip patterns within acceptable limits. It was found that the friction or impedance functions, which reflected the bahavior of tripmakers with regard to travel time, were not affected by zone aggregation, except for the intervals that applied to intrazonal trips. Even so, these trips did not cause any particular problem in the calibration process, provided that the intrazonal travel times for large zones were carefully determined. Even a small change in intrazonal times (e.g., 1 min) can have a significant effect on the number of trips leaving and entering a zone when 50,000 or more trip ends need to be distributed.

The investigation evaluated assignment results from 1-purpose and 4-purpose trip distributions. It found that the rms errors resulting from both trip distributions for links crossing 16 check lines were only marginally larger than the errors from the assignment of O-D survey data. The errors ranged from 2 to 10 percent, the smallest errors being associated with the least detailed zone plan and the largest errors with the most detailed zone plan. The errors of 1-purpose distribution assignments were consistently larger than those of 4-purpose assignments for each test system; however, the increase in error was much less than the increase in error over the O-D assignments.

The analysis of vehicle-miles and vehicle-hours traveled in 56 districts indicated similar results although, with the exception of system 263, the increase in errors from the 4-purpose distribution to the 1-purpose distribution was similar to the increase in assignment errors between the O-D survey and the 4-purpose trip distribution. Close analysis indicated, however, that most large errors were due to the effects of the zone plan rather than the trip distribution process.

In summary, the findings indicated that reductions in the number of zones did not appreciably diminish the accuracy of trip distributions. Further, 1-purpose trip distributions were only very slightly less accurate than separate distributions of several trip purposes. This finding is of major importance because of the substantial reductions in computer time and analysis effort that would result from the use of a 1-purpose distribution model.

CONCLUSIONS

The study has determined the levels of accuracy in traffic assignments and trip distributions that may be expected from the use of fewer analysis zones than normally used in conventional transportation studies.

Assignment tests with a total vehicle O-D survey trip matrix have identified the level of accuracy in traffic assignments for 5 widely varying zone plans, thus providing evidence to transport planners for the selection of zone numbers according to the objectives of their studies. For major route-planning purposes, adequate traffic assignments should be obtained from zone plans with an average of 10,000 to 15,000 trip ends per zone. For predictions of traffic growth within transportation corridors or segments of urban areas, zone plans with as many as 30,000 trip ends per zone should yield traffic assignments with sufficient accuracy.

The comparison of O-D traffic assignments with trip distribution output matrices indicated that the accuracy of the trip distribution process was not affected significantly by reductions in the number of analysis zones. Thus, decisions regarding the design

Figure 7. Comparison of 4-purpose and 1-purpose distribution assignments for system 263.

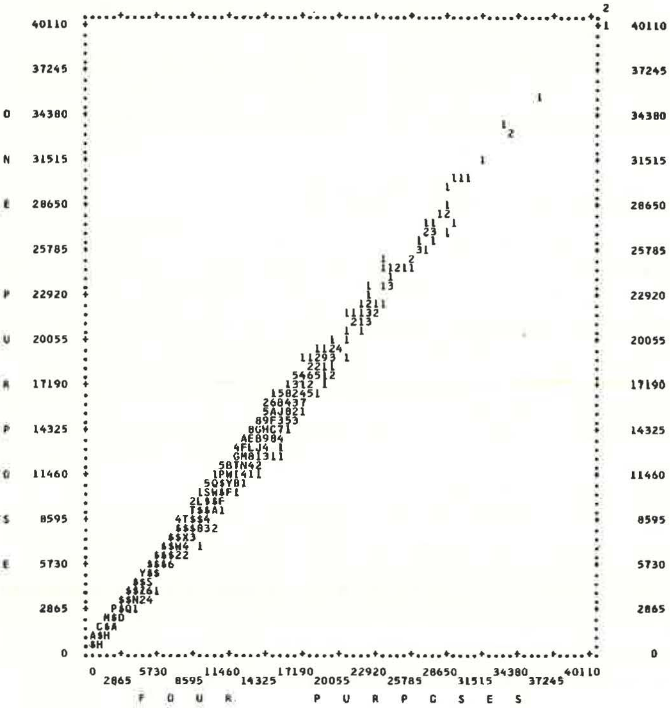
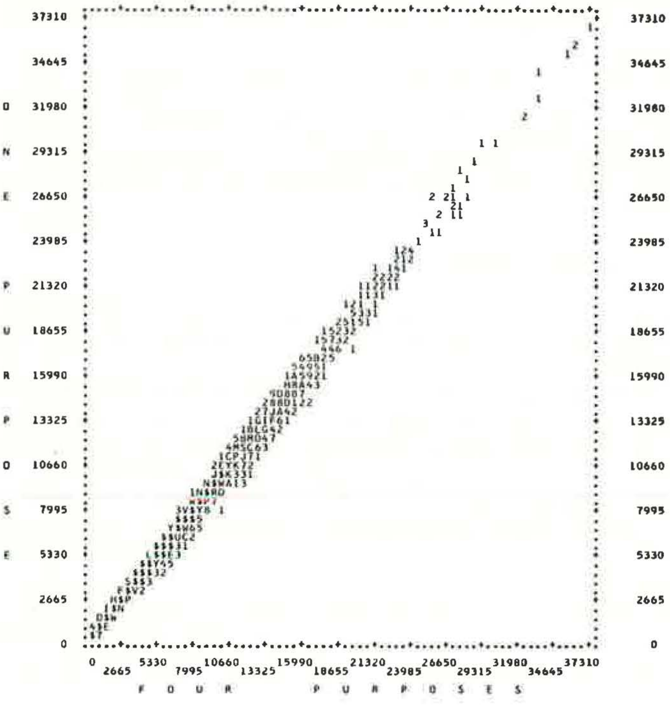


Figure 8. Comparison of 4-purpose and 1-purpose distribution assignments for system 144.



of zone systems should be based primarily on the level of acceptable accuracy for the traffic assignment phase alone. For the trip distribution phase of a study, the use of a 1-purpose trip distribution with partition of the matrix offers significant savings in computer processing and leads to little reduction in accuracy compared to the usual distribution of several trip purposes.

The data on traffic assignment errors presented in this study should serve as a guide for the selection of optimal zone size for transportation planning studies. The techniques developed are appropriate for testing the potential for aggregating the initial zone system of completed studies for use in continuing planning. Generally, zone aggregation offers significant reductions in the effort required to prepare input data and to analyze transportation study results. Particularly important savings may also be achieved in developing projections of land use activities for trip estimation. This could be welcomed by land planners, who prefer to work with larger zones than are normally used in transportation studies. Substantial savings can also be achieved in computer time for processing trip distributions and assignments.

We conclude from this study that for many aspects of transport planning, where accuracy at the level of individual arterial routes is not essential, standard traffic assignment and trip distribution methods can reliably be applied with substantially fewer zones and fewer trip purposes than used in conventional transportation studies.

In view of the potential savings resulting from the use of fewer zones, further research should be undertaken to investigate the effects of zone size on other aspects of the transportation planning process. For new or repeat studies, consideration should be given to the effect on the accuracy of trip-end estimations. Further work is also required to assess the effects with capacity-restraint assignments and with peak-hour assignments. Changes in the normal approach to assignment may prove feasible, so that individual route accuracy can be achieved while the advantages of large zones are retained. For example, intrazonal trips could be allocated uniformly to links within a zone as part of the assignment process. More fundamental innovations in traffic assignment may follow if it is appreciated that large numbers of zones are not necessarily basic to achieving acceptable levels of accuracy. This could open up new horizons for testing alternatives, for stage development planning, for economic evaluation, and for properly integrating road and public transport planning, if the present transport planning process is made less cumbersome, less time-consuming, and less costly.

DEVELOPMENT AND APPLICATION OF DIRECT TRAFFIC ESTIMATION METHOD

Yehuda Gur, Department of Systems Engineering, University of Illinois at Chicago Circle, and Chicago Area Transportation Study

The paper describes the following major problems encountered in the process of programming and testing the direct traffic estimation method: how to estimate contributions to volumes of individual zones, how to deal with assymetric networks, how domain overlap occurs and can be identified, and how to improve results by smoothing the boundaries between main and prime domains. The major conclusions of the first-stage testing of the model are as follows: (a) The model produces reliable expressway volume estimates that may be used for planning and analysis; (b) currently, it is not recommended that the model be used for estimating arterial volumes, and, in case such a use is made, the range of possible errors should be kept in mind; and (c) application of the model gives an excellent method for describing the probable sources of traffic on a given link and the function of the link in serving the travel demand as a part of the network. Visual representation of the results using the available auxiliary programs is of special help.

•WITHIN the framework of urban transportation planning, estimates of traffic volumes on major arteries have special importance. They are essential inputs for both planning and design and for evaluation of alternative plans and programs. Traditionally, link volumes on the highway network are estimated through the chain of land use-trip generation-conversion-distribution-assignment (4). This method is aimed at and is relatively efficient for estimating link volumes for the whole region or large parts of it. However, where volumes estimates are required for a small number of links, running the full assignment is quite inefficient.

In cases where information on sources of traffic on particular links is required, the same problem occurs. The assignment package can supply this information but only through a tedious and expensive process. The direct traffic estimation method (DTEM) is efficient in the analysis of separate links, and can be used where such an analysis is needed.

This method was developed by Schneider in 1965 (1). It was computer programmed and applied by the Tri-State Transportation Commission (TSTC) in 1967 (2). In this method, link volumes are estimated by examination of the link itself, not the interchange it serves. The link volume is estimated as a function of the potential of the region to generate traffic on the link and the availability of parallel links to serve this traffic. The method also gives a measure of the relative importance of different zones in the region as sources for traffic on the link. Analysis of each link is done completely independently of analysis of other links.

After careful examination of the program being used by TSTC, it was found that the use of this program by the Chicago Area Transportation Study (CATS) would require major changes either in the CATS data system or in the program. This was due to differences in the geographical identification system and network coding rules used in the 2 studies. Since DTEM was intended to be used by CATS, together with the existing assignment program, it was decided to reprogram DTEM so that it would accept the existing CATS assignment input tapes.

A program was written in SIMSCRIPT for the CDC 6000 series computers. (Currently, a FORTRAN version of the program is being written.) Current size limitations are 19,999 nodes maximum, 9,999 zones maximum, and

$$CM = 12{,}000 + N^{*} \{[2 + (C/2)]\} + 2Z$$

where

CM = central memory size (words) (50,000 for the CDC 6400 and 130,000 for the CDC 6600),
N = number of nodes,
Z = number of zones, and
C = maximum number of links connected to 1 node.

Together with the main program, a system of auxiliary programs, mainly in FORTRAN, was prepared. These programs enable efficient calibration, preparation of input, printing of summaries, and plotting of output. A users' and reference manual was also prepared for these programs.

THE THEORY

The theory of the direct traffic estimation method was developed in another paper (1). (In the Appendix some theoretical difficulties are discussed, and modifications to the original theory are suggested.) Here it is intended to give only the logic of the method, definition of terms, basic assumptions, and major results of the theory. This section also describes a number of problems in application of the theory and the way these problems are dealt with.

Terminology

Figures 1 and 2 show graphical description of some of the following terms:

1. Point of interest (POI). A point on the road network (not an intersection) whose volume is being estimated (Fig. 1).
2. Link of interest (LOI). The link on which the POI is located. For convenience in reference, the direction of the link is designated as north-south.
3. Main nodes. The nodes at the ends of the link of interest.
4. Turning links. The links connected to the main nodes, excluding the main link. The turning links end with the turning nodes.
5. Path value (F_{ij}). The generalized cost of travel between 2 points, i and j. It is defined as the value of the minimum path between the points on the coded network.
6. Main boundary. A line through the POI, approximately perpendicular to the link of interest, that divides the region into north and south domains. The path value from the POI to any point on the main boundary is the same going north or south.
7. Domain (d). A definite area within the region (Fig. 2).
8. Main path. The minimum path from the POI (to a point).
9. Prime path. The minimum path from the main boundary (to a point).
10. Main domains (n, s). The north domain (n) is part of the region that is most easily reached from the POI going north. (That is, the best path from the POI to any point in the north domain passes through the main node.) Similarly, the south main domain is defined. Hence, the main boundary may be defined as the indifference line between the northbound and the southbound main paths (Fig. 2).
11. Prime domains (n′, s′). The prime domains are parts of the main domains (n, s) for which the main path is also the prime path.
12. Decay function G(·). A decreasing function of the friction between 2 points.
13. Domain integral (Id).

$$Id = \sum_{j \in d} G(F_{oj}) \times V_j \tag{1}$$

where

F_{oj} = path value between zone j and POI, and
V_j = number of trip ends at zone j.

The summation is done over the finite area d (a zone, subregion, or region).

Assumptions

1. Complete symmetry exists in volumes and path values throughout the systems; i.e., $Q_{ij} = Q_{ji}$, where Q_{ij} is 1-way volume between i and j, and $F_{oj} = F_{jo}$, where F_{oj} is directional path value between POI and zone i.
2. Each trip end sends and receives exactly 1 trip daily.
3. A northbound vehicle at the POI has its origin in the south domain and its destination in the north domain.

(In all statements following, northbound can be replaced by southbound by replacing n by s, s by n, n′ by s′, and so on.)

4. A northbound trip originating in s′ is defined as free. Its destination may be anywhere in the north.
5. A northbound trip originating out of s′ has by definition a prime path better than the main path. If it uses the link of interest, it implies that its destination is in n′. Such trips are called fixed.
6. The relative possibilities of trips through the POI terminating in different domains are proportional to the domain integrals. That is,

$$P\{D = j \mid D \in d,\ j \in d\} = I_j/I_d \tag{2}$$

where D is the destination of the trip.

Calculation of Volumes

Based on the previous assumptions, the daily 2-way volume through the POI has been found to be

$$(I_n' \cdot I_s + I_s' \cdot I_n)/(I_n + I_s) \tag{3}$$

Partial Volumes

Equation—Let d be a subdomain of the north domain, so that $d \in n$, $d' \in n'$, $d' \in d$, and $d' \geq \phi$. The total 2-way volume from this domain through the POI is

$$Q = (I_d \times I_s' + I_d' \times I_s)/(I_n + I_s) \tag{4}$$

If we sum up the partial volumes to all the subdomains in the north domain, we get

$$Q = [1/(I_n + I_s)] * [I_s' \cdot I_n + I_s \cdot I_n'] \tag{5}$$

This equation is used to calculate turning volumes and the contribution of each zone in the link volume.

Turning Movements—Clearly, the path from any point in the region to the POI passes through exactly 1 turning link (as defined earlier). Hence, the subdomains (parts of the prime and the main domains) that belong to each turning link may be found, and the turning movement calcuated by using Eq. 5.

Volumes for Individual Zones—Equation 4 may be used to find 2-way volumes from individual zones through the POI. As usual, a zone is represented on the network as a point (load node) with a given number of trip ends on it. In this case, the complete zone belongs or does not belong to a prime domain. This might cause some inaccuracies in the estimation of the distribution of individual zones to the links volume. In order to partly overcome this problem, the program enables a breakdown of zones between the prime and nonprime domains. In this case, a zone may belong partly to the prime domain. Equation 4 covers the 2 cases.

Asymmetric Networks

Clearly, the assumption of symmetry of volumes and path values should be approximately true in order to allow successful application of this method. This is especially important near the POI. Hence, estimation of directional peak volumes or of volumes on 1-way links in the network is impossible. On the other hand, inclusion of 1-way links in the network is allowed as long as they do not cause substantially different path values to and from the POI.

Direction of Paths

Theoretically, the path values between any 2 points are the same in both directions. Unfortunately, 1-way links do exist and cause a number of small problems that are described below.

One-Way Turning Links—The main tree is built from the POI outward (Fig. 3). Therefore, if a turning link is 1 way inward (case 1), the calculated turning volume on this link will always be 0. If a turning link is 1 way outward (case 2), then its calculated volume may be positive.

One-Way Boundary Links—The main boundary is defined by a building a tree from the POI. The main domain of each node is determined according to the main node it is connected to. For the 2-way boundary link A-B, the following relation always holds:

$$| F_{oA} - F_{oB} | \leq T_{AB} \qquad (6)$$

where F_{oI} is the path value of I from POI, and T_{AB} is the link friction. If the boundary link is 1 way, this is not always true. In this case, it may happen that the boundary passes around a 1-way boundary link.

Sensitivity of Volume to Path Values

Errors in path values from the POI to different zones appear due to errors in coding the network or in link friction values. These errors may be classified as either systematic or random errors. They may affect the estimated volume by changing the domain integral values of zones or by transferring zones between domains.

Random Errors—In the case of random errors in link friction values, the errors due to changing zone values are generally very small. The errors due to transferring zones may be quite large. These errors usually consist of a transfer of a zone in or out of the prime domains. This happens quite often, because generally the path value from the POI to a zone is never much smaller than the best alternative path from the main boundary. A small increase in the "main path" value may easily cause a zone to leave the prime domain. Errors due to transfers between domains are especially serious when only a few zones are included in the prime domain (in this case, the inclusion of any zone makes a large percentage difference in the calculated volume) and when one of the main domains is much larger than the other (a common case when the link is located near region boundaries). This problem is overcome by breaking zones between main and nonmain domains as follows:

$$Z' = \begin{cases} 0 & \text{if } Z'' \leq 0 \\ Z'' & \text{if } 0 < Z'' \leq 1 \\ 1 & \text{if } Z'' > 1 \end{cases} \qquad (7)$$

where

$Z'' = 0.5 + k\,[(F'_A - F_{oA})/F_{oA}]$,

Z' = percentage of the zone (i.e., the trip ends in the zone) that belongs to the prime domain,

F'_A = path value on the prime tree,

F_{oA} = path value on the main tree (from the POI), and

k = specified parameter (currently used $k = 0.1/0.2$).

By this method, errors in volume due to misspecification of domains still cause considerable error, but much smaller than before. In the computer program, this formula may be applied through a special optional routine. Experimenting with this method shows substantial improvements in estimation of volumes on minor links with almost no effect on high-volume links. The problem of 0-volume links (due to nonexistence of prime domain) vanished.

Systematic Errors—Systematic errors in the path values cause changes of domain integral values. The danger of such errors exists especially when the network used for calibration is substantially different from the network used for estimation. This problem is discussed later.

Domain Overlap

Sometimes, due to peculiarities in networks, a certain area belongs by definition to the prime domains of 2 parallel adjacent links. This is called domain overlap. An example of such a condition is shown in Figure 4. There, area A belongs to the prime domains of the 2 points of interest I and II. Clearly, the existence of domain overlap is in contradiction with the logic of the theory. It causes overestimation of volumes on the links involved.

The existence and the extent of domain overlap in the analysis of any specific link are practically impossible to predict. They may be checked by plotting the domain boundaries of adjacent links. Generally speaking, areas with irregular networks are more prone to have difficult overlap problems. (Thus, in the Chicago area, which has relatively uniform network configuration, the problem is not especially serious.)

The domain overlap problem may be solved by merging 2 links into 1, estimating the joint volume, and then externally or otherwise redistributing the volume. A detailed study and implementation of correction procedures for this problem are yet to take place.

Concluding Remarks

The direct traffic estimation method uses practically the same inputs as the traditional method: trip ends, decay function, and coded network. Yet, the 2 methods are significantly different in the ways used to analyze the data. The DTEM is advantageous in solving individual links. But, it has a number of disadvantages. The more important are the unavailability of capacity constraints and the inability to apply the method for assymetric loadings. Only careful and continual use of the method will enable its practicality as a planning tool to be evaluated and its most effective place within the set of available travel demand prediction models to be defined.

CALIBRATION AND TESTING

DTEM was calibrated and tested with CATS 1965 data. The major findings are as follows:

1. The negative exponential function (Eq. 8) may be used as the decay function.

$$G(F_{oj}) = \exp(-\beta * F_{oj}) \qquad (8)$$

where

F_{oj} = travel friction between the point of interest and zone j,
β = coefficient, to be determined by the calibration; and
$G(\cdot)$ = decay function.

2. The optimal β for estimation are $\beta(1) = 0.06\ f^{-1}$ for expressways, and $\beta(2) = 0.08\ f^{-1}$ for arterials, where f is the CATS friction measure. (A unit friction represents approximately 24 sec on a free network.)
3. Expressway estimated volumes are close to counted volumes with a standard error of less than 12 percent of the mean.
4. Estimated arterial volumes have a standard error of 47 percent of the mean. The error is correlated negatively with the counted volume and number of lanes.

Method and Analysis

Ideally, the model may be calibrated by analyzing the distribution origins of vehicles traveling on the link of interest. Comparing the actual distribution with the distribution implied by the model (as calculated by Eq. 4) allows both the form of decay function and its parameters to be found and checked. Such an analysis may increase substantially our understanding of urban travel and the performance of the model. This analysis requires data on origins of trips using the link. These data are available only through a roadside interview. Unfortunately, such data were not readily available for the present analysis (and are practically unavailable for expressways because of technical difficulties in collection.)

In the calibration procedure used here, the decay function was chosen externally. Volumes on sample links were calculated by using different coefficients in the function. The optimal coefficients were chosen so as to minimize the sum of deviations between the estimated and the counted volumes. Later, the volumes estimated by using the chosen coefficients were analyzed to find the level of accuracy and the error properties of the estimates.

The Data

The model was calibrated and tested with CATS 1965 data. These data include an updated coded highway network, estimated number of zonal trip ends, and counted volumes on major roads. The volume counts on this set of data were carefully prepared and intensively checked. A sample of 11 expressway links and 22 arterial links was chosen for analysis. The sample covers a wide range of locations, geometry, and link volumes.

The Decay Function

Two decay functions were tested for use: the A function, developed by Schneider (3), and the negative exponential function. The two were shown earlier to fit DTEM (1, 3). In preliminary runs, no advantage of either of the 2 functions was apparent. Hence, it was decided to use the simpler one, i.e., the negative exponential function (Eq. 8).

It should be remembered that it is possible that other functions may prove to be better. Until more analysis is done (preferably by using extensive data from roadside interviews), no final conclusion can be made. However, it should be noted that the negative exponential is the only decay function that is completely consistent with the theory (see Appendix).

Analysis of Results

The basic theory presumes that 1 decay function may be used for all route types. However, it was found both by CATS (Figs. 5 and 6) and by TSTC (2) that the use of 1 decay function causes underestimation of expressway volumes or overestimation of arterial volumes or both. Use of 1 decay function for the 2 cases would require manual adjustment of results. Such adjustments may be avoided, or at least decreased, by a separate analysis of the 2 link types. Such a separation was done in this analysis.

Figure 1. Part of region around point of interest.

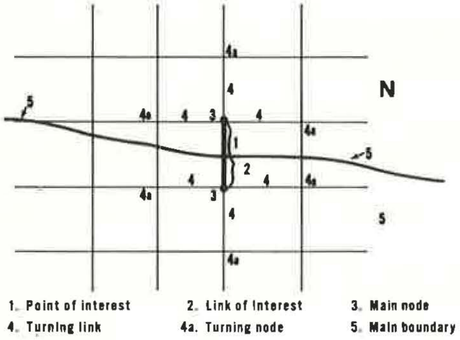


Figure 2. Definition of domain.

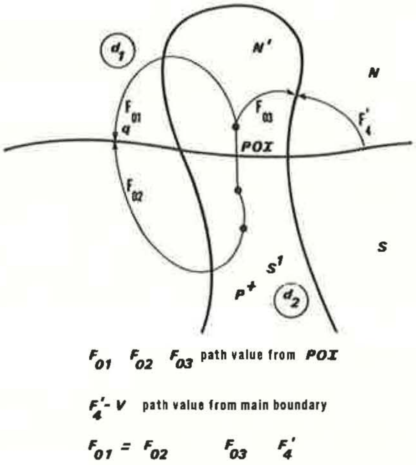


Figure 3. One-way turning links.

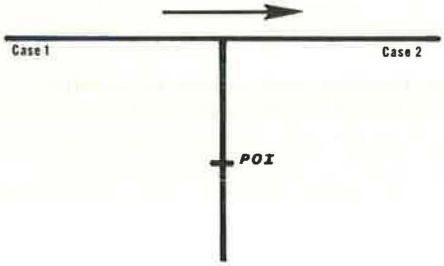


Figure 4. Domain overlap.

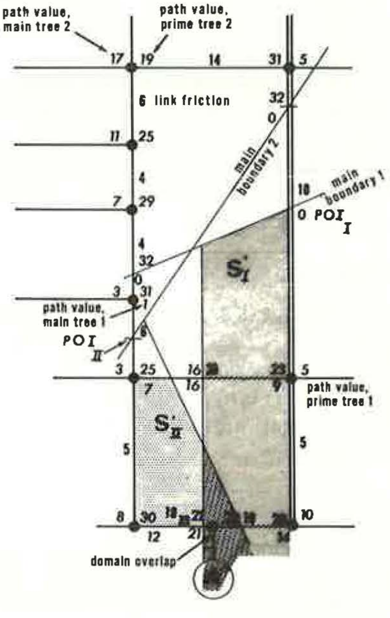


Figure 5. Difference between mean estimated volume and mean counted volume as function of β for expressways.

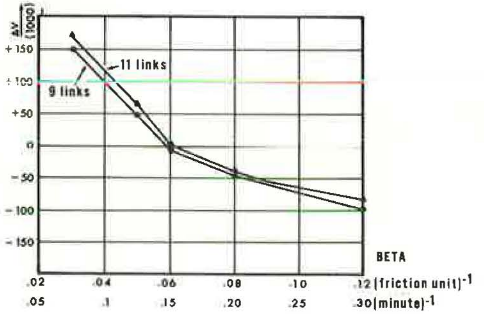


Figure 6. Difference between mean estimated volume and mean counted volume as function of β for arterials.

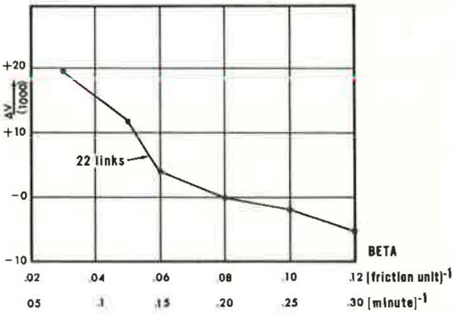


Table 1. Test results.

Item	Expressways		Arterials, 22-Link Sample
	11-Link Sample	9-Link Sample[a]	
Mean volume, vehicles per day, 2-way			
Counted	120,850	123,630	16,240
Estimated	124,200	117,880	16,090
Mean error	3,350	-5,750	-150
Standard deviation of volumes, vehicles per day, 2-way			
Counted	59,200	63,800	9,340
Estimated	52,100	51,800	9,300
Standard error of estimate, vehicles per day, 2-way	23,900	14,600	7,810
Standard error, percent	20.0	11.8	47
Range of volumes, vehicles per day, 2-way	8,600 to 216,600	8,600 to 216,600	200 to 38,300
Range of errors, vehicles per day, 2-way	-17,300 to 59,400	-17,300 to 17,300	-14,000 to 13,100
Correlation of error with counted volumes			-0.42

[a] The 9-link sample is identical to the 11-link sample after 2 links are dropped that show large domain overlap.

Choice of Optimal β—Figures 5 and 6 show that the mean error is quite sensitive to the value of β. This sensitivity is greater in expressways than in arterials. Because of this sensitivity, the choice of β is dictated by the results and is straightforward. Directly from the graphs we get $\beta(1) = 0.06$ (f^{-1}) for expressways, and $\beta(2) = 0.08$ (f^{-1}) for arterials.

Domain Overlap—The original sample includes 11 expressway links. The errors in estimates of these links are given in Table 1. Two of these links were respectively the express and the local lanes of the same section of an expressway (Dan Ryan Expressway on the south side of Chicago). Analysis of the domain boundaries of the links revealed a strong domain overlap (as described earlier). This caused overestimation of the volumes on these 2 links.

Because this situation (of parallel local and express lanes) is unique and identifiable, it was decided to drop the 2 links from further analysis. It is expected that using the means suggested earlier, i.e., estimation of the joint volume on the 2 links, will enable solution of this problem.

Error of Estimate, Expressways—A summary of the error properties is given in Table 1. The following points are of interest:

1. A relatively wide range of volumes was checked.
2. The standard error of estimate is 14,600 vehicles/day, 2-way or 11.8 percent of the mean volume. The maximum error in the sample is 17,300 vehicles/day.
3. The variance of the estimated volumes is lower than the variance of the counted volumes. This may be caused by a tendency of the model to underestimate high-volume links and overestimate low-volume links.

The error of 17,300 vehicles/day (the maximum observed) is less than the capacity of 1 expressway lane. Such a level of accuracy is well within the range acceptable for planning. Although the accuracy of estimates depends also on the errors in other stages of the modeling (network speeds and zonal trip end estimates), it seems that this total approach is relatively reliable for estimating expressway volumes.

Error of Estimate, Arterials—A summary of errors in estimates of arterial volumes is given in Table 1. The following observations are of interest:

1. The range of volume tested is quite high—200 through 38,000 vehicles/day. The results show that further breaking of this group may be helpful.
2. The standard error of estimate is 47 percent of the mean volume or 7,810 vehicles/day.

3. The range of errors is from -14,000 to 13,100 vehicles/day.

4. A significant negative correlation (-0.42) was found between the error and the counted volumes. This shows, as expected before, that the model underestimates high-volume links and vice versa.

Although this level of accuracy may be sufficient for certain purposes in transportation analysis, it is by no means satisfactory. More work on the model is required in order to improve these estimates.

Network Speeds—Multiplication of link friction values by a constant is equivalent to multiplying by the same constant. Seemingly, a systematic increase or decrease in link friction values is equivalent to these multiplications. Such systematic changes in friction occur while a transfer is made among theoretical, free, loaded, and congested networks. (These are not accurate terms but refer to the travel friction under different link loadings.) Because of the high sensitivity of the volumes to β (as shown in Fig 5), it is important that, in both the network used in the calibration stage and the network used for prediction, the friction corresponds to the same level of load on the network. The lack of capacity constraints in DTEM will always cause certain inaccuracies in volume estimates. However, systematic bias, which may be caused by using 2 differently loaded networks, is at least as erroneous.

Correlation Analysis of Errors—The errors in the arterial volume estimates were analyzed in order to find the reasons for the high errors and, possibly, to find ways of correcting them. The following results are of interest:

1. Negative correlations exist between the error and the counted volumes, capacity class, or number of lanes. This fact may be explained, at least partly, by lack of capacity constraints. It is reasonable to assume that, given the increase in link friction due to increase in volume, the prime domains of the small links will decrease and cause the total volume on these links to go down faster than on other links.

2. The error is negatively correlated with distance to the nearest parallel expressway.

Correction for these 2 variables through the regression equation causes a decrease in the standard error of estimates of approximately one-half. Nevertheless, it is not recommended that these corrections be used without further study of the causes of errors in the model.

Conclusions

1. DTEM produces reliable estimates of expressway volumes. These estimates may be used for planning and analysis.

2. Currently, it is not recommended that DTEM be used for estimating arterial volumes. In case such a use is made, the range of possible errors should be kept in mind.

3. Application of the model gives an excellent method for describing the probable sources of traffic on a given link and the function of the link in serving the travel demand as a part of the network. Visual representation of the results using the available auxiliary programs is of special help.

Recommended Further Research

Some theoretical problems and directions for research are suggested elsewhere (3 and the Appendix). In this section, concentration is on operational research in the framework of the existing theory.

1. It is suggested that the decay function be carefully examined by using data from roadside interviews when and where such data are available.

2. An intensive examination of error properties of arterial estimates should be made. Large reductions in the existing error levels are very likely.

3. A procedure that enables merging of links with overlapping prime domains into 1 equivalent link should be developed and tested. Such a procedure may be included within the program by a separate auxiliary subroutine.

4. This model is the first one used by CATS in which absolute path values and not relative values only are used for predicting link volumes. A comparative study of scaling of the 3 main networks (1965, current, and 1985) should be carried out in order to ensure compatibility among the networks.

5. Use of the available graphic output for production of predicted flow diagrams on major links should be studied.

It is expected that a large part of the existing problems in the model may be solved by conducting the research recommended here. Besides the solution of the specific problems, this research may serve as an excellent tool for improving understanding of urban travel characteristics.

REFERENCES

1. Schneider, M. Direct Estimation of Traffic Volume at a Point. Highway Research Record 165, 1967, pp. 108-116.
2. Brown, G. J. H., and Woehrle, A. H., Jr. Program, Inputs and Calibration of the Direct Traffic Estimation Method. Highway Research Record 250, 1968, pp. 18-24.
3. Schneider, M. Access and Land Development. HRB Spec. Rept. 97, 1968, pp. 164-177.
4. Final Report. Chicago Area Transportation Study, Vols. 1, 2, and 3, 1962.

APPENDIX

MODIFICATION OF THE VOLUME EQUATION

Schneider developed the volume formula as follows (1):

1. A hypothetical trip end is put near the point of interest. It is shown that

$$Q = 1/(P_n + P_s) \qquad (9)$$

where P_n and P_s are respectively the probabilities that any northbound or southbound vehicle at the POI will choose the hypothetical trip end as its destination. Q is the daily 1-way volume through the POI.

2. "Free" vehicles are defined as vehicles that begin their trips inside the prime domain. The remaining vehicles are called "fixed." The proportions of the free vehicles (called A_s and A_n) in the southbound and northbound traffic streams are found. This is done by equating the number of trips originating in the south prime domain with the number terminating there (1, Eqs. 12, 13, and 14).

3. The probabilities P_n and $\bar{P}_s$ are later calculated as the weighted average of the probabilities that either a fixed or a free vehicle will occupy the hypothetical destination (1, Eqs. 9, 10, and 11).

4. P_n and P_s are substituted in Eq. 1 to get the solution (1, Eq. 16).

In the original representation of the direct traffic estimation method, Schneider (1) found the 1-way daily volume on a link to be

$$Q_1 = [(I_n - I_s)/(I_n + I_s)]\,[1 - (1 - r_n) \cdot (1 - r_s)] \qquad (10)$$

where

I_d = the domain integral of d,
n, n′, s, s′ = 4 domains,
$r_n = I_{n'}/I_n$, and
$r_s = I_{s'}/I_s$.

Or, substituting the r's gives

$$Q_1 = [1/(I_n + I_s)]\cdot[I_n' I_s + I_s' I_n - I_s'] \tag{11}$$

Brown and Woehrle (2), in applying the method, give the formula for calculation of 2-way link volume as

$$Q_2 = [1/(I_n + I_s)]\cdot[I_n' I_s + I_s' \cdot I_n] \tag{12}$$

They refer to Schneider (1) as the source of their formula but give no explanation for the differences. Schneider himself suggested (1, p. 115) that modifications to his Eq. 16 were required and were being implemented but did not specify why and how they were to be done.

The analysis here results in a third volume formula as shown below.

Choice of Destination and Route by Northbound Free Trip

A typical situation is presented and is shown in Figure 2.

A domain d_2 is located in the south prime domain. A point p is found so that every trip from d_2 to the north has to pass through it. Here, the choice of the exact destination and the route of such a trip are discussed. Because all the vehicles going to the north from d_2 pass through point p, it will be used as a point in which the trips are observed, ensuring that no trips are eliminated from consideration. Equation 13 gives

$$P\{D \in d\} = I_d^*/I_n^* \tag{13}$$

where D is the destination of the trip, d is some domain in the north, and I* is the domain integral with respect to p.

Clearly, at p, the trip is completely free to choose a route. If d is in the north prime domain, it is very likely that the trip will use the path through the POI. When d is not in the north prime (for example, $d = d_1$ in Fig. 1), other paths may be considered. In particular, the possibility of crossing the main boundary at q (Fig. 2) cannot be ignored. q is the point on the main boundary nearest to d_1. This possibility should be accounted for because it is considered a valid choice while the symmetric case is analyzed (q is the POI, trip originates at d_1). Without unnecessary complications, it may be assumed that the probabilities of crossing at the POI and at q are each equal to ½.

Putting this in probability notation gives the following:

1. For any trip originating in s′ and going to n,

$$P\{X = POI \mid D \notin n'\} = 1/2 \tag{14}$$

$$P\{X = POI \mid D \in n'\} = 1 \tag{15}$$

$$P\{D \in d\} = I_d^*/I_n^* \tag{16}$$

where X is the crossing point of the main boundary. Taking d = n′, gives

$$P\{D \in n'\} = I_{n'}^*/I_n^* \tag{17}$$

$$\begin{aligned} P\{X = P\emptyset I\} &= 1/2 \cdot (1 - I_{n'}^*/I_n^*) + I_{n'}^*/I_n^* \\ &= (I_n^* + I_{n'}^*)/(2 \cdot I_n^*) \end{aligned} \tag{18}$$

2. For those trips that also pass through the point of interest,

$$\begin{aligned} P\{D \in n'\} &= (I^*_{n'}/I^*_n) \cdot P\{X = \mathrm{POI}\} \\ &= (2 \cdot I^*_{n'})/(I^*_{n'} + I^*_n) \end{aligned} \tag{19}$$

In case the decay function being used is the negative exponential, it always gives

$$I^*_{n'}/I^*_n = I_n'/I_n$$

and Eq. 19 can be modified into

$$P\{D \in n\} = (2 \cdot I_n')/(I_n + I_n') \tag{20}$$

without any loss of generality. For convenience of notation, define

$$\begin{aligned} r_n &= (2 \cdot I_n')/(I_n + I_n') \\ r_s &= (2I_s')/(I_s + I_s') \end{aligned} \tag{21}$$

Volume Formula

Using the steps described in the preceding section gives

$$\begin{aligned} Q_n A_n &= Q_s(1 - A_s) + Q_s r_s A_s \\ Q_s A_s &= Q_n(1 - A_n) + Q_n r_n A_n \end{aligned} \tag{22}$$

(parallel to 1, Eqs. 12 and 13).

The solution of this system (clearly, $Q_n = Q_s$) gives

$$\begin{aligned} A_n &= r_s/[1 - (1 - r_s) \cdot (1 - r_n)] \\ A_s &= r_n/[1 - (1 - r_s) \cdot (1 - r_n)] \end{aligned} \tag{23}$$

(parallel to 1, Eq. 14.)

$$\begin{aligned} P_n &= I_o\{A_n[2/(I_n + I_n')] + (1 - A_n) \cdot (1/I_n')\} \\ P_s &= I_o\{A_s \cdot [2/(I_s + I_s')] + (1 - A_s) \cdot (1/I_s')\} \end{aligned} \tag{24}$$

(parallel to 1, Eqs. 10 and 11).

Putting Eq. 23 into Eq. 24 gives

$$\begin{aligned} P_n &= [2I_o/(I_n + I_n')]/[r_n + r_s - r_n \cdot r_s] \\ P_s &= [2I_o/(I_s + I_s')]/[r_n + r_s - r_n \cdot r_s] \end{aligned} \tag{25}$$

To conclude, take $I_o = 1$ and put Eq. 25 into Eq. 12, which gives

$$Q_3 = [I_n \cdot I'_s + I_s \cdot I'_n]/[I_n + I_s + I'_n + I'_s] \tag{26}$$

where Q_3 is the 1-way volume through the POI.

Some Observations

A major shortcoming of this approach is the necessity of analyzing the problem at the origin instead of at the POI. But this procedure seems necessary to make the theory of DTEM completely valid. Because of this procedure, it is necessary to use the negative exponential function as the only possible decay function (for the crucial passage from Eq. 19 to Eq. 20). Only the negative exponential function has a "complete lack of memory" required to assume that the residual trip length distribution is always equal to the total trip length distribution.

TOWARD AN EFFICIENT HIGHWAY SERVICE ESTIMATOR FOR CONGESTED NETWORKS

Hubert Le Menestrel and Bruno R. Wildermuth, Wilbur Smith and Associates,
Columbia, South Carolina

This paper presents a new approach to obtain estimates of highway services on congested networks. Specifically, a 1-pass incremental capacity-restraint procedure is detailed that generates logical paths for reuse with the average network to obtain travel-time and travel-cost estimates. The method is based on the multiple-routing principle for path generation and Schneider's 1-pass capacity-restraint technique. The operation of the procedure is illustrated with an example, and results of several analyses are also presented. Suggestions are made for further research to establish reliable procedures for the estimation of average speeds and to investigate further the validity of the capacity-restraint paths in relation to the loading sequence used by the procedure.

•THE RAPIDLY growing expenditures for urban facilities at all levels of government have been paralleled by increased competition for public funds and by public awareness of the need to set community goals and objectives within which the full range of possible alternatives must be evaluated. Consequently, more demanding evaluation and decision criteria need to be established to facilitate complete consideration of the impacts of alternative transportation systems on the total urban system.

The transportation planning process provides the basic structure with which the characteristics of alternative networks can be measured and their respective abilities to satisfy objectives and goals evaluated. It provides the travel demand forecasts that serve to evaluate the potential cost and benefit patterns. However, this process is often limited and cumbersome in its capabilities for estimating service aspects of alternative transportation systems.

Within the past 15 years, urban transportation planning has relied increasingly on mathematical models and electronic computers as its basic tools. During this period, both the models and the computer systems have steadily been improved, although advancements in the design of computers have far outstripped those of analytical models.

Primary emphasis in the development of models and associated computer programs has been on analysis and estimation of travel demands and capacity evaluations. Increasingly, added emphasis is also placed on the modal-choice decision process and systems evaluation procedures.

A recent report (1) identified the following means that should be provided by transportation systems evaluation procedures:

1. Measuring and valuing the effects of transportation system changes on travelers and on community residents;
2. Estimating the system costs of transportation changes; and
3. Relating transport changes to high-level goals and to overall framework for decision-making.

To operate efficiently, service estimators and other systems evaluation models must be designed to accept travel demand projections and network descriptions from other transportation analysis models. Furthermore, their outputs must be structured for complete compatibility with other models to facilitate the explicit incorporation of service variables into trip generation, distribution, modal-split, and assignment models. Only then will it be possible to supply these models with measures of transportation services (e.g., travel times and travel costs), which are consistent with the anticipated levels of network congestion.

To date, major emphasis in evaluating alternative transportation systems has been on travel benefits. Estimates of transportation systems services, therefore, have been limited largely to measures of place-to-place travel times. Such estimates are also used most often to provide inputs and feedbacks to many travel demands and modal-choice projection models.

Not infrequently, however, analysts have ignored the feedback effects of congestion on the generation of demand and the projection of modal-choice decisions. The main reason for ignoring these feedback effects undoubtedly are the cumbersome and costly efforts required to obtain service estimates from existing analysis tools. Existing packages for urban transportation planning provide a number of programs that, when properly sequenced, can produce estimates of zone-to-zone travel times and travel costs for congested networks.

For highways, the programs required are capacity-restraint, link-costing, and skim-impedance-path programs. Capacity restraint in itself requires a series of 3 programs to be run in repeated succession. However, application of the capacity-restraint series of programs will only produce the data that, when averaged, will reflect the network conditions under the given level of congestion. This network can then serve to determine minimum-impedance paths from which zone-to-zone travel times and travel costs can be estimated.

The remainder of this paper describes a fast and efficient method that produces reliable estimates of zone-to-zone travel times and travel costs for congested highway networks. This procedure, which is easily operated at a fraction of the cost of other methods, has been tested on a number of highway networks for different urban areas.

METHODOLOGY

To obtain estimates of highway services for congested networks requires an efficien' capacity-restraint method. The method presented here produces, with 1 application and without iterations, an average network description (reflecting the level of congestion) and a set of valid "best path" trees along which zone-to-zone travel times and travel costs can be accumulated from the average network. The technique is based essentially on 2 concepts: a recently developed "multiple-routing" best path procedure that utilizes a stochastic approach and a modified version of Schneider's incremental 1-pass capacity-restraint model.

Multiple-Routing Principle

A procedure has been devised that eliminates the most critical problem of the minimum-path procedure and routes directional paths between pairs of nodes over all routes whose times differ from that of the shortest route by less than a given proportion. This procedure, called multiple routing, has made it possible to ensure proper balance of flows between alternate routes of equal merit. Theoretically, the multiple-routing principle allows the directional path between any pair of nodes to be different for each path generated. In practice, of course, the number of different paths is limited by the acceptable alternatives available within the network configuration. Multiple routing assumes that drivers associate a "perceived" travel time with each link of the network and that not all drivers will necessarily perceive the same time for a link.

Stochastic Approach—The multiple-routing principle is based on a stochastic approach that adds a long-sought element of realism to traffic assignments. Use of the stochastic process for the multiple-routing principle is limited to the generation of perceived link times that are drawn at random from a normal distribution of times for each link. Thus, maximum use can be made of existing concepts by relying on the fact that in the average network each pair of nodes will be traversed many times because of the large number of centroids normally used.

This approach requires very little additional computer time because only 1 tree is generated for each centroid. The only extra time is that required to calculate each time a link is being considered, a random sample from a normal distribution with a mean equal to each link's given impedance value.

Because most routes consist of a string of links, it is not possible for the stochastic process to generate unreasonable routings. The sum of perceived link times will differ on the average from the sum of mean link times by a smaller percentage for long routes than for short routes with few links. This corresponds well with the real world. For example, a driver might choose a route that takes 7 min, when the best available route takes only 5 min, and thus accept a 40 percent increase in time. On the other hand, it seems much less likely that anyone would choose a route that takes 70 min when the best route available takes only 50 min. Clearly, the multiple-routing principle reflects likely actual route choices by drivers.

Effect of Alternate Routes on Link Flow—In the application of the multiple-routing principle as presented here, all trips from 1 centroid to another centroid will use the same path; in effect, this is the all-or-nothing principle. However, not all trips passing from 1 node to another will use the same path, provided, of course, that the network consists of a reasonable number of centroids. The incidence of multiple-source centroids, therefore, serves to distribute the traffic flow onto all acceptable alternative routes. In the method described here, the route chosen for a particular trip may to some degree be arbitrary, in the sense that it is determined by a process of chance. The total traffic flows, however, on each link are clearly insensitive to the stochastic process.

One-Pass Incremental Capacity-Restraint Procedure

Previous research and experience with capacity-restraint procedures have indicated that to achieve stable speed-flow relations in congested networks requires either an incremental procedure or an average of the results of several iterations with the all-or-nothing method. Stability of speed-flow is attained when the use of one or more iterations or increments would not significantly alter the flow or the speed of any link in the network. The application of the multiple-routing method of building best paths will reduce the amount of imbalance generated for each assignment step. Hence, a reduction in the number of assignment increments or the number of iterations can be justified.

Alternatively, in combination with multiple routing, it is possible to devise a more efficient way of selecting increments of the total traffic demand for assignment at each step. The 1-pass incremental procedure based on Schneider's basic approach presents such an alternative.

Schneider's Assignment Concept—Schneider's basic philosophy was founded on the assumption that each unit of traffic seeks to travel along the best available path. Furthermore, the addition of each unit of flow to a link would, in effect, reduce the speed of travel on that link and, therefore, create a different best path situation for each additional unit of traffic demand desiring to move through the network.

The method, originally developed for the Chicago Area Transportation Study (2) combined trip distribution and traffic assignment into a single comprehensive process. Only the assignment concept, which substantially reduces the sometimes troublesome effects caused by changes in the loading sequence, is used here. It is a 1-pass incremental process whereby the fraction of the total traffic demand loaded during each increment is the total demand from 1 origin centroid.

Trips from the first origin centroid in the randomly selected loading sequence are loaded on the minimum-path tree determined on the basis of the input travel times. Link times are then updated according to an exponential time versus volume-capacity function. The travel demand from the next centroid in the random sequence is then loaded along the minimum path determined from the revised network times. This process is repeated until the entire traffic demand matrix has been assigned. A randomly selected loading sequence is used to minimize the effects of restraint on routings and travel times for traffic of the same geographic origin.

An Approach Using Multiple Routing—Schneider's incremental method was efficient in its use of only 1 tree per origin centroid. However, it required an excessively large number of network modifications (1 per origin centroid). Apart from conceptual reasons, the frequent modifications served to reduce the problems associated with the use of minimum-path, all-or-nothing procedures. Multiple routing has inherent balancing qualities that allow, in combination with the incremental 1-pass technique, a substantial reduction in the frequency of network modifications without loss of stability.

The increment of traffic demand that can be allocated between successive network updates, without excessively increasing the flow on any 1 link, can readily encompass traffic demands from several centroids as long as drastic changes in travel times are being avoided. The probability of any link being used by 2 or more successive trees is very small if a random-loading sequence is used because each origin centroid is normally separated spatially from the preceding and the following centroid. In addition, the characteristics of multiple routing further reduce the probability that successive trees will use the same path between pairs of nodes.

As the flow on many links reaches the practical capacity level, it is more important to load smaller fractions at each increment. This is readily controlled by selecting fewer origin centroids from the predetermined sequence. The total number of network modifications, which account for the increase in computer time above that of the all-or-nothing technique, is reduced by a factor of 10 or more compared to Schneider's original concept.

Speed-Flow Relation—Schneider used an exponential function to modify travel times. This assumption does not conform to actual speed-flow relations observed on highways but yields increasingly lower speeds with increasing volume-capacity ratios. A "real-life" speed-flow function, on the other hand, would approach a vertical tangent at maximum attainable capacity. The restraint function, however, must not be allowed to preclude altogether the assignment of flow beyond the maximum capacity level. This is because one of the functions of traffic assignments is to identify potential capacity deficiencies within travel corridors of transportation networks.

The function used must first of all satisfy the principle of the method rather than direct itself to a precise simulation of the real world. Links that are loaded with a volume reaching capacity after only a small percentage of the total demand has been processed require more drastic modifications than links that reach the same volume-capacity ratio at the end of the loading process. In short, the model must anticipate the effects of the additional demand still to be loaded and react in a way that will prevent (to a certain extent) its volumes from exceeding its capacity by 2 or 3 times.

Stability and Convergence of Procedure—The 1-pass incremental method permits estimates of speeds and flows in congested networks with a minimun of computer time. In addition, it avoids the nonconvergent unbalanced estimates of flows that characterize iterative methods, as illustrated by the following example.

Figure 1 shows a portion of a larger network in which many trees are routed between points A and B. Three alternative paths exist between A (a bridge head) and B (the main access point to a major center). Path 2 is an old road with a directional capacity of 2,500 vehicles, while paths 1 and 3 are newer, wider facilities with directional capacities of 6,000 vehicles each. Travel times at practical capacity are 36, 34, and 37 integer time units respectively (most network analysis computer programs work with integer values). Volumes entering from 5 assumed source-centroids are as shown.

Table 1 gives the progression of calculated travel times and volumes loaded on each path for 4 iterations of the FHWA capacity-restraint procedure. The first iteration (free flow) loads the total demand of 5,500 vehicles onto path 2, the old narrow road. This leads to drastic overloading and consequently a substantial increase in travel time for path 2. Times for paths 1 and 3, on the other hand, are reduced because of the absence of flow. The second iteration loads all traffic onto path 1. This leads to a downward adjustment of travel times for paths 2 and 3. In iterations 3 and 4, traffic is loaded alternatively on path 3 and 1, indicating the beginning of a nonconvergent shifting process. The example clearly establishes the need to average the flows of all iterations to achieve reasonably stable flows. In fact, for the particular example, 4 iterations are not sufficient to attain stability.

Figure 1. Network configuration with 3 alternate paths between A and B.

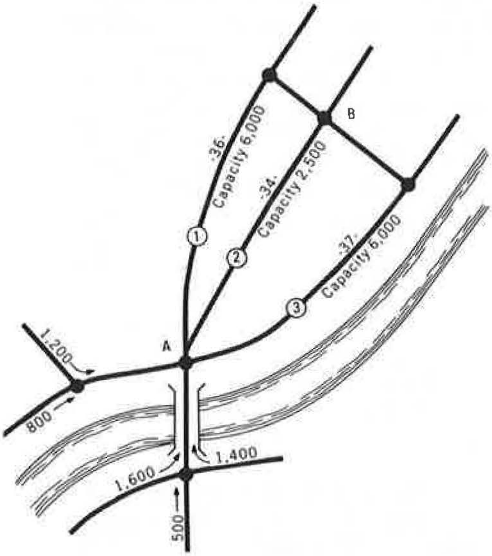


Table 1. Path times and flows for 4 iterations with averaging method.

	Path 1		Path 2		Path 3	
Iteration	Time	Flow	Time	Flow	Time	Flow
1	36	0	34	5,500	37	0
2	35	5,500	59	0	36	0
3	35	0	57	0	35[a]	5,500
4	34	5,500	55	0	35	0
Avg		2,750		1,375		1,375
Final V/C ratio	0.46		0.55		0.23	

[a] It is assumed that, in the case of equal times, path 3 rather than path 1 will be chosen. If path 1 were chosen first, path 3 would be selected in iteration 4. In either case, overall results would not be affected.

Table 2. Path times and flows for 5 centroids with 1-pass incremental method.

	Path 1			Path 2			Path 3		
Increment	Time	Flow	Cumulative Flow	Time	Flow	Cumulative Flow	Time	Flow	Cumulative Flow
1	31	0	0	30	1,200	1,200	32	0	0
2	31	800	800	38	0	1,200	32	0	0
3	32	0	800	38	0	1,200	32	1,600[a]	1,600
4	32	500	1,300	38	0	1,200	36	0	1,600
5	34	1,400	2,700	38	0	1,200	36	0	1,600
Final V/C ratio		0.45			0.48			0.27	

[a] If the stochastic process would have selected path 1 instead of path 3, final flows would be 2,400 for path 1 and 1,900 for path 3.

Table 3. Path times and flows for 100 centroids with 1-pass incremental method.

	Path 1			Path 2			Path 3		
Increment	Time	Flow	Cumulative Flow	Time	Flow	Cumulative Flow	Time	Flow	Cumulative Flow
1	31	385	385	30	605	605	32	110	110
2	32	440	825	34	165	770	32	445	605
3	34	495	1,320	36	55	825	33	550	1,155
4	36	330	1,650	36	330	1,155	35	440	1,595
5	36	495	2,145	38	220	1,375	37	385	1,980
Final V/C ratio		0.36			0.55			0.33	

Table 2 gives the progression of volumes and travel times with the 1-pass incremental method. For the example, with only 5 centroids, each increment's traffic demand is assumed to be that of 1 centroid. The initial times for all paths are originally set equal to 0.87 times the value given at a flow level equal to practical capacity. (This value corresponds to the ratio of the travel time under free-flow conditions versus the travel time at a flow equal to practical capacity as defined by the Highway Capacity Manual.) Centroids are assumed to be selected for loading from left to right. A flow of 1,200 vehicles is assigned first to path 2 and results in a substantial increase in time on that path. The demand for increment 2 is loaded onto path 1 and increases that path's travel time. The third increment will choose either path 3 or path 1 depending on the random samples generated. This results in a condition where the other path will receive the flows from the remaining 2 increments as illustrated.

This example demonstrates the economical operation of the proposed method. With only 5 trees and 4 network modifications, a speed-flow condition was achieved that appears more stable than that obtained with 20 trees (4 iterations with 5 trees each) and 3 network modifications. Because network modifications require only marginal amounts of computer time compared to building sets of trees, it is obvious that the 1-pass method achieves substantial savings. The added cost of building multiple-routing best path trees (approximately 15 percent more than minimum path trees) does not substantially affect these cost savings.

The smoothness with which the 1-pass method increases the volumes on each path from 0 to their final flow is barely indicated by this example. In a real network with several hundred origin centroids, the smoothness of increase in flows would be much more pronounced because the volume loaded along each tree would represent a smaller fraction of the final flow. If there were 100 centroids instead of 5 and an average volume of 55 trips per centroid, the loading process for the sample network would proceed as given in Table 3. With 20 trees being generated and loaded during each increment, the benefits of multiple routing are now clearly apparent. Modifications of travel times are more gradual and each path receives some flow in every increment.

Calculation of Average Link Speed—The average speed on any link in a congested network is determined from its flow and the total vehicle time traveled on the link. Total vehicle time can be expressed by the sum of the products of the number of vehicles assigned during each increment (between 2 network modifications) and the travel times at which those vehicles have been assigned. This is not easy to calculate because core storage limitations preclude keeping a separate accumulator for each link.

Most capacity-restraint assignment procedures solve this problem by calculating the sum of the products of the total assigned volumes by an average travel time calculated from the input time and the time used during the last increment or the last iteration. Analysis of this problem suggests that the area under the time-versus-volume curve would yield an acceptable approximation of vehicle time. Use of the area under the curve will result in a slight overestimation of the correct value caused by the assumption of a smooth curve instead of the resulting step function (Fig. 2); however, with proper adjustment, the calculated value presents a reliable estimate.

Travel-time calculations for paths 1, 2, and 3 of the 5-centroid example previously illustrated would result in average values of 34, 33, and 33 time units. For the 100-centroid example (Table 3), average values would correspond to 33, 34, and 34 time units respectively. The corresponding estimates for the iterative method, derived from the average volumes and the fourth-power relation, are 32, 30, and 32. If the estimates are made on the basis of speeds used for each iteration weighted by the volume loaded, the values would be 35, 34, and 35 time units. The resulting values for the 1-pass method clearly fall within the range of estimates from the iterative method, indicating a need for further research on this topic.

Fraction of Demand Assigned by Each Increment—The accuracy of assignment results is related to the amount of traffic assigned to each path and the fraction of the total traffic demand assigned between network modifications. This fraction in turn is controlled by the volume originating at each centroid and the number of centroids. The procedure, therefore, allows the user to choose the number of trees built between modifications.

Figure 2 shows the accumulation of volume for a typical link located centrally within a small network. The volume buildup is shown for assignments loading 5 and 20 trees respectively between network modifications. Results obtained in each case are quite similar, although a more reliable average speed estimate would have been expected from the loading based on 5-tree increments.

The program has been prepared with an additional routine to decrease the fraction of traffic demand loaded at each increment whenever the accumulation of flow exceeded a set level between successive network modifications. This procedure evaluates, after each modification, the average factor by which link times have been modified. Whenever the difference between this factor and the one obtained in the previous increment reaches the set level, the number of trees is decreased by a percentage calculated from a parabolic function of the difference.

Validity of Individual Trees

The 1-pass incremental procedure, in addition to its efficient operation, generates trees that follow realistic and logical routes and can be used for subsequent purposes such as analysis of selected links and accumulation of service parameters along their paths. As a result of the randomly selected loading sequence, most links start receiving flows during early increments. In a typical run with 19 increments, more than 45 percent of all links showed a flow with only 8.2 percent of the total demand loaded. With 20.2 percent of the volume processed, almost 67 percent of all links had received flows; and with 39.1 percent of the total demand loaded, more than 80 percent of all links showed flows.

The multiple-routing procedure, as illustrated in the 100-centroid example (Table 3), will load flows on competing routes at an even pace and thereby ensure the smoothness of the 1-pass capacity-restraint method and the validity of its individual trees.

Comparison of Highway Service Estimates—Service estimates accumulated along the paths generated by the capacity-restraint procedure were verified by comparing zone-to-zone travel times and travel distances with estimates obtained along the minimum-time paths of the average network. Values accumulated along the capacity-restraint paths were expressed as a percentage of those obtained from minimum-time paths and grouped according to 6 value ranges.

The data from a 165-centroid network with a high degree of congestion showed travel-time values that were 1 percent higher on the average for capacity-restraint paths. More than 75 percent of all capacity-restraint paths had travel-time values that were within 2 percent of minimum-path times. Only 12 paths contained differences greater than 5 percent with extreme values of 94 percent and 114 percent respectively.

Distance values accumulated along capacity-restraint paths were found to be 3 percent higher than those along the minimum-time path. However, this does not necessarily cause concern because assignments based on minimum-time paths have been found to underestimate vehicle-miles of travel. Table 4 gives the frequency of capacity-restraint paths by percentage difference of travel-time and travel-distance values.

Of more interest perhaps than the variation in average values is the occurrence of differences within the loading sequence of the capacity-restraint procedure. Figure 3 shows a plot of the average percentage difference of time and distance for each group of 10 successive paths. The plot of travel-time differences seems to indicate a slight but definite trend toward increasing values for the paths generated in the second half of the loading sequence. The same trend, but much more pronounced, is found for the distance values.

Detailed analysis shows that group 16, for which the accumulated distances over the capacity-restraint paths are 10 percent higher than those of the minimum-time path, contains 2 of the 3 most extreme paths. Another extreme path is located in group 15 and, like the other two, originates from a zone bounding the external cordon. Further investigation of the few extreme paths has not yet been undertaken, but a preliminary analysis indicates that peculiar network conditions may be the major cause of the differences. The findings suggest, however, a definite need for further detailed analysis, including plots of individual path traces.

Figure 2. Progression of V/C for loading of centrally located link for different numbers of trees built between network modifications.

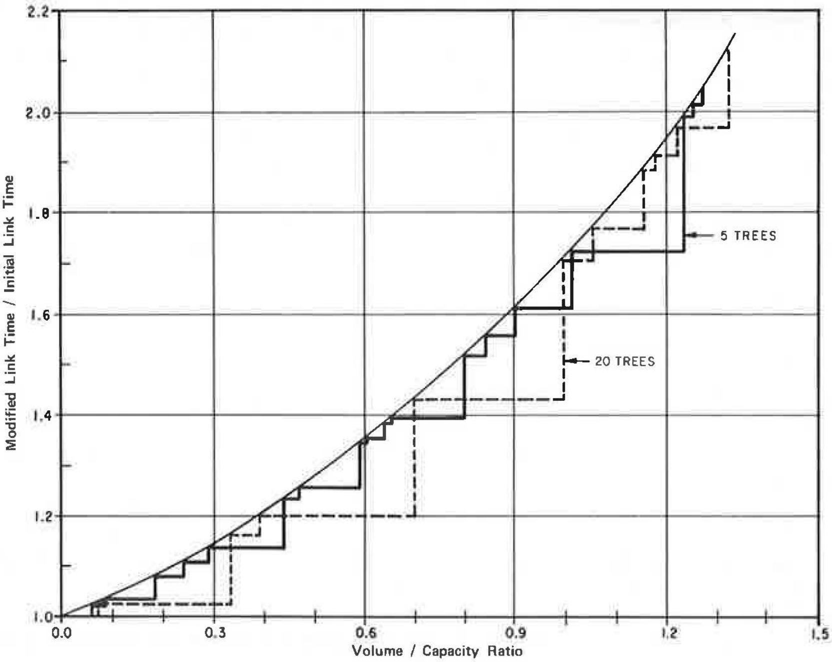
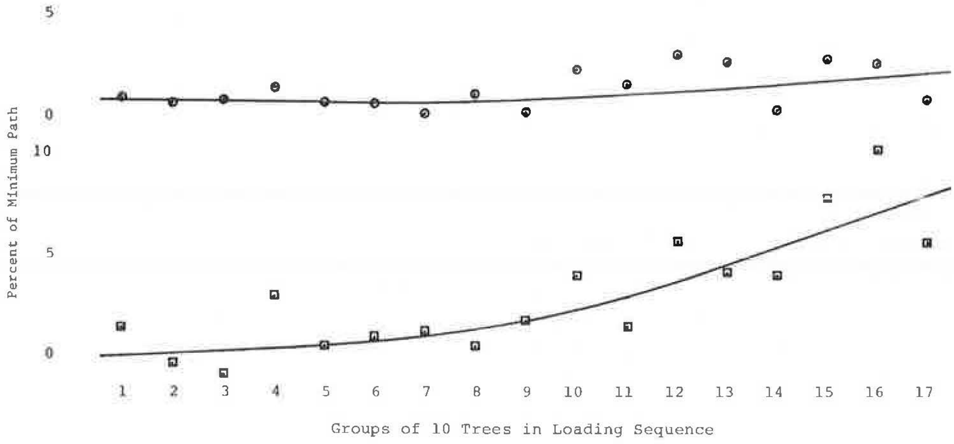


Table 4. Frequency of capacity-restraint paths by percentage difference in time and distance.

Percent of Minimum Path	Travel Time (number of paths)	Travel Distance (number of paths)
<95	5	2
95	0	1
96	0	2
97	2	3
98	0	1
99	6	16
100	42	17
101	52	29
102	27	25
103	9	16
104	8	10
105	5	9
106	1	9
107	1	10
108	2	3
109	1	3
>109	4	9
Total	165	165
Mean	101.4	103.1

Figure 3. Percentage difference in accumulated time and distance for capacity-restraint path by loading sequence groups.



CONCLUSIONS

The 1-pass incremental capacity-restraint program presented in this paper provides a flexible and economical tool for the estimation of highway service measures on congested networks. Computer costs experienced with this method indicate that stable estimates of speeds and flows can be obtained at costs that are, at most, 40 percent more than those required for a minimum-path all-or-nothing assignment. This compares with an increase of 350 to 400 percent when other capacity-restraint methods are used, not counting the cost to generate final sets of trees that properly reflect the level of congestion.

A number of areas have been identified that suggest the need for additional research including the calculation of average speeds for congested links and the detailed analysis of individual tree traces in relation to the loading sequence and the level of congestion.

The 1-pass procedure provides an efficient and realistic method to obtain service estimates for the evaluation of networks and modal choice under conditions of congestion. Because this procedure does not require cumbersome sequencing of many different programs, it should be of interest to those concerned with analyses and evaluation of alternative transportation networks.

REFERENCES

1. Transportation Network Evaluation: Costable Criteria. Stanford Research Institute, Nov. 1969.
2. Data Projections. Chicago Area Transportation Study, Vol. 2, July 1960.

ACTIVITY-ACCESSIBILITY MODELS OF TRIP GENERATION

T. Z. Nakkash and W. L. Grecco, Joint Highway Research Project, Purdue University

This paper examined the effect of activity-accessibility variables on trip generation. A second aspect of accessibility was studied by stratifying the zones of the study area by location. Data from the transportation study in Indianapolis were used to develop 4 sets of trip-generation regression equations for each of 13 trip purposes. A comparison of the developed models revealed that location always improved the statistical strength of the trip-generation models and relative accessibility variables improved the statistical strength of trip-attraction models more than that of trip-production models. Stratification alone improved the models more than relative accessibility variables only. 1985 forecasts of demographic, socioeconomic, and land use variables together with estimates of travel time on the proposed future network were inputs to the 4 sets of developed models to forecast trip productions and attractions. More trips for zones of the noncentral area and fewer trips for zones of the central area were forecast by stratified models than by unstratified models.

•IN recent years urban transportation planning activities have increased in many American cities. This increased activity was in response to the challenging urban transportation problem, one of the major problems of contemporary cities. The safe and efficient intracity movement of goods and people is very essential for the social, cultural, and economic health of an urban area.

The urban transportation problem is the product of many interacting factors. The enormous population growth in urban areas and their expanding areal extent as a result of the redistribution of population, the improved standard of living due to increased affluency, and the subsequent greater reliance on private automobiles are only some of those factors. Together with those size-related features of the problem, the temporal aspects induce periodic high demands for transportation. This, of course, is due to the interdependency of human activities that occur essentially during the 8-hour workday, starting and ending at rather definite times.

A recognition of the immense complexity and the vast dimensionality of the urban transportation problem is a prerequisite of any attempt to solve the problem.

Apart from the socioeconomic, demographic, and land use forecasts, trip generation constitutes the first step toward establishing the future travel pattern. The accuracy of the future trip distribution in forecasting design-year trip interchange cannot be any better than the accuracy of the trip-generation forecasts.

The ultimate purpose of the trip-generation analysis is to arrive at an estimate of the trip ends generated at each analysis unit of the study area. Trip-generation techniques try to establish a relation between the demographic and socioeconomic characteristics of the population of an analysis unit and its trip generation. Similarly, the intensity, character, and location of different land uses are related to trip-making of the analysis units. These procedures are based on the hypothesis of a causal relation between population characteristics, land use, and trip-making behavior of people.

Traditionally, trip-generation forecasts are established independently of any direct consideration of the transportation network. This, of course, assumes that trips produced at or attracted to a zone are a function only of the attributes of the zone itself and are not directly a function of the transportation network on which the trips were made. The traditional trip-generation process is shown in Figure 1.

PURPOSE AND SCOPE

The purpose of this research was to study the trip-generation process and specifically to investigate the effect of the transportation system on the rate of trip-making. Conceptually, there is no strong basis for assuming that trip-making is independent of the transportation system. On the contrary, it seems that trips produced by or attracted to a zone should be a function of the relative accessibility of the zone to different land uses in addition to the characteristics of the zone itself.

Trip-making is a product of the desire for human interaction and the necessity for having to perform different daily activities at different locales. Basically, the rate of trip-making is a function of 2 categories of variables: One tends to increase the potential of trip-making, and the other tends to restrict it. The availability of vehicles to the residents of a zone, the percentage of the residents in the labor force, the number employed in a zone, and the amount of floor area of different land uses are examples of the first category of variables. They measure the potential of trip production or trip attraction. The penalties incurred by travel measured in cost, travel time, or travel distance are variables that belong to the second category.

This study utilized data obtained from the surveys for the Indianapolis Regional Transportation and Development Study (IRTADS). Multiple linear regression predictive models of person-trip productions and attractions by purpose were developed. The developed models differ from the traditional trip-generation models. The independent variables were not restricted to socioeconomic and land use measures of the zones but included also measures of the relative accessibility of the zone to different activities and land uses.

The locational aspects that affect trip generation were also investigated. It was hypothesized that central locations in the study area, generally, afford greater accessibility; and the convergence of the street network on the city center favors the core location. The zones of the study area were stratified into 2 groups: central and noncentral. This stratification was entered as an independent dummy variable in the trip-generation analysis.

A comparison was made of the forecast trip generation by the suggested approach and the forecast by the traditional approach.

ACTIVITY-ACCESSIBILITY CONCEPT IN TRIP GENERATION

As stated earlier, the rationale of trip generation has been based on the demographic, economic, and land use characteristics of the zone; no consideration has been given to the status of the transportation system. This research was based on the concept that trips generated by a zone of the study area are also a function of the status of the transportation subsystem that serves a zone and connects it to the other zones of the study area. The effect of the transportation system on trip generation was investigated in the light of the relative accessibility of each zone to various urban activities and of the spatial relation of the different zones to each other.

There are many instances where researchers have realized the existence of a feedback from the transportation system to the trip-generation phenomenon (7, p. 18; 6, pp. 201-202; 11, p. 75; 1, p. 66; 15, p. 98; 9, p. 166; 2, p. 38; 3, pp. 73-74; 10; and 8). In spite of the large number of references noted above, actually little has been done to measure those effects.

The hypothesis proposed by this research is that the number of trips generated by a zone is a function of the transportation subsystem that connects the zone under consideration with the other zones of the study area. For this purpose, the term "relative accessibility" will be defined conceptually and operationally.

It was hypothesized earlier that the trips produced by or attracted to a zone are a function of the causal or symptomatic variables modified by the relative ease in overcoming space between that zone and all other zones. Zones with relatively more accessible destinations should, in general, produce more trips; similarly, zones that are relatively more accessible to origins should, in general, attract more trips. The term "relatively" refers to the zone under consideration as compared to all other zones of the study area. This implies a competitive consideration among zones in generating

Figure 1. Traditional trip-generation process.

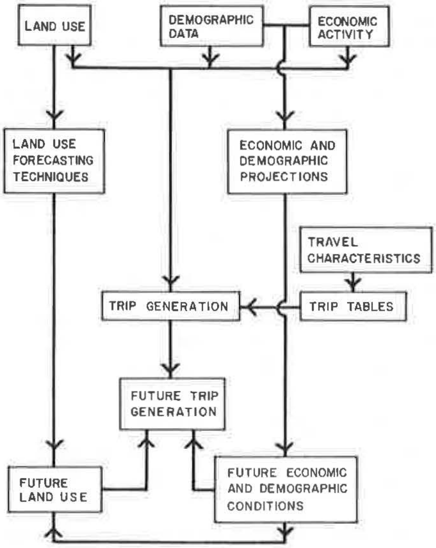


Table 1. Dependent variables for which trip-generation equations were developed.

Person-Trip Production		Person-Trip Attraction	
Purpose	Number	Purpose	Number
Home-based work	1	Home-based work	8
Home-based shop	2	Home-based shop	9
Home-based school	3	Home-based other	10
Home-based other	4	Non-home-based work	11
Non-home-based work	5	Non-home-based nonwork	12
Non-home-based nonwork	6	Total	13
Total	7		

Figure 2. Generated relative accessibility variables.

MEASURES OF ACTIVITY	TRIP PURPOSE					
	HBWK 1	HBSHP 2	HBSCL 3	HBOTR 4	NHBWK 5	NHBNWK 6
Total Employment	P			P - A	P - A*	
Retail Employment						
Service Employment						
Retail Floor Area		P		P - A	P - A	P - A
Educational Floor Area			P	P - A	P - A	P - A
Dwelling Units		A		P - A		
Labor Force	A			P - A		
Population		A		P - A		
Cars				P - A		
Single Family Dwellings		A		P - A		

* P (and/or A) in cells indicates that the corresponding relative accessibility variable was considered in developing models of trip productions (and/or attractions) for the indicated trip purpose

trips. Zones of similar-sized activities will attract trips differently according to their locational and accessibility advantages.

As a measure of the ease or difficulty of overcoming space, this study used the set of friction factors developed from the calibrated gravity model of trip interchange for the study area. This had the advantage of avoiding the use of a constant exponent of distance or time. Friction factors as developed by the calibration of the gravity model are a function of travel time and are classified by trip purpose.

This study's definition of relative accessibility is a modification of Hansen's. In notational form, relative accessibility is computed as follows:

$$A_{i\,.\,k(\ell)} = \sum_{j=1}^{n} S_{jk}\, F_{ij(\ell)}$$

$$RA_{i\,.\,k(\ell)} = A_{i\,.\,k(\ell)} \Big/ \left[\sum_{i=1}^{n} \sum_{j=1}^{n} S_{jk}\, F_{ij(\ell)}\right] \cdot 100 = A_{i\,.\,k(\ell)} \Big/ \left[\sum_{i=1}^{n} A_{i\,.\,k(\ell)}\right] \cdot 100$$

where

i = zone under consideration ($i = 1, 2, \ldots, n$);
j = any zone in the study area, including zone i ($j = 1, 2, \ldots, i, \ldots, n$);
k = activity under consideration ($k = 1, 2, \ldots, m$);
ℓ = trip purpose ($\ell = 1, 2, \ldots, p$);
S_{jk} = size of activity k in zone j;
$F_{ij(\ell)}$ = friction factor corresponding to the travel time from zone i to zone j for purpose ℓ;
$A_{i\,.\,k(\ell)}$ = accessibility of zone i to activity k for purpose ℓ; and
$RA_{i\,.\,k(\ell)}$ = relative accessibility of zone i to activity k for purpose ℓ.

The value of the relative accessibility of zone i to activity k for purpose ℓ, as it is possible to infer from its formulation, could be different for a future year if any or all of the following study area parameters change:

1. Interzonal travel time, consisting of interzonal driving time, terminal time, and intrazonal time; or
2. Size of activity k in any or all zones of the study area.

The value of the relative accessibility of a zone to the same activity could be different for different trip purposes. The reason is that the friction factors, corresponding to a certain travel time, are usually different for different trip purposes.

Another aspect of accessibility, directly related neither to travel time nor to size of activity, is considered. The relation to trip-generation characteristics is investigated by stratifying the zones of the study area into 2 sets: central and noncentral. The conceptual basis is that "central sites afford maximum accessibility..." (2, p. 108). The central area is also, more or less, equally accessible to the various zones of the study area because of the convergence of the street system on the city center. This stratification introduces a qualitative factor describing the general arrangement of the land uses and the configuration of the street system.

STUDY DESIGN AND ANALYTICAL PROCEDURES

The following methodology was used in this investigation. Trip-generation models that take into account accessibility variables were developed from data of an operational transportation study. Those models were then compared with the conventional models developed as part of the transportation study. Both sets of models were used to forecast 1985 trip generation. The 2 sets of forecasts were compared by testing for any significant differences, on a zone-by-zone basis, between the 2 forecasts.

Models that take into consideration the stratification of the zones of the study area into central and noncentral sets were also developed. This stratification was investigated for the models developed by the transportation study and those developed by this investigation.

The main purpose of this investigation was to compare the sets of developed models that used accessibility variables and stratified zones into central and noncentral sets with the traditional models developed by an operational transportation study. Care was taken to keep any factors that might disturb the comparison out of the developed models so that the comparisons would be most valid. Furthermore, the decision to develop multiple linear regression models of trip generation by using data summarized by zone was mainly in the interest of keeping the results of this investigation comparable with those of the operational transportation study.

Data Preparation

The data used in this investigation were obtained in the IRTADS surveys. The study was at a stage where most, if not all, of the analyses were completed, the forecasts established, and the proposed networks evaluated. In the trip-generation analysis, all the models were developed for total person trips (except for truck and taxi equations). The dependent variables were in the form of productions and attractions suitable for distribution by the gravity model. Nineteen trip-generation equations were developed by IRTADS; 2 were for truck and taxi trip ends, 4 were for control totals, and the others were for person-trips productions and attractions by purpose.

Dependent Variables—This investigation was limited to 6 trip-production purposes, 5 trip-attraction purposes, and 2 control totals (1 for all productions and 1 for all attractions). Trip-generation equations were developed for the dependent variables given in Table 1. It was not possible to develop an equation for home-based school person-trip attractions because the key independent variable, school enrollment, was not available.

Socioeconomic and Land Use Variables—Twenty-nine socioeconomic and land use variables were originally considered by IRTADS in its trip-generation analysis; only 15 were, however, eventually retained in the final equations. Moreover, only 10 of these were available for this study for both the survey year and the forecast year. The following socioeconomic and land use variables were available and used by this study as independent variables in developing multiple linear regression models of trip generation: total employment, retail employment, service employment, retail floor area, educational floor area, dwelling units, labor force, population, cars, and single-family dwelling units.

Accessibility Variables—Different measures of relative accessibility to be used as independent variables in trip-generation regression models were established. The operational definition of relative accessibility stated earlier was used. The required inputs are the size of various activities in each zone and skim zone-to-zone friction factor trees.

Size of Activities—The definition of activity was extended for this purpose to include all the 10 socioeconomic independent variables that were available from the IRTADS surveys. If the information had not been available from that source, the size of activities would have been collected from various sources, e.g., employment data from the State Employment Securities Division, floor area information from the land use survey, and other data from the home interview survey or census records.

Skim Zone-to-Zone Friction Factor Trees—A set of friction factors for each of 6 trip purposes was available from the results of the IRTADS calibrated gravity model.

A binary zone-to-zone tree tape was made available by IRTADS for this investigation. This binary tape was updated (intrazonal and terminal times added) and skimmed to give skim zone-to-zone travel time binary trees. Six skim zone-to-zone friction factor trees were built, 1 for each of 6 trip purposes.

Only the highway network was considered in developing relative accessibility variables. Although the trip-generation models were for person trips, using the highway network only would not introduce any appreciable bias in the case of IRTADS mainly because transit passenger trips constituted only 4.1 percent of all the person trips (3). Moreover, the transit in IRTADS area was entirely bus service on the city streets.

Generated Relative Accessibility Variables—With 10 activity measures (the available independent variables) and 6 sets of friction factors (1 for each trip purpose), 60 measures of relative accessibility could be generated. However, not all 60 possible combinations were generated; only those that were meaningful to the trip generation analysis were used. For example, the relative accessibility of a zone to retail floor area could be meaningful in conjunction with home-based shopping person-trip productions. Also, the relative accessibility of a zone to single-family dwellings would be meaningful in conjunction with home-based shop person-trip attractions. Twenty relative accessibility variables were generated and considered in the analysis; these are shown in Figure 2. The same measure of relative accessibility could be meaningful in conjunction with both the productions and attractions of some of the trip purposes.

Delimiting the Central Area

As indicated earlier, this investigation considered stratifying the study area into central and noncentral areas. It was assumed that the central and noncentral areas might reflect 2 different trip-generation patterns because of the shape of the study area, its historical quasiannular urban growth, and the configuration of the transportation system. This differentiation of the central and noncentral areas was categorical instead of numerical and could best be treated through stratification.

The rationale behind the procedure developed to delimit the central area was tied to the expected character and attributes of a central area. A large proportion of the land in a central area is expected to be in urban use. A relatively small proportion of the land in a central area is expected to be devoted to residential uses. Conversely, a relatively high proportion of the land in the central area was expected to be in uses that are known to seek central location.

Land was considered to be in urban use if it did not belong to any of the following classifications: quarry and mining, automobile junkyard, water, agriculture, or vacant. The percentage of urban land in residential use was the measure of the intensity of residential activity. The percentage of land in urban use was the measure of urbanity. Among the different trade and service uses on which IRTADS had floor area information, the following were chosen as those that seek a central location: wholesale trade (without warehousing); general retail trade; automobile retail; apparel, furniture, and appliance retail; retail use not otherwise classified; finance, business, and professional services; contract construction services; governmental services; personal services; and services not otherwise classified. Educational services were excluded because schools do not necessarily seek central locations. The floor area of each use in hundreds of square feet per acre of land in urban use was calculated for each district. This ratio measures not only the amount but also the intensity of use.

The measures given above were used to set the following conditions for delimiting the central areas:

1. The delimitation should be performed at the district level (a district is a contiguous group of zones).
2. The central area should probably include all of the central business districts and some of the qualifying surrounding districts.
3. The districts of the central area should all be contiguous and connected.
4. A qualifying district must be at least two of these: in the lower quartile of all the districts of the study area in percentage of urban land in residential use, in the upper quartile of all the districts of the study area in the percentage of land in urban use, or in the upper quartile of all districts of the study area in the ratio of hundreds of square feet of uses usually seeking central location to acres of urban land in each district.

The study area was stratified into central and noncentral areas (Fig. 4). The districts of the central area comprised 105 zones out of the 395 in the study area.

MODEL BUILDING

Guidelines for Model Building

Multiple linear regression models of trip generation were developed by using the computer program BMD-2R, stepwise regression. In addition to the desired statistical qualities of the developed models, other important factors were also considered.

Conceptual Validity—The consideration of relative accessibility was mainly to achieve a sounder conceptual basis for trip generation. In addition, only independent variables that were logically related to the specific dependent variable under consideration were allowed to enter when regression equations were developed for that dependent variable. The causal-logical relation was considered prior to the mere statistical correlation analysis. Association and correlation do not prove causality; causality should only be hypothesized on theoretical or conceptual grounds.

Model stability is one of the desirable products of conceptual validity. Relations that are not conceptually valid, if established from today's data, are more apt not to hold in the future. Predictive equations of trip generation should hold for the future to have any forecasting capability.

Another facet of conceptual validity is the sign of the regression coefficient. Because of collinearity in the variables, the coefficient of one of the independent variables could be contrary to the theoretical relation, and this condition might be statistically acceptable. In spite of this, it was decided to delete those variables whose coefficients had signs contrary to conceptual expectations. This should increase the statistical validity of the model as it tends to reduce the effects of collinearity.

Simplicity—The models were kept as simple as possible by avoiding unnecessary transformations of and interactions among the original independent variables. Interactions beyond the product of 2 independent variables were considered difficult to interpret and thus were avoided unless the third variable of the product was the dummy variable defining the location of a zone in the central or noncentral areas.

Keeping the structure of the model as simple as possible by not going to higher order interactions curtails the propagation of measurement errors. Another aspect of simplicity is parameter parsimony. Although it is valuable to include all relevant independent variables and thus reduce specification errors, it is doubtful that it would be advantageous to do so when, as is the case for transportation studies, the input data are inherently plagued by measurement errors. As an emphasis of this research, the number of independent variables in the model was kept to a minimum.

Stability—So that the developed models would be stable during a time period, the prerequisite of allowing a variable to enter the model was a hypothesized causal relation rather than a mere correlation.

Stability was also sought over the range of the values of the independent variables. This could be quite a difficult criterion to account for during model building. A study of the range of the independent variables for the forecast year was undertaken, and possible problem zones were identified. Recommendations will be made to ameliorate this condition.

Sensitivity—It is desirable that the response of the dependent variable be sensitive to changes in each of the independent variables in the model. The cost of adding 1 more independent variable would not be justifiable if the dependent variable is not sensitive to changes in the added independent variable. The sensitivity of the dependent variable to each of the independent variables in the model was tested by calculating the standardized regression coefficients (the regression coefficients multiplied by the ratio of the standard deviation of the independent variable under consideration to the standard deviation of the dependent variable).

Statistical Considerations

Stepwise Regression—The computer program used by this research, as mentioned earlier, was BMD-2R, stepwise regression. Several procedures are available to develop multiple regression models. The "tear-down" or "backward elimination" method starts with a model containing all the available independent variables and subsequently eliminates some of the independent variables until a model with predescribed statistical features is reached. The "build-up" or "forward selection" procedure strives for a similar final outcome but works in the opposite direction by inserting 1 more independent variable at a time. Stepwise regression is an improved version of forward selection procedure. The independent variables in the model are reexamined at the end of each step. The variable that might have been the best single variable to enter at an earlier step might prove to be unnecessary at a later stage because of the relation between it and other variables now in the equation. Thus, at each step, the partial F-test for each variable in the equation was evaluated and compared to a preselected percentage point of the appropriate F-distribution. Stepwise regression evaluates the contribution of each independent variable in the model at the end of each step, regardless of whether the independent variable has entered at the last step or at any earlier step.

Partial F- or Sequential F-Test—By far, the most important statistic in conjunction with regression analysis is the multiple coefficient of determination, R^2. It measures the proportion of total variability in the dependent variable explained by the regression model. R^2 varies between 0 and 1: A value of 0 indicates a complete lack of fit, while a value of 1 implies a perfect correlation. In stepwise regression, a test is needed at each step to check whether the increase in R^2 contributed by each added independent variable in the equation is significantly different from 0. The following F-statistic tests whether the contribution of the k independent variables is significantly greater than 0.

$$F_{k,\, n-k-1} = (R_k^2/k)/[(1 - R_k^2)/(n - k - 1)]$$

where

n = number of observations,
k = number of independent variables, and
R_k^2 = coefficient of multiple determination of a model with k independent variables.

The calculated F-statistic is compared to a tabulated $F_{k,\, n-k-1,\, 1-\alpha}$, where α is the probability of a type I error, or the level of significance. The level of significance chosen should depend on the consequences of rejecting a true hypothesis. The level of significance was set at 0.010 for including a variable and at 0.005 for deleting a variable. The selection of these values is based on acceptance of a relatively high risk of including a variable that does not belong. Once this variable has been accepted, there is a lower risk acceptable for its retention in the equation based on the entry of other independent variables.

The blind use of the F-test may result in the development of a regression model that involves more independent variables than are of practical significance. In transportation studies, the number of observations is large and results in an F-statistic that is statistically significant even when the absolute increase in R^2 is very small. The criterion of a significant increase in R^2 proved to be superfluous in most cases; other criteria such as simplicity, parsimony, and reasonableness controlled the number of variables to be included in the model.

Standard Error of Estimate—Another statistic of interest is the standard error of the estimate, s. It is the square root of the residual mean square. The smaller the value of this statistic is, the more precise the predictions will be. The criterion of reducing s must be used cautiously because s can be made small by including enough parameters in the model, just as R^2 can be increased. As more independent variables are included in the equation, the decrease in s will be at a decreasing rate. Reduction of s is desirable if many degrees of freedom for error are remaining.

Another way of looking at the reduction in s is to consider it in relation to the dependent variable, namely, as a percentage of the mean value of the dependent variable. Standard error of estimate as a percentage of the mean of the dependent variable is referred to as the coefficient of variation.

t-Test on Regression Coefficients—It is sometimes desirable to test whether each of the estimated regression parameters is significantly different from 0. The ratio of each regression coefficient to its standard error is distributed as student-t. If the regression coefficient of one of the independent variables does not pass the t-test, it can be deleted from the equation.

The 3 criteria of R^2, s, and significance of the regression coefficient are not independent. Usually, the decision can be made on the basis of R^2 alone.

Model Identification

The first set of models that was developed by this investigation was a rerun for each of the 13 dependent variables using the same independent variables established by IRTADS for its equations (3). In the interest of compatibility and comparability, data from all the 395 zones were used to reestimate the parameters of the models developed by IRTADS. Those models, essentially developed by IRTADS, were used as a basis for comparison with other developed models.

A second set of models that include relative accessibility variables was attempted for each of the 13 dependent variables. The first 2 sets of models were developed with data from the 395 zones with no distinction relative to location in the central or noncentral areas. Two more sets of models were developed: one corresponding to the set developed by IRTADS and the other to the set of models developed by this investigation. These latter models contained a dummy variable defining the location of a zone in the central or noncentral areas.

Basically, 4 sets of models were developed. Two had no relative accessibility variables among their independent variables: one of those, set W-U, was developed by the traditional procedures for IRTADS and the other, set W-S, contained a dummy variable that defined the zone location or some of the interaction of the dummy variable with the other independent variables in the equation or both of these. Of the remaining 2 sets, set A-U had relative accessibility variables, and set A-S also had relative accessibility variables and was calibrated with stratified data. Figure 3 shows a system for the identification of the developed models. It was not possible to develop models for each dependent variable in every set, as indicated.

COMPARISON OF THE DEVELOPED MODELS

The statistical strength of the models was compared by comparing their corresponding coefficients of multiple determination. The model sets compared to determine improvements by the actions indicated were as follows:

1. A-U versus W-U, introducing relative accessibility variables to the basic IRTADS models;
2. W-S versus W-U, calibrating the models with data stratified according to the zone location over the basic IRTADS models; and
3. A-S versus W-U, introducing both relative accessibility variables and calibrating the model with stratified data over the basic IRTADS models.

The results of the statistical tests of the significance of the increase in R^2 of each developed model over the corresponding basic IRTADS model are given in Table 2. Comparative summary statistics of all developed models are given in Table 3.

CONCLUSIONS

Based on the preceding results and analyses, the following conclusions can be drawn:

1. Among all of the relative accessibility variables considered, the following variables were included in the trip-generation models that were developed: accessibility to employment in conjunction with home-based work productions and non-home-based work attractions; accessibility to labor force in conjunction with home-based work attractions; accessibility to single-family dwellings in conjunction with home-based shop attractions and home-based other attractions; and accessibility to educational floor area in conjunction with home-based school productions and non-home-based nonwork attractions. The preceding accessibilities were each calculated with the friction factor corresponding to the same trip purpose as the model under consideration.
2. Relative accessibility variables in trip-generation models improved the statistical strength of models of person-trip attractions more than that of models of person-trip productions. Competition is a more important locational consideration for high-attraction zones, which indicates their need for greater accessibility.
3. Calibrating trip-generation models with data stratified according to the location of the zone in the central or noncentral areas always improved the statistical strength

Figure 3. Trip-generation models developed.

THE DEPENDENT VARIABLES	WITHOUT ACCESSIBILITY W		WITH ACCESSIBILITY A	
	UNSTRATIFIED U	STRATIFIED S	UNSTRATIFIED U	STRATIFIED S
1. HOME_BASED WORK PERSON_TRIP PRODUCTIONS (HBWKP)	W_U_1	W_S_1	A_U_1	☒
2. ,, ,, SHOP ,, ,, ,, (HBSHPP)	W_U_2	☒	☒	☒
3. ,, ,, SCHOOL ,, ,, ,, (HBSCLP)	W_U_3	W_S_3	A_U_3	A_S_3
4. ,, ,, OTHER ,, ,, ,, (HBOTRP)	W_U_4	☒	☒	☒
5. NON HOME_BASED WORK_ORIENTED PERSON_TRIP PRODUCTIONS (NHBWKP)	W_U_5	W_S_5	A_U_5	A_S_5
6. NON HOME_BASED NON WORK_ORIENTED PERSON_TRIP ,, (NHBNWP)	W_U_6	W_S_6	A_U_6	A_S_6
7. TOTAL PERSON_TRIP PRODUCTIONS (TOTP)	W_U_7	W_S_7	A_U_7	A_S_7
8. HOME_BASED WORK PERSON_TRIP ATTRACTIONS (HBWKA)	W_U_8	W_S_8	A_U_8	A_S_8
9. ,, ,, SHOP ,, ,, ,, (HBSHPA)	W_U_9	W_S_9	A_U_9	A_S_9
10. ,, ,, OTHER ,, ,, ,, (HBOTRA)	W_U_10	W_S_10	A_U_10	A_S_10
11. NON HOME_BASED WORK_ORIENTED PERSON_TRIP ATTRACTIONS (NHBWKA)	W_U_11	W_S_11	A_U_11	A_S_11
12. NON HOME_BASED NON WORK_ORIENTED PERSON_TRIP ,, (NHBNWA)	W_U_12	W_S_12	A_U_12	A_S_12
13. TOTAL PERSON_TRIP ATTRACTIONS (TOTA)	W_U_13	W_S_13	A_U_13	A_S_13

☒ INDICATES THAT NO SATISFACTORY MODEL WAS DEVELOPED.

Table 2. Results of comparisons of sets of models.

Trip Purpose	A-U Versus W-U	W-S Versus W-U	A-S Versus W-U	Trip Purpose	A-U Versus W-U	W-S Versus W-U	A-S Versus W-U
1	N	S	*	8	N	N	**
2	*	*	*	9	**	S	S
3	S	S	S	10	S	S	S
4	*	*	*	11	**	S	**
5	**	S	**	12	**	S	**
6	**	S	**	13	**	S	S
7	**	S	**				

Note: N = increase in R^2 is not significant at $\alpha = 0.0005$, S = increase in R^2 is significant at $\alpha = 0.0005$, * = no satisfactory models were developed, and ** = models were developed but no statistical testing was possible.

Table 3. Comparative statistics for all sets of models.

Trip Purpose	Coefficient of Multiple Determination				Coefficient of Variation (percent)				Number of Independent Variables in Model			
	W-U	W-S	A-U	A-S	W-U	W-S	A-U	A-S	W-U	W-S	A-U	A-S
1	0.974	0.975	0.977	*	16.30	16.00	15.52	*	1	2	3	*
2	0.887	*	*	*	38.38	*	*	*	2	*	*	*
3	0.630	0.659	0.670	0.699	80.28	77.14	75.95	72.66	1	2	2	3
4	0.903	*	*	*	31.99	*	*	*	2	*	*	*
5	0.748	0.767	0.748	0.790	53.83	51.89	53.89	49.41	4	7	5	9
6	0.650	0.806	0.643	0.797	70.43	52.69	71.10	53.91	5	8	5	9
7	0.961	0.965	0.958	0.962	17.89	17.15	18.61	17.83	4	6	4	6
8	0.839	0.840	0.840	0.842	54.93	54.74	54.72	54.54	1	2	2	3
9	0.442	0.705	0.553	0.639	158.64	115.63	141.99	127.92	2	4	2	4
10	0.679	0.716	0.682	0.719	53.05	50.04	52.90	49.94	5	7	6	9
11	0.738	0.757	0.739	0.780	54.01	52.22	54.15	49.84	4	8	7	10
12	0.636	0.799	0.645	0.801	73.67	55.05	72.90	54.70	5	8	6	7
13	0.771	0.844	0.801	0.848	46.68	38.71	43.58	38.27	5	8	5	9

* = no satisfactory model developed. Degrees of freedom ranged from 382 to 392.

Figure 4. Noncentral area zones where basic IRTADS models underforecast trips.

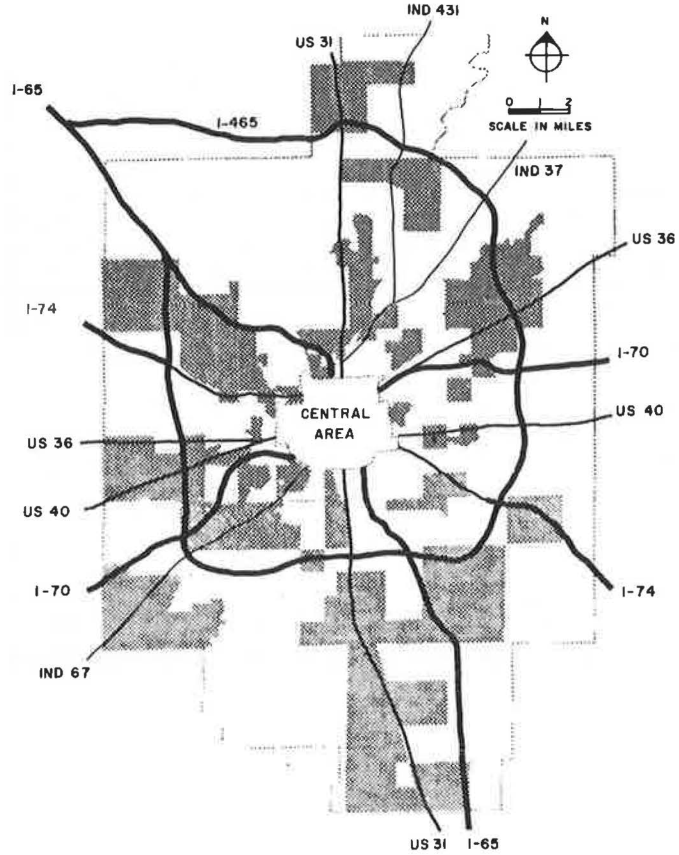
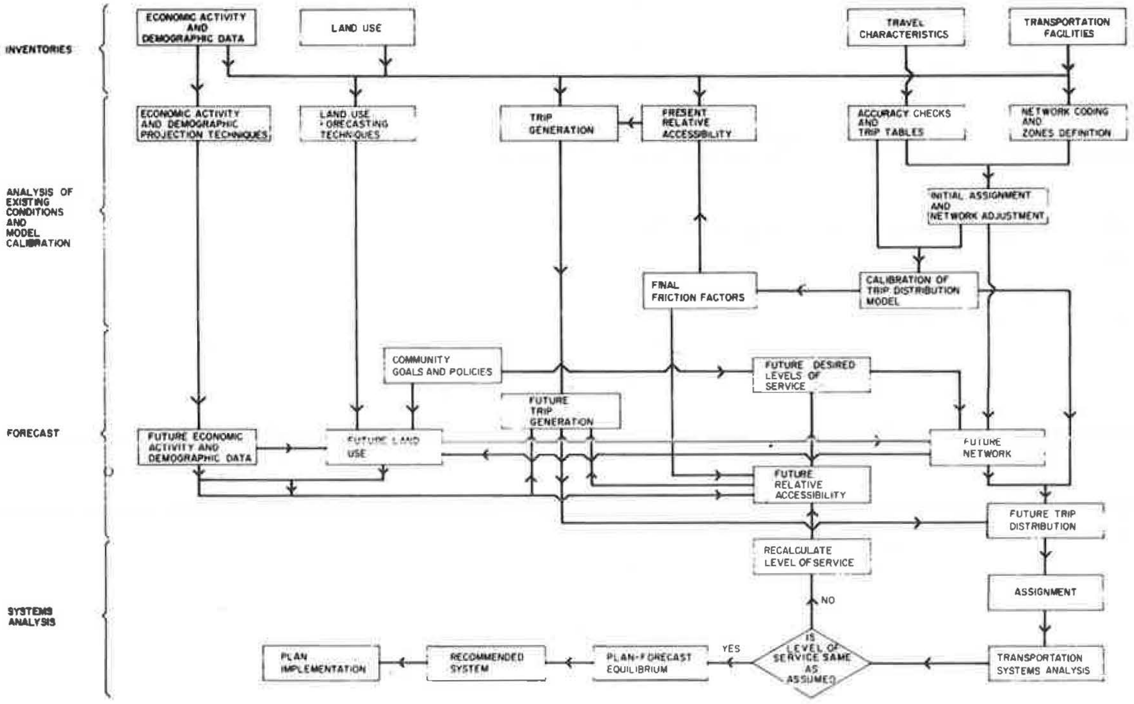


Figure 5. Proposed trip-generation process.



of the models whether the models had accessibility variables or not. Models of home-based work person-trip productions or attractions were improved least by including relative accessibility variables or stratification or both. This is expected because work trips are inelastic to trip length because of their regularity and necessity. Furthermore, it indicates substantially similar attracting characteristics for work trips by zones in the central and noncentral areas.

4. In general, the statistical strength of the developed models was better achieved by stratification alone than by including relative accessibility variables only.

The 4 sets of developed models were solved with the 1985 forecast values of the independent variables. The forecasts were analyzed to identify comparative forecasting trends of the different models. The following conclusions were drawn.

5. It was observed that stratified models consistently forecast more trip productions and attractions for zones of the noncentral area and fewer for zones of the central area than models without stratification. Stratified models are thus sensitive to the situation of equilibrium and saturation being reached in the central area and also the faster rate of traffic growth in the noncentral area.

6. 1985 forecasts of person-trip productions and attractions by models that had relative accessibility variables and that were calibrated with stratified data were significantly different from forecasts by basic IRTADS models. There was not a detectable trend as to the sign of the mean difference between zones of the central and noncentral areas. Further analysis indicated that stratified models with relative accessibility variables forecast more productions and attractions than were forecast by basic IRTADS models, in general, for zones located in the vicinity and along corridors defined by the major thoroughfares of the study area. This reflects a possible locational aspect of trip generation in addition to the central-noncentral stratification. This is shown in Figure 4.

The trip-generation models proposed by this research are functions of the status of the transportation system. In an operational transportation study, future forecasts of trip generation would then be affected by the nature of the proposed transportation network. Because the proposed network should be designed to serve future trip generation, an iterative process should be followed. It would be terminated when an equilibrium between the future supply of transportation (proposed plan) and the demand for transportation (travel forecast) is reached. This iterative process is shown in Figure 5.

ACKNOWLEDGMENTS

The authors are grateful to the Joint Highway Research Project of Purdue University and the Indiana State Highway Commission for the financial support that, in addition to the assistance of the Federal Highway Administration, made this research possible. The opinions, findings, and conclusions expressed in this paper are those of the authors and not necessarily those of any of the supporting agencies.

REFERENCES

1. Domencich, T. A., Kraft, G., and Valette, J. P. Estimation of Urban Passenger Travel Behavior: An Economic Demand Model. Highway Research Record 238, 1968, pp. 64-78.
2. Haig, R. M. Regional Survey of New York and Its Environs. Committee on Regional Plan of New York and Its Environs, Vol. 1, 1927.
3. Hansen, W. G. How Accessibility Shapes Land Use. Jour. of the American Institute of Planners, Vol. 25, No. 2, May 1959, pp. 73-77.
4. Transit Inventory. Indianapolis Regional Transportation and Development Study, Tech. Work Paper, 1966.
5. Trip Generation Analysis. Indianapolis Regional Transportation and Development Study, Tech. Work Paper, 1966.
6. Martin, B. V., Memmott, F. W., and Bone, A. J. Principles and Techniques of Predicting Future Demand for Urban Area Transportation. M.I.T. Press, Cambridge, 1961.

7. Mitchell, R. B., and Rapkin, C. Urban Traffic a Function of Land Use. Columbia Univ. Press, New York, 1954.
8. Savigear, F. A Quantitative Measure of Accessibility. Town Planning Review, April 1967, pp. 64-72.
9. Schneider, M. Access and Land Development. HRB Spec. Rept. 97, 1968, pp. 164-177.
10. Shindler, R., and Ferreri, M. G. Auto Ownership as Affected by Transportation System Alternatives. Traffic Engineering, Vol. 38, No. 1, Oct. 1967, pp. 24-28.
11. Shuldiner, P. W. Land Use, Activity, and Non-Residential Trip Generation. Highway Research Record 141, 1966, pp. 73-88.
12. Wilbur Smith and Associates. Baltimore Metropolitan Area Transportation Study Report, Vol. 1, 1964.
13. Stewart, J. Q. Demographic Gravitation: Evidence and Applications. Sociometry, Vol. 11, No. 1, Feb. 1948, pp. 31-58.
14. Swerdloff, C. N., and Stowers, J. R. A Test of Some First Generation Residential Land Use Models. Highway Research Record 126, 1966, pp. 38-59.
15. Walker, J. R. Rank Classification: A Procedure for Determining Future Trip Ends. Highway Research Record 240, 1968, pp. 88-99.

COMPARISON OF PROBABILISTIC MODAL-CHOICE MODELS: ESTIMATION METHODS AND SYSTEM INPUTS

Antti Talvitie, School of Civil Engineering and Environmental Science,
University of Oklahoma

Twelve models were formulated by segmenting the total travel time and total travel cost by rapid transit and by automobile in different ways or by leaving them out completely and including only socioeconomic variables in the model. These models were then estimated by using logit, probit, and discriminant analyses. The results were evaluated in 2 respects: Are there differences in performance among the methods of estimation? and Are there differences in performance among the 12 model specifications? The results indicate that there are no statistically significant differences either among the methods of estimation or among the model specifications themselves. A model that uses only 2 user characteristics, income and the number of working household members, and 1 system-related variable, a dummy variable for walk access to the transit station, performs no worse than a model that uses a whole set of system characteristics in addition to those 3 variables. Values of time significantly lower than those previously reported were found; the "best" estimate in this study is only 12 percent of the wage rate.

THERE ARE 3 objectives in this research: (a) to investigate the relative merits of the 3 methods most widely used in probabilistic modal-choice modeling or in value of time studies, logit, probit, and discriminant analyses (1, 2, 3, 4); (b) to investigate the need for trip segmentation, that is, Is there a need to differentiate among access, egress, and line-haul times and costs? and if so, How should this trip segmentation be done?; and (c) to obtain further evidence on the value of time.

Answers to the first objective are provided by estimating several different models by all 3 methods and then by assessing and comparing the forecasting accuracy of each method. For the second objective, these different models were designed by segmenting the travel costs and travel times in different ways. The forecasting accuracy of each model specification was then assessed and compared with the forecasting accuracy of the other model specifications. The third objective was accomplished as a by-product from several different models estimated by objectives 1 and 2.

BASIC MODEL AND METHODS OF ESTIMATION

A general probabilistic travel demand model can be expressed as

$$\Pr(ij, M) = \Pr(ij) \cdot \Pr(M|ij) \qquad (1)$$

where

$\Pr(ij, M)$ = probability that the event (ij, M) occurs, that is, an individual makes a trip between points i and j by using mode M;
$\Pr(ij)$ = probability that an individual makes a trip between points i and j; and
$\Pr(M|ij)$ = probability that an individual uses mode M, given that he makes a trip between points i and j.

Clearly, in modal choice, models are estimated for $\Pr(M|ij)$. In this research only binary choice, automobile versus rapid transit, is considered.

It is assumed that there is an index y that determines to which group (automobile or transit) an individual is likely to belong. This index y is constructed as follows:

$$y = a + b\,(L_{ij}^{M1} - L_{ij}^{M2}) + c\,(SE)$$

where

a = constant term;
b = $(1 \times k)$ vector, where k is the number of system characteristics;
c = $(1 \times \ell)$ vector, where ℓ is the number of user characteristics;
L_{ij}^{M} = $(k \times 1)$ vector of system characteristics describing the level of service between points i and j by mode M;
SE = $(\ell \times 1)$ vector of user characteristics; and
a, b, c = coefficients to be estimated.

Verbally expressed, it is hypothesized that modal differences in the level of service (i.e., differences in travel times) combined with the user characteristics are the determinants of choice of mode. [Also the ratios of system characteristics can be used (1, 5, for example).]

In the logit model, the probability that choice of mode M, the dependent variable, will equal 1 (automobile choice is denoted by 1, and transit choice is denoted by 0) is

$$\Pr(M = 1 \mid ij) = e^{y}/(1 + e^{y}) \tag{4}$$

and similarly

$$\Pr(M = 0 \mid ij) = 1 - \Pr(M = 1 \mid ij) = 1/(1 + e^{y})$$

No assumptions are needed about the distributions of the variables or of y.

The probit model uses the same linear function y. If, for any given individual, $y \geq y_{crit}$, then $M = 1$; and if $y < y_{crit}$, then $M = 0$. The assumption is made that y_{crit} is normally distributed over the population. The probability that M will equal 1 is

$$\Pr(M = 1 \mid ij) = \Pr(y_{crit} \geq y \mid ij) = 1/\sqrt{2\pi} \int_{-\infty}^{y} e^{-(t^2)/2}\, dt \tag{5}$$

and similarly

$$\Pr(M = 0 \mid ij) = 1 - \Pr(M = 1 \mid ij) = 1/\sqrt{2\pi} \int_{y}^{\infty} e^{-(t^2)/2}\, dt$$

The method of maximum likelihood is used to obtain estimates for a, b, and c in both logit and probit analysis. (A computer program originally written by John Gragg, Department of Economics, University of British Columbia, and modified by Peter Stopher for CDC 6400 was used in estimating the models.)

In discriminant analysis, no dependent variable exists. The objective is to find such linear combination of the explanatory variables that their joint distribution, the distribution of y, for the 2 groups would possess very little overlap. This discrimination rule classifies an observation at y as coming from population 1 if $f_1(y)/f_2(y) > k$, and otherwise from population 0. If we assume that y is normally distributed, as is conventionally done, then, after some serious manipulation,

$$\Pr(M = 1 \mid ij) = e^{y+\ln(p/q)}/[1 + e^{y+\ln(p/q)}] \tag{6}$$

and

$$\Pr(M = 0 \mid ij) = 1/[1 + e^{y+\ln(p/q)}]$$

where p and q are the a priori probabilities of group membership 1 and 0 respectively.

DATA

Data Source

The data used were originally prepared by Lisco (2) and are well documented in his thesis; only a couple of comments are in order here.

The data consist of 159 work trips made during the morning rush hour from Skokie to the Chicago Loop. Only a binary choice (automobile or rapid transit) was available to the trip-makers. The corresponding travel times and costs were compiled for each segment of an individual trip; these disaggregate figures formed the basis for the trip segmentation.

Trip Segmentation

The system shown in Figure 1 will help explain the trip segmentation procedures. (The actual system was more complicated because of train transfers needed by some of the travelers. This does not change the principles, however.) Links 1 through 7 describe the transit network, and links 8 through 10 describe the automobile network. Times and costs associated with each link were obtained for each element in the sample. These transportation system attributes are given in Table 1.

Three user attributes—income, number of workers in household, and automobile ownership—and 1 indirect system attribute—walk access to rapid transit station—were also included in the data. These socioeconomic attributes are given in Table 2.

Twelve models were estimated by combining the travel time and travel cost differences in different ways. Of the socioeconomic attributes, dummy 1 and dummy 2 are included in every model; but if income was used, then automobile ownership was not, and vice versa.

A description of the models is given in Table 3. Next to the description column, a relationship is given to indicate how the time and cost differences were combined. Access refers to the trip from home to station, egress refers to the trip from station to work, and total access means access and egress taken together. Excess refers to the time spent outside the vehicle, for either walking, waiting, or transferring. Travel time differences are transit minus automobile, and travel cost differences are automobile minus transit. Except for the socioeconomic variables, given in Table 2, model 7 is the same as the one used by Quarmby (3), model 8 is the one used by Lisco (2), and model 9 was used by TRC (6, the TRC model used ratios instead of difference of respective travel times and costs). The system variables in these models were excess time (not in Lisco's model), total time, and total cost or out-of-pocket cost (TRC model).

RESULTS

Two kinds of evaluations were made on the basis of the results: (a) What is the best method? and (b) What is the best model? Ideally, the evaluation of the methods and the models should be done with a set of data other than that used for the model calibration. However, in the present study the original body of data was already quite small (159 individuals), and splitting that would only have left too few either for the model estimation or for the control group. An alternative procedure was adopted. It involved taking random samples from the data, computing the corresponding probabilities for each individual, summing them up to the expected value, and comparing the actual and expected values. By taking enough samples and getting the expected values and their standard deviations for each method and model, one could perform statistical tests (t-tests) to see whether there are any differences among the 3 methods or among the 12 models.

Twenty samples of 20 individuals were drawn, and the t-tests were undertaken. Unfortunately, no differences either among the methods or among the models were detected this way. The hypothesis that the results are statistically equal could not thus be rejected.

It was, therefore, decided to engineer the answer to these 2 questions. First, the methods were checked for dominance. No model could be excluded because of dominance (based on all 3 methods). The decision was then made to rank the methods and the models. This ranking was based on multiple criteria. The ranking criteria were di-

Figure 1. Transportation system between Skokie and Chicago Loop.

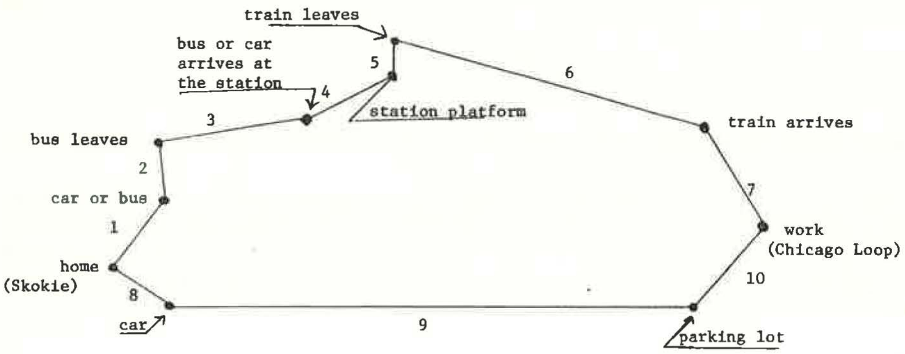


Table 1. Time and cost variables of transportation system.

Link	Description	Notation Time	Notation Cost	Note
1	Walk to car or bus	T1	—	
2	Wait for bus	T2	—	If car access to station, T2 = 0
3	Take car or bus to station	T3	C3	
4	Walk to station platform	T4	—	If walk access to station, T1 = T2 = T3 = C3 = 0
5	Wait for train	T5		
6	Line haul in train	T6	C6	C6 = train fare
7	Egress to work	T7	—	Walk was only egress mode
8	Walk to car	T8	—	
9	Drive to parking lot in Loop	T9	C9	T9 includes parking time; C9 excludes parking cost
10	Egress to work	T10	C10	Walk was only egress mode; C10 = parking cost

Table 2. Socioeconomic variables.

Description	Notation	Models Used in
Income (household or unrelated individual)	I	1 through 9
Workers per household (dummy 1, 0 if 1 and 1 if otherwise)	D1	All
Walk access to station (dummy 2, 0 if distance is 5 miles and 1 if otherwise)	D2	All[a]
Automobile ownership	A	10 through 12

[a] One-half mile was data supported walk access distance to station.

Table 3. Description of models.

Model	Time Description	Time Relationship[a]	Cost Description	Cost Relationship[a]
1	Access	T1 + T2 + T3 + T4 + T5 - T8	Access	-C3
	Line-haul	T6 - T9	Line-haul	C9 - C6
	Egress	T7 - T10	Egress	C10
2	Access	Same as model 1	Total access	C10 - C3
	Line-haul	Same as model 1	Line-haul	Same as model 1
	Egress	Same as model 1		
3	Total access	T1 + T2+T4 + T5 + T7 - T8 - T10	Total access	Same as model 2
	Line-haul	Same as model 1	Line-haul	Same as model 1
4	Excess	T1 + T2 + T4 + T5 - T8 - T10	Total access	Same as model 2
	In-vehicle access	T3	Line-haul	Same as model 1
	Line-haul	Same as model 1		
5	Excess	Same as model 4	Total access	Same as model 2
	In-vehicle	T3 + T6 - T9	Line-haul	Same as model 1
6	Excess	Same as model 4	Total access	Same as model 2
	Total	$\sum_{i=1}^{7} Ti - \sum_{i=8}^{10} Ti$	Line-haul	Same as model 1
7	Excess	Same as model 4	Total	C9 + C10 - C3 - C6
	Total	Same as model 6		
8	Total	Same as model 6	Total	Same as model 7
9	Excess	Same as model 4	Out-of-pocket	C10 - C6 - $\widehat{C3}$[b]
	Total	Same as model 6		
10	None		None	
11	None		Parking	C10
12	Excess	Same as model 4	Parking	C10

[a] See Table 1.

[b] $\widehat{C3}$ = Bus fare to station or parking cost at the station, no driving costs.

vided into 2 categories: classification criteria and criteria based on expected values. The classification criteria included 3 items:

1. Misclassified automobile users,
2. Misclassified transit users, and
3. Proportion classified correctly.

A group membership probability of ≥ 0.5 was used as a rule for correct group classification, and the correctly classified proportion was obtained based on these classification results. The classification criteria are given in Tables 4, 5, and 6 for the logit, probit, and discriminant methods respectively.

The expected-value criteria were computed based on the results of the 20 random samples and included the following:

1. Average absolute error, which = [(actual number of automobile users - expected| value of automobile users)/20];

2. Percent error, which = (1/20) [(actual number of automobile users - expected value of automobile users)/actual number of automobile users];

3. Standard deviation of error, and

4. Rank sum of the 20 predictions. [This item was obtained by ranking the results of 20 sample predictions (in case of a tie, the rank average was assigned for each tied value) and summing the ranks for each model. The lower the rank sum is, the better the model is.]

These results are given separately for each method in Tables 4, 5, and 6.

In the discriminant analysis method, 2 values have been given for classificatory measures. The upper ones were derived by using sample proportions as the a priori probabilities and the lower ones by using 0.5 as the a priori probability (Eq. 6). The latter a priori probability produces better classification results. The same does not hold true for the expected-value statistics (Tables 4, 5, and 6), however; but the sample proportions as a priori probabilities now give better results. In the ranking analyses reported below, the classification rankings correspond to those values obtained with 0.5 a priori probability.

Ranking Analysis of Estimation Methods

The methods were ranked separately by using the classification criteria and the expected-value criteria. Each column for each model was assigned a rank of 1 to 3. The best method was assigned the rank of 1. In case of a tie, the average of a rank sum was assigned for each tie. The ranks were then summed by column and row to yield a rank sum for each method. The inspection of the statistical measures given in Tables 4, 5, and 6 indicates that discriminant function always produces the lowest standard error. An experiment with another data set proved this is not always true. Hence, this criterion, expected-value criterion 3, was dropped from the ranking of the methods. The priority ranking of the 3 estimation methods is given in Table 7. The results of the ranking analysis indicate that the logit analysis is the best method. Examination of data given in Tables 4, 5, and 6 reveals, however, that the differences between logit and probit analyses are very small. In any case, logit analysis seems to slightly edge both probit and discriminant analyses.

Ranking Analysis of Model Specifications

A similar ranking as undertaken for the methods of estimation was performed for the 12 model specifications. In addition to the classification criteria and criteria based on expected values, the number of variables in the model was used in ranking the models to input (approximately, of course) the data collection and model estimation costs. Ranks run from 1 to 12; no averaging was done for ties, but the lowest tied rank was assigned for all ties. The results of this ranking analysis are given in Table 8.

Three comments may be made on the basis of these ranking results. First, models without any explicit system variables appear to be best accorcing to the classification criteria. Model 10 and model 11 rank as first and second. According to expected-value

Table 4. Ranking criteria for logit estimation method.

	Classification Criteria			Expected-Value Criteria			
Model	1	2	3	1	2	3	4
1	20	16	0.774	1.04	0.3	0.88	142.5
2	20	17	0.767	1.02	0.5	0.90	130
3	22	18	0.748	1.08	0.7	0.94	143
4	21	14	0.779	0.98	1.1	0.93	118
5	21	14	0.779	0.97	0.7	0.92	117.5
6	21	14	0.779	0.98	0.8	0.93	112
7	20	14	0.786	0.98	0.7	0.91	116.5
8	23	14	0.767	1.10	0.8	0.98	140.5
9	22	13	0.779	0.96	1.3	0.97	118.5
10	21	16	0.767	1.07	0.8	1.16	115.5
11	18	17	0.779	1.15	0	1.15	144
12	21	16	0.767	1.12	0.9	1.24	151

Table 5. Ranking criteria for probit estimation method.

	Classification Criteria			Expected-Value Criteria			
Model	1	2	3	1	2	3	4
1	22	16	0.762	1.05	0.5	0.87	138.5
2	21	16	0.767	1.02	0.5	0.89	129
3	22	18	0.748	1.08	0.7	0.93	141.5
4	22	14	0.773	0.99	1.1	0.92	118
5	22	15	0.767	0.99	0.9	0.91	121
6	22	14	0.773	1.00	0.9	0.91	122.5
7	20	13	0.792	0.99	1.1	0.89	119
8	23	14	0.767	1.10	0.9	0.97	144
9	22	13	0.779	0.98	1.6	0.96	119
10	21	16	0.767	1.06	0.4	1.11	110.5
11	23	15	0.762	1.15	1.4	1.10	149
12	21	16	0.767	1.11	0.1	1.20	148

Table 6. Ranking criteria for discriminant estimation method.

	Criteria			Expected-Value Criteria			
Model	1	2	3	1	2	3	4
1	38 12	8 30	0.710 0.736	1.09	0.8	0.48	134.5
2	37 12	7 29	0.729 0.743	1.08	0.7	0.50	128
3	36 13	7 29	0.729 0.736	1.08	0.5	0.55	119
4	44 13	5 31	0.691 0.724	1.05	0.7	0.44	122.5
5	44 15	3 29	0.704 0.724	1.06	0.8	0.42	131
6	44 14	4 31	0.698 0.717	1.06	0.8	0.43	137
7	40 15	6 29	0.710 0.724	1.07	0.5	0.44	132
8	40 13	11 34	0.680 0.704	1.11	0.3	0.59	123.5
9	37 14	10 28	0.704 0.736	1.06	0.3	0.52	130.5
10	58 21	1 16	0.630 0.767	1.08	0.7	0.53	124
11	44 17	5 18	0.691 0.779	1.10	0.9	0.50	135
12	47 14	2 23	0.691 0.767	1.09	0.9	0.49	146

Table 7. Ranking of estimation methods.

		Rank Sum by Criteria		
Method	Rank	Classification	Expected Value	Total
Logit	1	63.5	64.5	128
Probit	2	72	75.5	147.5
Discriminant	3	80.5	76	156.5

Table 8. Ranking of model specifications.

	Model by Criteria		
Rank	Classification	Expected Value	Total
1	10	7	7
2	11	5, 10	10
3	7	4	4, 5, 10
4	9	9	6
5	5	6	2

Table 9. Coefficients and standard deviations of model 5.

	Logit		Probit		Discriminant	
Variable	Coef.	S. D.	Coef.	S. D.	Coef.	S. D.
Constant term	-0.384	1.722	-0.337	1.040	-0.358	—
Time						
Out-of-vehicle	-0.316	0.114[a]	-0.173	0.0639[a]	-0.114	0.147
In-vehicle[b]	-0.0206	0.0559	-0.0165	0.0321	-0.0085	0.0037[a]
Total						
Cost						
Access[c]	0.0158	0.0088	0.0087	0.0052	0.006	0.0048
Line-haul	0.0414	0.0248	0.0249	0.0148	0.0170	0.0129
Total						
Income[d]	-0.0136	0.0056[a]	-0.0074	0.0033[a]	-0.0056	0.0018[a]
Automobile ownership						
2 workers per household	1.910	0.507[a]	1.087	0.284[a]	0.602	0.216[a]
Walk access	2.176	0.780[a]	1.277	0.447[a]	0.711	0.226[a]

[a] Significant at 0.05 level.

[b] Time differences are transit time minus automobile time in hundreds of seconds.

[c] Cost differences are automobile cost minus transit cost in cents.

[d] In thousands of dollars.

Table 10. Coefficients and standard deviations of model 9.

Variable	Logit Coef.	Logit S. D.	Probit Coef.	Probit S. D.	Discriminant Coef.	Discriminant S. D.
Constant term	1.075	1.143	0.584	0.694	0.288	—
Time						
Out-of-vehicle	-0.271	0.141	-0.145	0.0806	-0.103	1.34
In-vehicle[b]						
Total	-0.0327	0.0552	-0.0236	0.0319	-0.0157	0.0062[a]
Cost						
Access[c]						
Line-haul						
Total	0.0165	0.0088	0.0092	0.0051	-0.0068	0.0034
Income[d]	-0.0157	0.0054[a]	-0.0087	0.0031[a]	-0.0070	0.0023[a]
Automobile ownership						
2 workers per household	1.837	0.501[a]	1.032	0.278[a]	0.722	0.226
Walk access	1.996	0.769[a]	1.174	0.443[a]	0.684	0.217[a]

[a] Significant at -0.05 level.
[b] Time differences are transit time minus automobile time in hundreds of seconds.
[c] Cost differences are automobile cost minus transit cost in cents.
[d] In thousands of dollars.

Table 11. Coefficients and standard deviations of model 10.

Variable	Logit Coef.	Logit S. D.	Probit Coef.	Probit S. D.	Discriminant Coef.	Discriminant S. D.
Constant term	2.218	0.573	1.220	0.319	0.725	—
Time						
Out-of-vehicle						
In-vehicle[b]						
Total						
Cost						
Access[c]						
I[illegible]-haul						
[illegible]						
In[illegible]						
Automobile ownership	-1.460	0.343[a]	-0.797	0.184[a]	-0.592	0.126[a]
2 workers per household	1.660	0.476[a]	0.940	0.268[a]	0.623	0.194[a]
Walk access	1.624	0.663[a]	0.963	0.376	0.511	0.162[a]

[a] Significant at -0.05 level.
[b] Time differences are transit time minus automobile time in hundreds of seconds.
[c] Cost differences are automobile cost minus transit cost in cents.
[d] In thousands of dollars.

Table 12. Means of selected transit system variables.

Variable	Value[a]	Variable	Value[a]
Total travel time	3,624	Total travel cost	58.2
Out-of-vehicle time	749	Line-haul cost	38.2
In-vehicle time	2,875	Total access cost	20.0

[a] Time is in seconds, and cost is in cents.

Table 13. Effect of changes in travel time and cost on transit proportion.

Model	Transit Time Up 30 Percent	Automobile Time Down 23 Percent	Transit Cost Up 30 Percent	Automobile Cost Down 23 Percent	Remarks
Quarmby	- 9.0			-8.0	Reported by McGillivray
Wohl-Kraft	- 7.3	-6.2	-1.6	-2.0	Reported by McGillivray
McGillivray	-7.5		-5.6		Reported by McGillivray
Warner	-6.4		-6.4		Reported by McGillivray
Lisco	-34.6		-8.1		Change in total travel time and total travel cost
Model 7	- 8.8		-7.0		Change in total travel time and total travel cost
Model 7	-15.0				Change in excess time

Table 14. Value of time.

Model	Method	Value of Time ($/hour)	Standard Error of Value of Time	Percent of Wage Rate
7	Logit	0.71	1.17	12.5
	Probit	0.22	1.23	16.0
	Discriminant	0.83	—	14.0
8	Logit	3.49	3.02	62
	Probit	3.44	2.90	61
	Discriminant	3.60	—	64
9	Logit	0.62	0.82	11
	Probit	0.71	0.86	12.5
	Discriminant	0.65		11.5

criteria, however, models that have explicit system performance variables rank best. Model 7 and model 5 rank as the first and second.

Second, it is interesting that model 10, which has only 1 system variable (a dummy for walk access), and user attributes (car ownership, which probably is not system independent, and number of workers in the household), seems to be as good as models 5 and 7. Thus, the decision to buy a second car often signifies car choice, and the decision to reside near the transit station signifies transit choice. These choices are, of course, conditioned by socioeconomic factors such as the number of family members in the labor force and income. The effect of the extent of the public transportation system on modal choice may be negligible at present. The good performance of model 10 is not of spurious nature, as the actual survey proved (7, p. iii):

> This study indicated that diversion of Loop-bound trips ... is made mainly from other rapid transit modes, along with suburban railroads and buses, with only a small number being diverted from automobile trips.

Third, the results of the ranking analysis indicate that model 12, which included only those variables that were statistically significantly (at the 0.05 level) different from 0, consistently showed poor performance. The customary null hypothesis, b = 0, may not be a good one. There seems to be a reason to remember that 0 is a very particular coefficient; it implies no relation between the dependent and the explanatory variables. Why should a variable be excluded solely because its coefficient, which is the maximum likelihood estimate, has a wide standard error? In this study, all the variables were originally included because they should have an effect on modal choice and not because they were available in the data set. Coefficients and standard deviations are given in Tables 9, 10, and 11 for models 5, 7, and 9.

COMPARISONS WITH OTHER SIMILAR MODELS

McGillivray (5, p. 40, Table 13) computed changes in modal split for the policies ot increasing relative travel time by 30 percent and increasing relative travel cost by 30 percent for 4 models. These results are reproduced here along with 3 new results for the same policies. The new results are those for Lisco's original model, for model 7 when the change is in total travel time and total travel cost, and for model 7 when the change is in excess time (Tables 12 and 13).

With regard to a 30 percent increase in transit time, it appears that the 4 results reported by McGillivray are largely identical; models developed by Quarmby, Wohl-Kraft, McGillivray, and Warner estimate the change in transit proportion to be between 6½ and 9 percent. A much different result is obtained by using Lisco's model; there the same increase in transit time would decrease the transit proportion by 34½ percent. Model 7 estimates the decrease in transit proportion to be nearly 9 percent in response to the 30 percent increase in total transit time; a result similar to those reported by McGillivray. However, if the change is in excess time, then model 7 indicates a 15 percent reduction in transit proportion.

With respect to travel costs, the results reported by McGillivray indicate that a 30 percent increase in transit cost (or 23 percent decrease in automobile cost) would decrease the transit proportion by 6 to 8 percent. There is an exception: The Wohl-Kraft model estimates this change to be about 2 percent. The results obtained by Lisco's model and model 7 give support to the 6 to 8 percent figure; the percentages for Lisco's model and model 7 are 8 and 7 percent respectively.

Three comments are in order on the basis of these results. First, model 7, where the travel time was broken down into excess and total travel time components, suggests that travelers value excess time differently from in-vehicle time. This result was obtained also by Quarmby and Kraft and Wohl. These models, Quarmby, Wohl-Kraft, and model 7, also estimate an identical magnitude for the change in transit proportion in response to a change in total travel time. Second, it is somewhat surprising that models by McGillivray and Warner on the one hand and by Lisco on the other estimate such different responses, even though the models are largely similar and do not include the excess-time term. McGillivray and Warner models give results similar to the other models mentioned above, but Lisco's model estimates modal choice to be much more sensitive to changes in travel time. Third, changes in modal split in response to changes in travel cost appear to be equal by all the modal-choice models; however, Wohl and Kraft, who use a different type of model, estimate less sensitivity with respect to travel cost.

VALUE OF TIME SAVED

Values of time saved, or more commonly values of time, were computed from model 7 (excess time, total time, and total cost), from model 8 (total time and total cost), and from model 9 (excess time, total time, and out-of-pocket cost). The results, including the standard error of the value of time and value of time as a percentage of the wage rate, are given in Table 14.

Value of time, in dollars/hour, was computed as $(b_1/b_2) \times 0.36$, where b_1 and b_2 are the estimated coefficients of total travel time difference and total cost difference respectively. Variance (b_1/b_2) is computed from the following formula:

$$\mathrm{Var}\ (b_1/b_2) = 1/b_2^4\ [b_2^2\ \mathrm{var}\ (b_1) + b_1^2\ \mathrm{var}\ (b_2) - 2b_1b_2\ \mathrm{cov}\ (b_1b_2)] + 1/b_2^3\ [b_1\ \mathrm{var}\ (b_2) + b_2\ \mathrm{cov}\ (b_1b_2)]$$

It appears from the results that all 3 methods of estimation produce approximately the same values of time. Models 7 and 9, which both have excess and total travel time variables, with the cost term being total cost difference in the former and out-of-pocket cost difference in the latter, obtain largely equal values of time. The average value of time for model 7 is 82 cents/hour and for model 9 is 66 cents/hour. These ratios are about 14 and 12 percent of the wage rate respectively.

Model 8, which has total travel time and total travel cost variables but no excess time variable, obtains a much higher value of time: approximately \$3.50/hour or 62 percent of the wage rate. (In Lisco's study the time value was 40 percent of the wage rate, and in Quarmby's 20 to 35 percent.)

The standard errors of all value of time estimates are quite large. Data given in Table 11 show that the standard errors are about equal to the values of time themselves. Two comments may be made on the basis of the results. First, the out-of-vehicle and in-vehicle times must have extremely different values because of the large difference in value of time depending on whether it was computed from a model where excess time is explicitly accounted. This result was also indicated by the results obtained in the previous section, where choice of mode was much more sensitive to out-of-vehicle than in-vehicle times. From the coefficients of models 4, 5, 6, 7, and 9, it may be inferred that out-of-vehicle time is 6 to 15 times the in-vehicle time, most of the values being around 7. [Quarmby found the out-of-vehicle time to be 2.5 to 3.0 times the in-vehicle time. In Ergün's recent study the value of walking time was estimated to be between \$4.50 and \$11.50 (8). This result is in general agreement with the findings of this paper.] Second, in spite of the large standard errors estimated for the value of time in this study, it appears that the value of time savings may not be so large as previously believed. This concerns especially the in-vehicle time. Therefore, to be realistic any economic study of a transportation improvement must consider the out-of-vehicle and in-vehicle times separately.

CONCLUSIONS

Three major conclusions of this paper are as follows:

1. The methods of estimation, commonly used in probabilistic modal-choice models, probit, logit, and discriminant analyses, all yielded comparable results. Any of them can be used with equal success.
2. At present, the modal-choice behavior of travelers appears to be only marginally influenced by the travel times and travel costs. This, in turn, implies that travel service by automobile and transit are not perfectly substitutable services and should, therefore, be modeled separately.

3. The values of travel times obtained are substantially lower than those previously reported. In particular, the value of out-of-vehicle time is much different from the value of in-vehicle time. This fact should be recognized in any economic study of a transportation improvement if travel time is given a monetary value.

REFERENCES

1. Warner, S. Stochastic Choice of Mode in Urban Travel: A Study in Binary Choice. Northwestern Univ. Press, 1967.
2. Lisco, T. The Value of Commuters' Travel Time: A Study in Urban Transportation. Dept. of Economics, Univ. of Chicago, PhD dissertation, 1967.
3. Quarmby, D. Choice of Travel Mode for the Journey to Work: Some Findings. Jour. of Transport Economics and Policy, 1967.
4. Stopher, P. R. A Probability Model of Travel Mode Choice for the Work Journey. Highway Research Record 283, 1969, pp. 57-65.
5. McGillivray. Classification of Trips by Mode. Dept. of Economics, Univ. of California, Berkeley.
6. Hill, D. M., and Von Cube, H. G. Development of a Model for Forecasting Travel Mode Choice in Urban Areas. Highway Research Record 38, 1963, pp. 78-96.
7. Village of Skokie Mass Transit Demonstration Project. Chicago Transit Authority.
8. Ergün, G. Development of a Downtown Parking Model.

APPROACH TO PROBABILITY DISTRIBUTION OF VALUE OF WALKING TIME AND PEDESTRIAN CIRCULATION MODELS

Jean Francois Allouche, Bureau Central d'Etudes pour les Equipments d'Outre-Mer, Paris

•TIME is money. So goes the dictum, but not until engineers began to apply economic analysis to transportation plans was there any real concern about how much money it was worth. The quantitative concept of a value of time has been used for some time but mostly by engineers trying to improve and rationalize manufacturing processes. For those analyses, the value of time can be easily taken into account through salaries, prices, and interest rates; time saved is equivalent to labor saved or additional production or shorter hours and faster capital turnover.

But the transportation system moves people—people as consumers much more than people as agents of production—and only their behavior can tell what subjective value they assign to their own time. [In most U.S. urban areas, for instance, truck trips represent about 5 percent of all vehicle trips, and business trips amount to approximately 10 percent of person movements. Work trips, preempting leisure time and not production time, do not, at least under our present social organization, fall in the "production" category (12, p. 81).] This is more than an academic problem; indeed, one of the most important quantifiable benefits in highway programs, a part of the transportation system, consists of savings in travel time.

Attempts have been made to measure the value of time, or rather the values of time. Factual research as well as psychological inference indicates that our valuation of time is influenced by a variety of factors, ranging from the types of activities we pursue to our levels of income and the amounts of time being saved. If we restrict our discussion to time spent in transportation, the perceived value of time depends on the purposes and the conditions of travel as well as on personal factors.

Most investigations performed in the past (1, 2, 3, 4, 5, 6, 7) considered the value of time as a constant. Two papers only treated the value of time as a stochastic variable. Pratt (8), using the central limit theorem approach in statistics, hypothesized that the interplay of numerous subjective factors, including value of time and inconvenience cost, would result in a normal distribution of the "catch-all index" by which individuals compare 2 travel modes (the index being 0 when a person finds both modes equally attractive). St. Clair and Lieder (9), studying a toll versus free highway situation, assumed a normal distribution for both the value of time and the inconvenience cost (as measured by the number of speed changes on each alternative route). By successive trials, they determined, for the mean and standard deviation of both distributions, values that would yield the closest approximation to observed route choices. Their approach, however, does not allow for a statistical test of confidence of their assumption. A second round of analysis performed by Thomas on the data used in his original study (7) of commuter's values aims at defining the value of time as a variable function of income and of the amount of time saved, but it is not a probability distribution. Apparently, there has been no attempt to determine directly an empirical probability distribution for the value of time.

Also, from a different standpoint, it is remarkable that most researchers have confined their investigations to the value of driving or riding time. With the recent upturn of interest toward public transit in urban areas and the growing attention given to people-mover systems, it becomes equally important to estimate the value of walking time. (People-mover systems are very short-haul facilities and can best be described as distributor systems within major activity centers such as the CBD, shopping centers, campuses, and airports.)

The present paper proposes a method that can be used to determine the actual probability distribution of the value of walking time to motorists and has no assumption as to its mathematical form. The resulting curve then either can be used directly in modeled-choice procedures or can be submitted to statistical tests in order to determine the best mathematical approximation for incorporation in a predictive model of pedestrian behavior.

THE MODEL

The model is based on the behavior of car drivers selecting a parking location before they walk to their destinations. In a first-stage approximation, we will assume that only parking fee and walking time influence the decision-maker. This eliminates factors such as weather, environmental quality, or street gradient. It is legitimate to disregard grade if the study area is reasonably level, and, except with shoppers, it is unlikely that parking decisions take much account of the environmental quality along the path followed to on-foot destination.

Formulation

We hypothesize that a driver tends to minimize his total generalized cost (money and time).

$$C = c + xd \tag{1}$$

where

C = total cost,
c = parking fee for desired parking duration,
d = distance from garage to on-foot destination, and
x = disutility cost of walking 1 unit of distance.

If a driver selects parking facility 1, parking fee c_1, and distance d_1 from his destination, rather than facilities 2, 3, . . . , N, we assume that

$$c_1 + xd_1 \leq c_j + xd_j \qquad \text{for all } j = 2, 3, \ldots, N \tag{2}$$

Three cases are possible:

1. If $d_1 - d_j > 0$, then Eq. 2 implies that

$$x \leq (c_j - c_1)/(d_1 - d_j) \tag{3}$$

2. If $d_1 - d_j = 0$, then Eq. 2 implies that

$$c_1 \leq c_j$$

i.e., if our original assumption of rational trade-offs between time and walking is correct, the driver should select the cheaper facility. But this tells nothing about x.

3. If $d_1 - d_j < 0$, then Eq. 2 implies that

$$x \geq (c_1 - c_j)/(d_j - d_1) \tag{4}$$

Each parking decision then results in a list of inequalities. The whole set of inequalities imposes on x a lower bound or an upper bound or both. An additional constraint on x is that, inasmuch as the model is valid, no negative value should be accepted for the distance disutility cost. This assumption is validated by most people's behavior: They like to park close to their destinations. Therefore, in all cases we have

$$L \leq x \leq M \tag{5}$$

or

$$L \leq x \tag{6}$$

Graphically, the argument goes as follows (Fig. 1): Total cost for a driver parking in facility i and walking to building j is

$$g_{ij} = c_i + xd_{ij} \tag{7}$$

For a given parker, j is fixed and Eq. 7 can be represented by a straight line in the coordinate system x, g. All parking facilities can be so represented. Whatever x, the disutility cost to him, the driver will consider only the minimum cost curve [PRST (F), Fig. 1].

If the driver selects facility C [line C, Fig. 1], then x, his valuation of distance disutility, must belong in the range r to s, for this is where C is the minimum cost curve. Selection of F is the case exemplified by inequality (Eq. 6). It will deserve special treatment. Selection of A would mean that the driver has a negative distance disutility cost. Selection of E is merely unaccountable by the model in its present form. E is never the minimum-cost solution and is, therefore, termed a noncompetitive facility under the circumstances. Both of these aberrant cases will be briefly discussed hereafter.

As already mentioned, a negative disutility cost of distance is at odds with the empirical observed behavior of an overwhelming majority of drivers. Negative disutility cost values or selection of a noncompetitive facility can be interpreted without throwing the model away, however. They can arise from a nonuniform distortion in distance perception on the part of the driver. If there are only a few occurrences of drivers with negative costs, we will just dismiss the datum and tally only the percentage of people falling in that category for later use in an assignment model.

The 2 aberrant cases can also arise from what we might term blurred rather than biased perception. People are probably little sensitive to small differences in distances and may, therefore, make decisions apparently at odds with the "numbers." If negative disutility cost values are observed in significant number, we propose to investigate whether it can be traced to some slackness in sensitivity to distance by defining a sensitivity threshold, for instance.

Histogram

A histogram is a statistical representation describing the number of observations falling within a certain range (a to b). These observations are represented by a rectangle, the area of which is proportional to the number of observations and one side of which is the interval (a to b) on the horizontal axis of the graph.

Figure 2 shows a distribution with 3 groups of observations: In group 1, 10 observations are between 0 and 5; in group 2, 20 observations are between 5 and 10; and in group 3, 20 observations are between 10 and 20. Group 3, being spread over an interval twice as large as that of group 2, is assigned an ordinate twice as low.

The limit of a distribution's histogram is a probability density function when intervals multiply ad infinitum and their width tends toward 0. In other words, the histogram constitutes an approximation to the probability density function after the area under its perimeter has been normalized to 1 (by rescaling ordinates).

Such a histogram could be constructed step by step from the inequalities (Eq. 5) attached to each surveyed parker. Each parker is considered as being one observation, and this procedure does not affect the final configuration of the histogram. Parkers with the same parking location and the same on-foot destination could be considered as a group at this stage, but later considerations will call for individual processing. The range for the group is assumed to be the interval defined by the set of inequalities. If, as probable, the data come from a sample survey, each observed parker has to be weighted with the appropriate factor to expand sampled data to the whole population of parkers.

The problem of parkers with an upper unbounded disutility cost of walking (inequality, Eq. 6) can be dealt with in 2 ways.

1. By estimating an absolute highest disutility cost based on the highest upper bounds observed among other parkers and based also on common sense rationales,

Figure 1.

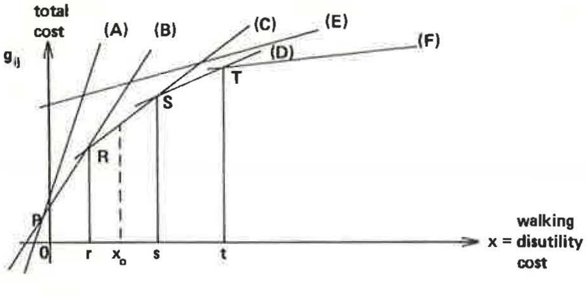


Figure 2.

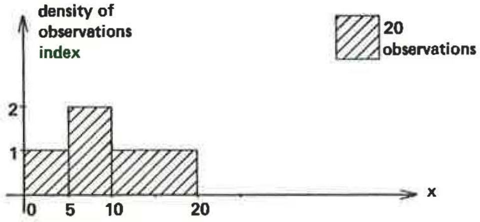


Figure 3.

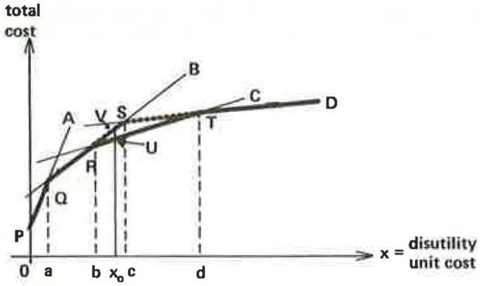
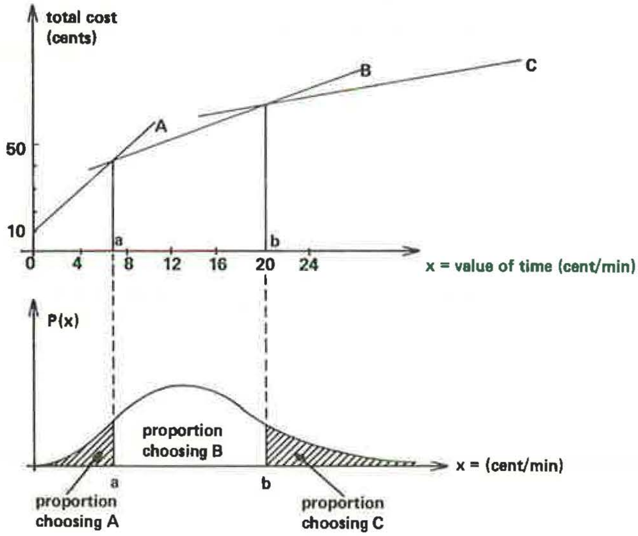


Figure 4.



e.g., a value that would make a taxi ride a preferable alternative; and

2. By calibrating a value of the upper bound that, in turn, would produce a best fit when the histogram is used for predictive purposes (this will be developed later in this paper).

Distribution of the Value of Time

In the preceding section we have proposed a method to build the distribution of the disutility cost of walking based on distance. An assessment of walking speed is necessary to derive the distribution for the value of time.

Existing studies (11) show that pedestrian travel speeds (averaged over a complete portal-to-portal walk because speeds can vary significantly from block to block during the same walk) can be considered normally distributed, with significantly different means for men (4.93 ft/sec) and women (4.53 ft/sec). Standard deviations are similar (approximately 45 percent of the value of the mean). Here again, the problem can be approached in several ways:

1. The speed is assumed to be constant and equal to the mean of the observed distributions for men and women separately, if this information is reported in the survey;
2. Walking speed is considered to be a normally distributed random variable with known mean and standard deviation and is assumed to be statistically independent of walking disutility cost; and
3. Walking speed is considered to be linearly correlated in a positive way with distance disutility cost (the rationale on which this assumption is based is developed in the brief discussion hereafter).

Regarding the first approach (constant average speed), once a unique speed has been set (possible one for each sex and purpose), multiplying the horizontal scale by that speed factor will transform dollars/ft into dollars/min, thereby yielding the distribution of the value of time. Some studies have indicated a variation of walking speed with time of day, but this seems to be related primarily to trip purpose. If warranted, the model can be applied separately to each trip purpose.

Regarding the second approach, we briefly discuss the assumption of independence between walking speed and distance disutility cost, on which it is predicated. A rapid and lively pace is often correlated with a certain liking for walking. It seems logical to assume that a great liking for walking is linked to a lower distance disutility cost. Conversely, a slow walking speed would be associated more often with a dislike for walking and a high distance disutility cost. The assumption of independence appears at best as a convenient simplification, pending a careful and much-needed test of its validity. If the kind of statistical correlation depicted above does exist but has to be overlooked for convenience reasons, the distribution of time value derived under the assumption of statistical independence will be flatter and more widespread than the "true" distribution. The computations required to develop the distribution under the assumption of statistical independence is described in the Appendix.

The rationale in favor of a correlation (i.e., the third approach) is developed in the preceding paragraph. If the first and second approaches yield unsatisfactory results, it is possible to single out some of the observed parkers and track them later to measure their average walking speed. (In a later section, tests will be suggested to evaluate the reliability of the distribution produced.) Conditional distributions of distance disutility cost could be constructed for people with given walking speeds, and their correlation with speed analyzed. This is a long and costly operation, but there is good reason to think that, if such a correlation does exist, it depends primarily on the culture or the distribution of temperamental features among the population and therefore is rather constant from place to place, at least within a culturally homogeneous domain such as the whole of North American cities. Several definitions for stability are suggested in the Appendix. At any rate, once this tedious investigation is performed, the student of different cities could dispense with this special kind of tracking survey and use only standard parking surveys for data gathering.

PRELIMINARY EVALUATION OF THE MODEL

Possibly the utmost advantage of the approach outlined in this paper is the very low cost of data collection. Practically all of the information required is available from standard parking surveys conducted at intervals in major American city center areas. Minor changes such as additional questions dealing with sex, age, or income could be accommodated at practically no cost if one wanted a stratified application of the model. Data processing would be slightly more sophisticated than for a standard parking study, but the basic program for histogram building is neither sophisticated nor very long and can be used again in different studies.

Many situations in which walking can be traded off for money in a downtown environment are of a multiple-choice nature. More than 2 modes are often available. More than 2 routes can easily be envisioned. The parking decision on which this paper is predicated also involves multiple competition as soon as there are more than 2 parking facilities open to the public. This multiple-choice character inherent to the problem thus precludes the use of discriminant analysis. On the other hand, competition among many garages is beneficial to our approach; the more facilities there are, the better our chance is to define a narrow interval containing the driver's (unknown) value of time.

The model proposed here has conceptually more explanatory power than models based on discriminant analysis, although the latter may produce relevant predictions at the aggregate level. A discriminant analysis type of model "does not specifically estimate the route choice for the individual motorist; rather, it predicts the action expected of an 'average' motorist when faced with the given route choice situation" (7 p. 59, Thomas comments about his own model only, but the statement can be applied to any model of the discriminant analysis type). On the other hand, if supplied with a decision-maker's actual value of time, our model is capable of predicting that person's decision. In our model, probabilities do not reflect people's indeterminacy or "unpredictability" (i.e., people completely and identically defined in terms of Thomas' mod formulation have to be arbitrarily assigned opposite decisions in order for his model to work satisfactorily) but rather uncertainty on our part as to their (well-defined) time valuation or other similar parameters.

Using a parking fee as the dollar element of our "total cost" function eliminates the trap of "perceived costs" in which so many previous studies have fallen when, for instance, car-operating costs are used.

We now turn to the shortcomings or difficulties we see involved in the model.

The first problem involves time perceived versus time actually spent (or, for that matter, distance perceived versus actual distance). All models share this problem. However, if the decision-maker systematically overestimates or underestimates all actual time durations by the same factor, this merely amounts to scaling down or up by that factor the value of time computed from actual duration. The revised value is the person's value for each actual unit of time and can be used sensibly for evaluation purposes. In that case, adding a question about perceived walking time in the parking survey should take care of the problem inasmuch as people are sensitive to short-time durations and capable of estimating them within a reasonably narrow range. (We refer here not to people's biases but to their own indeterminacy. It is a case of "blurred" perception.) This procedure would eliminate the need for walking speed analysis, as developed above. Both could be later synthesized in an improved version of the model, as will be discussed later. Perception biases that depend on circumstances cannot be integrated systematically in the model and will be assumed away, together with the multitude of other random influences.

The influence of the weather on the disutility cost of walking (which often takes place outdoors) does not need demonstration. Using data gathered on a rainy day would lead to the value of time spent carrying an umbrella or dodging raindrops. Although this is not uninteresting, it is suggested that the first analysis be conducted on data pertaining to moderate weather conditions. Results will be usable for a larger set of circumstances, including semicovered malls and indoor facilities.

It has been argued that short-term and long-term parkers do not have the same value of time. Although it is obvious that they do not attach the same value to parking time, is not evident that their valuations of walking time are different. A discussion in the Appendix demonstrates that, as parking duration increases, cheaper and more distant facilities become preferable. This establishes the necessity of treating each interviewed parker separately to assess his exact parking fee, either by direct questioning or by using rates and recorded parking duration. Short-term and long-term parkers could, of course, be processed separately, if so desired, and separate histograms prepared.

Drivers occupy parking facilities on a first-come, first-served basis. At certain times during the day, some facilities are saturated and the driver is faced with only a restricted supply. Figure 3 shows how this may bias the estimated range for the driver's distance disutility cost to the point where the actual value lies outside the estimated range. If all facilities were available, the minimum-cost domain is the line PQRTD, and a driver with a disutility cost x_0 selects facility C. As in the earlier section, we reason that his disutility cost range is b to d. When C is removed from the supply, the minimum-cost curve becomes PQRSTD, and the driver selects facility B. If we do not know that C is out of the supply, we will interpret his decision as evidence that his disutility cost range is a to b, which is erroneous. One could keep track of facility saturation by hours of the day in the survey, but it does not seem practical at this point. Moreover, saturation is not a stable condition: C may be saturated a moment and then become open again as B becomes saturated, and so on. The bias tends to smooth out the histogram rather than change its balance. Figure 3 shows that range a to b was mistaken for the "true" range b to d. But it is so only because x_0 is inferior to c. Had x_0 been superior to c, the driver would have selected facility D, thereby leading us to mistakenly assume range d to $+\alpha$ instead of the true range b to d. The net result is that people who should have been distributed in the b to d interval will now be represented on either side of the histogram.

MODEL APPLICATION

The probability distribution of the value of time (or walking distance disutility cost) can be used in many applications, which can be roughly categorized in 2 groups.

1. Economic analyses estimating value of time savings accruing to pedestrians due to an improvement in pedestrian circulation, such as an overpass, a trail system, or a people mover, or evaluating alternate plans for a primarily pedestrian-oriented facility, such as a hospital, university campus, or civic center. (In case of a people mover, our curve would give a lower limit because the action of walking is considered a hardship by many, regardless of the time involved; a people mover mitigates this hardship.)
2. Pedestrian assignment models based on a total-cost-function assignment of drivers with known on-foot destinations to parking facilities (enabling a comprehensive treatment of parking schemes and pedestrian systems) and an estimation of pedestrian traffic diverted to a new facility, such as an overpass, or a people mover.

We use drivers' assignment to parking facilities as the example because it will help explain calibration procedures. Assume an average walking speed determined by the driver's physical characteristics (sex and age). Assume also 3 parking facilities A, B, and C and a group of drivers who have identical characteristics, have destinations in the same building, and are willing to purchase 1 hour of parking. Parking rates, walking distance, and time are as follows:

Facility	Parking Rate ($/hour)	Walking Distance (ft)	Walking Time (min)	Total Cost Function
A	0.10	1,250	5	$0.10 + 5x$
B	0.30	500	2	$0.30 + 2x$
C	0.50	250	1	$0.50 + x$

Figure 4 shows the minimum cost curve and the probability density function of x, the value of time, as functions of x. People with a value of time lower than a will select facility A. People with a value of time greater than b will select C. People with a value of time between a and b will select B. Their proportions are represented by the areas delimited by the probability density curve P (x), the x axis, and abcissas a and b. Knowing P (x), we can compute these proportions and can accordingly assign the group of drivers under study to the various facilities.

In summary, the inputs required for this application are the pattern of final trip destinations and parking durations; the locations of parking facilities; parking rates for each facility; and the probability density function of the disutility cost of walking, or the probability density function of the value of time, plus some information about walking speeds (average or probability distribution). All inputs but the last are currently available. The output is an assignment to parking facilities of drivers with destinations in given buildings.

CALIBRATION

So far we have mentioned only 1 degree of liberty for the model: the absolute maximum M, the value of time to be used in cases where inequalities (Eqs. 5 and 6) do not define an upper bound for x. (This section deals with value of time, but is readily applicable to the calibration of a model based on walking disutility cost.)

It is possible to define an index of performance for the model with respect to which parameter M can be adjusted. To that end, we apply the model to a parker-assignment problem in a situation where a parking survey with origin-destination data is available.

Let Y_{ij} be the number of people observed in the survey walking from parking facility i to building j (the pair i-j is called an interchange). Let X_{ij} be the number of people assigned by the model to interchange i-j. Ideally $X_{ij} = Y_{ij}$. If $X_{ij} > Y_{ij}$, then $X_{ij} - Y_{ij}$ people have been erroneously assigned, thus creating an imbalance $X_{i'j'} < Y_{i'j'}$ in another pair i′-j′. Thus,

$$I = \left(\sum_{ij} |X_{ij} - Y_{ij}|\right) \Big/ 2\sum_{ij} Y_{ij}$$

is the percentage of people erroneously assigned. (The number of pairs i-j may be very large. In some cases, it may be more practical to group interchanges by volume categories or into screen lines and apply correlation analysis to the aggregate.)

Each value of M will yield a certain value I (M) for the performance index. Trying several values of M over a reasonable range will give evidence of a trend for I (M). It can be shown that I (M) has an asymptote when M tends to infinity. This in turn will enable us to determine a value of M and to maximize I. If the fit between observed and synthesized interchange patterns is not satisfactory, another degree of liberty can be provided.

So far we have used 0 as the lower bound for ranges that lack one. But there is probably a nonzero minimum value, L, to people's value of time.

To calibrate the model with respect to M and L at the same time requires that some maximizing techniques be borrowed from nonlinear programming (such as steepest ascent or reduced gradient) so that the I (M, L) surface can be "climbed" on. They generally entail much computer time. Besides, in our problem the computation of 1 point on the surface already involves a complete assignment run plus a performance analysis (the latter is relatively inexpensive). This makes the feasibility of a double calibration dependent on the cost of running the basic assignment program.

POSSIBLE IMPROVEMENTS

Stratification of the data by trip purpose, sex, age, and income may uncover significant variations in the probability distribution of the value of time with respect to these factors. Logical considerations suggest the direction, if not the extent, of the influence of these factors on the value of time.

A joint study of walking speed v, time perception bias b, and distance disutility cost z would give information about the mutual correlations of these 3 characteristics in a given individual. Correlations between v and z and between b and v could be considered as stable among cities (within a given culture, e.g., Northern Europe or South America). They depend mainly on cultural, noneconomic features. The mathematics of a model based on such premises is developed in the Appendix.

Street grade plays a double part in decisions entailing walking. First, it increases the actual walking time, and, second, it makes walking more exhausting and thereby increases its disutility cost. In a first approximation, we may assume that this increase Δz in disutility cost is directly related to the street gradient g.

$$\begin{aligned} \Delta z &= ag && \text{if } g \geq 0 \\ \Delta z &= 0 && \text{if } g < 0 \quad \text{(walking down a street)} \end{aligned} \tag{8}$$

and total cost for a parker would now be

$$C = k\theta + tz + t(g)\,\Delta z \tag{9}$$

where t(g) is the time effectively spent walking on a street with gradient g, t is total walking time, and θ is parking duration. The ideal situation to test this model and calibrate the parameter a would be a city with several dense nuclei, one in flat terrain and another in a hilly area. A first study conducted in the flat area would provide the distribution of z. A second study in the hilly core would determine the probability distribution of a, assuming the previously derived z distribution. We are working at this time on a model that could determine directly both the z-distribution and the a-distribution.

It is interesting to note that the latter approach, if successfully developed, can be applied to any variable teamed with value of time (or distance disutility cost), provided that a scale is available for that variable. "Street attractiveness" or "environmental quality" in a flat area can be treated within this framework. The quality scale required could be based on the variations in walking speed. Hoel (11) has observed that the same pedestrian walks at varying speeds in the course of a trip, depending on the type of block he is walking along (shop windows, bank, factory, or parking lot).

CONCLUSIONS

This paper proposes a model of pedestrian behavior that can be tested in a real situation at a minimal cost, if coupled with a standard parking survey for data collection.

Most of the paper deals with how to determine the probability density distribution of the value of walking time, which is believed to be the central element for a pedestrian behavioral model applicable to a variety of situations (parking location selection, utilization of short distance people movers, or evaluation of pedestrian circulation improvements).

Valuable improvements to the model can be introduced, if necessary, by adequately designing the parking survey questionnaire.

The approach proposed allows for an incremental study design, concerned first with the value of time and then with other elements of the choice procedure such as street gradients or environmental quality.

Advantages of the model in its basic form are the low cost of data collection; its multiple-choice nature, covering a wider range of situations than binary-choice models; its "explanatory" orientation in that it proposes a rationale for pedestrian behavior instead of a more numerical correlation and remains meaningful at the level of an individual decision-maker; and its avoidance of the use of dollar costs that are ill-defined in the decision-maker's mind (such as car operating costs).

Among the model's shortcomings are the problem of time perceived versus actual time; the variability of walking speeds; the influence of the weather, which is unaccounted for; and the first-come, first-served rule of operation in the situation selected to calibrate the model that tends to distort the distribution of the value of time. Some of these drawbacks can be mitigated through investigation of particular interrelations such as that between walking speed and time perception bias, considered as permanent personal characteristics. Others are still beyond the range of analysis.

Evaluation criteria are proposed to test the model's reliability. This paper suggests that a first test be made in a medium-sized city that has a flat and rather uniform CBD. Subsequent tests can then be designed, if warranted, depending on the insufficiencies evidenced by the first one. This paper has tried to anticipate some of the problems and to suggest solutions.

REFERENCES

1. LeFevre, P. The Economic Justification of Road Construction. Road International, Summer 1956 and Autumn 1956.
2. Hummel, C. M. A Criterion Designed to Aid Expenditure Programming. HRB Bull. 249, 1960, pp. 52-60.
3. Haikalis, G., and Joseph, H. Economic Evaluation of Traffic Networks. HRB Bull. 306, 1961, pp. 39-63.
4. Warner, S. Stochastic Choice of Mode in Urban Travel: A Study in Binary Choice. University Microfilm, Inc., Ann Arbor, Mich., 1962.
5. Lisco, T. E. The Value of Commuters' Travel Time: A Study in Urban Transportation. Univ. of Chicago, Dept. of Economics, PhD dissertation, June 1967.
6. Quarmby, D. A. Choice of Travel Mode for the Journey to Work: Some Findings. Jour. Transport Economics and Policy, Vol. 1, No. 3, Sept. 1967.
7. Thomas, T. C. The Value of Time for Passenger Cars: An Experimental Study of Commuters' Values. Stanford Research Institute, Palo Alto, 1967.
8. Pratt, R. H. A Utilitarian Theory of Travel Mode Choice. Highway Research Record 322, 1970, pp. 40-53.
9. St. Clair, G. P., and Lieder, N. Evaluation of Unit Cost of Time and Strain-and-Discomfort Cost of Non-Uniform Driving. HRB Spec. Rept. 56, 1959, pp. 116-129.
10. Anderson and Bancroft. Statistical Theory in Research. McGraw-Hill, 1952.
11. Hoel, L. A. Pedestrian Travel Rates in Central Business Districts. Traffic Engineering, Jan. 1968.
12. Future Highways and Urban Growth. Wilbur Smith and Associates, 1961.
13. Oort, D. J. The Evaluation of Traveling Time. Jour. Transport Economics and Policy, Vol. 3, No. 3, Sept. 1969.

APPENDIX

DERIVATION OF DISTRIBUTION OF VALUE OF TIME ASSUMED STATISTICALLY INDEPENDENT OF WALKING SPEED

Let

v = walking speed,
t = walking time,
δ = walking distance,
x = value of time,
z = distance disutility cost,
$p(z)$ = probability density function of z among the population,
$q(v)$ = probability density function of v among the population, and
$r(x)$ = probability density function of x among the population.

$p(z)$ is assumed known through the histogram method, $q(v)$ is known from prior studies, and $r(x)$ is to be determined.

We have, by definition,

$$\delta = vt \tag{10}$$

Also, the overall cost of walking must be the same, whether computed on the basis of time or on the basis of distance.

$$xt = z\delta \tag{11}$$

Equations 10 and 11 imply that

$$x = zv \text{ for } x,\ z,\ v \geq 0 \tag{12}$$

x is the product of 2 independent stochastic variables, and its distribution is

$$r(x) = \int_{z_1}^{z_2} p(z) \cdot q(x/z) \cdot dz \tag{13}$$

z_1 and z_2, the limits of integration, are functions of the boundaries of the ranges permitted for z and v, which in turn define the range permitted for x.

If p(z) and q(v) are step functions, Eq. 13 is changed into Eq. 14.

$$r(x) = \sum_{i=1}^{N} p[(z_i + z_i+1)/2] \cdot q[2x/(z_i + z_i+1)] \ (z_{i+1} - z_i) \tag{14}$$

where z_1, z_2, . . ., z_{N+1} are the abcissas at which p(z) jumps from one step to the next. Equation 14 is particularly relevant to the determination of r(x) by approximation, when p and q are only empirically defined, and therefore not amenable to theoretical calculus.

DEFINITION OF STABILITY OF THE RELATION BETWEEN WALKING SPEED AND DISTRIBUTION OF DISTANCE DISUTILITY COST

As mentioned earlier, there is the possibility that fast walkers are more walk-loving than average and thereby have a low distance disutility cost (or, better said, show a distance disutility cost distribution shifted toward the low values). Conversely, slow walkers would have a distance disutility cost distribution shifted toward the high values. This correlation can be formulated as follows:

$$\bar{v} = f(z) \tag{15}$$

$$\sigma_v = g(z) \tag{16}$$

where

$\bar{v}$ = mean value of the average walking speed v, and
σ_v = standard deviation of v

both for a subpopulation of given distance disutility cost z.

Equations 15 and 16 are (at least theoretically) sufficient to define the distribution of v, shown by Hoel (11) to be normal.

It is hypothesized that the correlation between v and z is stable. One definition of stability is to assume that the functions f and g are identical from city to city. But this would mean in turn that people with a given walking speed have the same average disutility cost in city A and city B. This is not obvious.

Although, it seems reasonable that people's physical characteristics (exemplified here by "walking speed") are distributed in approximately the same way in various cities, differences in social and economic conditions may influence distance or time valuations by those physically similar people located in different cities.

For instance, if wages, prices, or dividends were doubled overnight while people's preferences stayed unchanged, their distance disutility cost would also have doubled, although their walking characteristics would still be the same as before. Equations 15 and 16 would then read

$$\bar{v} = f(z') \tag{17}$$

$$\sigma_v = g(z') \tag{18}$$

where

$z' =$ new distance disutility cost = twice old cost = $2z$.

However, for everyone, disutility cost relative to the population's mean would be undisturbed by the overnight change.

$$z'/\text{mean } z' = 2z/\text{mean } (2z) = 2z/2 \text{ mean } z = z/\text{mean } z \tag{19}$$

Consequently, one way of expressing the stability of the correlation between v and z, regardless of (intercity) differences in socioeconomic conditions, is to replace Eqs. 15 and 16 with Eqs. 20 and 21, valid in both the before and after situations of the example given above.

$$\bar{v} = F(Y) \tag{20}$$

$$\sigma_v = G(Y) \tag{21}$$

where

$Y = z/\bar{z}$, and

$\bar{z} =$ mean value of z among the city population.

Taking the functions F and G to be the same in different cities is now a relatively safe assumption.

A different but somewhat similar rationale could lead to $\bar{v}$ and σ_v being functions of $Y = (z - \bar{z})/\sigma_z$; these functions, like P and G, would then be considered valid for all cities.

IMPACT OF PARKING DURATION ON TIME AND COST TRADE-OFFS

The total cost C_i to a driver selecting parking location i is

$$C_i = \theta k_i + x t_i \tag{22}$$

where

$k_i =$ hourly rate of facility i,

$t_i =$ walking time from facility i to on-foot destination,

$\theta =$ parking duration, and

$x =$ value of time.

The cost differential between 2 facilities for the given driver is

$$\Delta C = C_1 - C_2 = \theta \cdot \Delta k + x \cdot \Delta t \tag{23}$$

where

$\Delta k = k_1 - k_2$, and

$\Delta t = t_1 - t_2$.

Assume that facility 1, which is optimal for duration θ_o, is competing with facility 2, which is less expensive ($\Delta k > 0$) but more distant ($\Delta x < 0$). For θ_o, ΔC is negative; i.e.,

$$\theta_o \cdot \Delta k + x \cdot \Delta t < 0 \tag{24}$$

But ΔC increases when θ increases, and when

$$\theta > -(x \, \Delta t/\Delta k) \tag{25}$$

ΔC becomes positive and facility 2 is the preferred one.

This shows that cheaper, more distant facilities become preferable when parking duration increases, if a constant value of walking time x is used. Of course, the marginal value of parking time k_i tends to decrease. This is well in accordance with observations.

DERIVATION OF DISTRIBUTION OF VALUE OF TIME, ASSUMING THAT THERE IS A CORRELATION BETWEEN WALKING SPEED AND DISTANCE DISUTILITY COST AND A PERSONALIZED SYSTEMATIC PERCEPTION BIAS

Let

v = walking speed,
T = actual walking time,
t = perceived walking time,
D = actual walking distance,
δ = perceived walking distance,
x = perceived value of time,
z = perceived distance disutility cost,
Z = actual distance disutility cost,
b = perception bias factor,
$p(Z)$ = probability density function of Z among the total population,
$r(x)$ = probability density function of x among the total population,
$q(v/Z)$ = probability density function of v among the subpopulation with distance disutility cost Z, and
$s(b)$ = probability density function of b among the total population.

The perception bias b is assumed to be statistically independent of all other personal characteristics z, v, x. The correlation between z and v is expressed through q(v/Z), as developed earlier. The distribution p(Z) is known from model application (histogram). The distribution s(b) is known from answers to a special question in the parking survey or prior studies on perceived time. Basic definitional relationships are

$$t = b \cdot T \tag{26}$$
$$\delta = b \cdot D \tag{27}$$
$$D = V \cdot T \tag{28}$$
$$xt = zd = ZD \tag{29}$$

The model described in this paper enables us to construct the histogram of Z. We now want to relate x (value of perceived time) with Z (actual distance disutility cost). Equations 26 through 29 lead to

$$x = (Z \cdot v)/b \tag{30}$$

Therefore, the probability density distributions are related as follows:

$$r(x) = \int_{Z_1}^{Z_2} \int_{V_1}^{V_2} p(Z) \cdot q(v/Z) \cdot s(Zv/x) \cdot dZ \cdot dv \tag{31}$$

As before, bounds can be imposed on v, Z, and b that will restrict the range of x. For each (permitted) value of x, the limits of integration are functions of the bounds imposed on v, Z, and b.

As shown earlier Eq. 31 can be transformed to fit the case of step functions, particularly relevant to the empirical determination of r(x).

TRAVEL TO OUTDOOR RECREATION AREAS IN KENTUCKY

John A. Deacon, University of Kentucky; and
Jerry G. Pigman and Robert C. Deen, Kentucky Department of Highways

ABRIDGMENT

•THE PURPOSE of this investigation was to examine the nature of travel to outdoor recreation areas in Kentucky. Basic data were obtained by means of a license-plate origin-destination survey at 160 sites within 42 major recreation areas in Kentucky. These data were supplemented by means of a continuous counting program at 10 sites. The O-D survey was conducted on Sundays during the summer of 1970; outdoor recreation travel and other rural travel typically reach a combined peak on summer Sunday afternoons. Overall results indicate that the license-plate O-D survey is a most satisfactory way to gather data of the type required, particularly because it enables maximum utilization of personnel, does not require voluntary participation of the traveler, and allows a large sampling rate. The time selected for the survey, 10 a.m. to 8 p.m. on Sundays, proved to be completely acceptable.

Modeling efforts concentrated on the simulation of distributed travel flows from each of 190 origin zones comprising the entire continental United States to each of the 42 recreation areas. The actual distributed flows, treated as the dependent variable in the analysis, were the 10-hour departing vehicular flows on an average summer Sunday Factors were developed to convert the 10-hour flows to peak-hour and 24-hour flows. Distributed traffic flows are known to be sensitive to demand for recreation at the origin zone, supply of recreation opportunities at the recreation area, price of the recreation experience, competition among available sources of supply and demand, and various other factors. Primary independent variables that were chosen for analysis and that reflect to a sufficiently accurate first approximation the most important of these sensitivities are population of the origin zone in thousands, p_i, attractiveness of the recreation area, a_j, and spatial separation in miles between the origin zone and the recreation area, d_{ij}, which is used as a measure of the price of the recreation experience.

The attractiveness of recreation areas of varying types and sizes can be reasonably approximated by the number and types of facilities available. The following facilities, listed in the order of highest to lowest significance, were identified as having important effects on attractiveness and were judged to be essential for encompassing the wide range of recreation areas studied: water area, picnic tables, swimming pools, horseback trails, beach, golf, hiking trails, overnight accommodations, and outdoor drama. The relative importance of those facility types was evaluated by using regression analyses.

Others who have attempted to simulate distributed recreation travel flows have utilized gravity models, opportunity models, system theory models, and single-equation models. Because the available literature revealed no distinct preference for any model type, efforts of this study were concentrated on various single-equation models evaluated by regression techniques and on a cross-classification model. The cross-classification model was found to be an acceptable means for simulating and predicting outdoor recreation travel flows and was decidedly superior to any of the single-equation models evaluated. From the cross-classification model, per capita distributed flows were found to decrease at a decreasing rate with increasing population, increase at an increasing rate with increasing attractiveness, and decrease at a decreasing rate with increasing distance.

The best single-equation model for simulating distributed flows, V_{ij}, was of the form

$$V_{ij} = k_1 d_{ij}^{k_2} p_i^{k_3} a_j^{k_4}$$

in which the k's are constants. This nonlinear flow equation as well as all others investigated had to be evaluated by using nonlinear regression analysis. Linear regression analysis using transformed (linearized) equations proved totally unsuitable.

Other data, in addition to the distributed flows, were also available from the O-D survey. The average vehicle occupancy rate was found to be 3.13 persons/vehicle. However, rates at specific locations were found to depend on the length of the trip and the nature of the recreation area and were smallest for the short trips to predominantly day-use facilities. Vehicle classification (percentages of the various vehicle types) was also found to depend on both trip length and the nature of the recreation area.

DEVELOPMENT AND TESTING OF A MULTIPATH-ASSIGNMENT TECHNIQUE

P. M. Dalton and M. D. Harmelink, Ontario Department of Transportation
and Communications, Downsview

ABRIDGMENT

•IN MOST conventional traffic-assignment techniques, the travel between 2 points is assigned to the minimum or shortest path between the 2 points. Even in capacity-restraint assignment, in which a capacity function or speed-flow relation is used in successive assignment runs to modify link speeds for use in the assignment, minimum-path assignment is used in each run. A distribution of trips over several paths is usually obtained by combining the results of successive assignment runs.

A multipath-assignment technique has several advantages over the conventional minimum-path technique. First, in networks without capacity limitations (unrestrained or demand assignments), empirical studies have shown that travel between 2 points usually distributes itself over several routes. Further, minor network changes can often produce major changes in assigned volumes when minimum-path assignment is used. Second, in capacity-restrained assignments, the multipath technique enables each assignment to be considered on its own, so that (a) all the volume on a given link is assigned at the same speed, and the consequent link volumes are consistent with the speed-flow relation for the links, and (b) extreme speed oscillations and unrealistic paths produced by them are largely excluded.

In the method developed and tested, the algorithms used for building the multiple-route trees are the same as those used in most conventional minimum-path assignment programs, except that a different set of link costs is used to build each tree. Each link time is chosen at random from a distribution of 8 values having a mean value equal to the specified link cost and a mean deviation specified by the user according to link type and cost. When a large number of trees are built, the paths will be divided among the feasible routes with the largest number normally on the minimum path and the numbers on alternative paths decreasing as the extra cost involved in using them increases. This principle also applies between any pair of nodes along a route, thus making trivial alternatives unlikely where the extra cost is high in proportion to the distance covered between the points where the alternatives exist. This overcomes a problem with some previous multipath techniques that assign to the n shortest routes, which may only be minor deviations of the same basic route.

In its simplest form, the assignment may consist of building just 1 tree for each origin zone. Although this permits only 1 path between a specific pair of zones, there may still be multiple choices of routes between pairs of nodes because large numbers of different zone-to-zone movements may pass through the same nodes. Building 1 tree per zone may create problems near the origin zone if the volumes assigned because of that zone are a major part of the total link volume. This problem was solved, without the computer cost of building several complete trees, by introducing a cost cordon around each origin zone and by building several "inner trees" inside the cordon for each complete "outer tree" outside the cordon.

Ideally, the probability of a given path between 2 points being chosen should be independent of the number of links constituting it, and this is achieved by making the mean deviation of the link times within each link class proportional to the square root of the specified link times.

The method of capacity restraint adopted is one of successive assignments and speed adjustments, the object of which is to reach an equilibrium point where the volumes assigned to each link in the network are consistent with both the speeds at which they were assigned and the link capacities. Speed-flow relations are specified by link class and a capacity indicator. After each assignment, the link speeds are adjusted according to the assigned volume-capacity ratio and the speed-flow relations. The program includes several options as to the formula used to derive the new link speed.

The speed adjustment procedure used for the capacity restraint may be 1 of 2 methods. The first of these is an iterative procedure where the complete trip table is reassigned at each stage and all previous assignments are ignored. This method has always produced poor results when used with a minimum-path assignment technique because it may produce large oscillations in assigned volumes for very small speed changes on some links while a similar change on other links has no effect. Multipath techniques should produce better results because they permit trips to be diverted from 1 route to another in small increments.

The second method of capacity restraint is an incremental one where a proportion of the trip table, specified by the user, is assigned at each stage and the speed adjustment is based on the total assigned volume in all the previous increments and the link speeds at which the latest assignment was made. The speed adjustments may be made (a) according to the total capacity of the link or (b) according to a proportion of it equal to the proportion of the trip table already assigned (effectively making all speed adjustments on the basis of a fully loaded network). (Some planners have called this latter technique an iterative method because of the repeated adjustments to the link speeds. In this paper, it is called the incremental method and is distinguished from the iterative method in which the assignment process is applied more than once, but each assignment run is considered complete on its own.) The latter technique is more dynamic and responsive to volume buildups at an early stage in the assignment process and is therefore preferred to the former.

The testing and evaluation of the program were carried out in 2 stages. The first series of tests consisted of unrestrained assignments on 3 Ontario networks of different types for comparison with minimum-path assignments. These were a small urban network, a rural area network, and a comprehensive regional network from the 1964 Toronto Area and Region Model Study (TARMS).

In each case, the first assignment was made on the final networks that were used for the base years of the respective studies and were assumed to have been calibrated for a reasonable minimum-path assignment. On all 3 networks, the link volumes obtained by the multipath were closer to observed volumes than those given by the minimum-path assignment with a very substantial improvement on the TARMS network. Further improvements were obtained with a better calibration for the multipath.

The second series of tests was designed to evaluate different combinations of capacity-restraint and assignment techniques. The network used for this was the metropolitan area of the 1969 TARMS network. The methods tested were as follows:

1. Minimum-path iterative,
2. Multipath iterative,
3. Minimum-path incremental,
4. Multipath incremental, and
5. Multipath iterative followed by multipath incremental.

Early tests eliminated method 1 as being unworkable. In methods 2, 3, and 4 it was found that the iterative method produced link speeds closer to those observed and much higher on the average than the incremental methods; nor were the extreme values of link speeds so pronounced. These differences can be explained by the manner in which the 2 techniques work and were to be expected.

The greater differences in link speeds from run to run in the incremental assignment caused some large detours from the minimum path, raising the total vehicle mileage by a considerable amount at each stage. The iterative method was found to give a stable average speed very quickly, and the total vehicle-miles assigned also changed very little between speed adjustments.

The comparison of the assigned volumes with observed volumes for 1,378 links where counts were available was the principal basis of evaluating the methods. All of the methods produced roughly the same error when compared to the observed volume and, because there are many sources of this error other than the assignment technique (network representation, trip table, capacity functions, and errors in the counts), the only conclusion that could be drawn was that no method was significantly better than any other for producing accurate assigned volumes. Method 4 gave marginally better results than methods 2 or 3. Method 5 was tested because preliminary tests had indicated that better results could be obtained at the initial stages by using a cost function of time and distance instead of just time for the tree building. This was found to be true at the early stages, but the benefits diminished rapidly in subsequent stages.

With all the incremental methods it was found that, although the assigned volumes came closer to the speed-flow relations as the capacity restraint proceeded, the comparison with observed link volumes started to deteriorate after 3 or 4 speed adjustments. The fact that the assigned volume appears to be a closer fit to the speed-flow relation is artificial because the assigned volume has been obtained at several different link speeds that cannot be represented by a single speed-flow point. The value of making a large number of speed adjustments for an incremental assignment in order to obtain settlement would seem questionable.

Some of the conclusions indicated by the results were that the multipath-assignment technique can be used to advantage for most types of network with or without capacity-restraint assignment. When the multipath technique for capacity restraint is used, an iterative method should be adopted to produce realistic link speeds; otherwise, the method of capacity restraint used has little effect. With either method of capacity restraint, there is likely to be little benefit gained by making more than 4 or 5 speed adjustments.

The multipath assignment and capacity restraint techniques are contained in 1 program but may be used independently if required. The program is written in FORTRAN IV for use with an IBM 360/65 computer and will accept networks with as many as 2,000 zones, 6,000 nodes, and 14,000 links.

USE OF SOCIOECONOMIC INDICATORS IN TRIP ATTRACTION OF LARGE WORK CENTERS

George E. Mouchahoir, Urban Transportation Project, Atlanta University; and
Paul H. Wright, School of Civil Engineering, Georgia Institute of Technology

•IN trip-generation analysis, transportation planning has been concerned with the establishment of a functional relation between trip-end volumes and the land use and socioeconomic characteristics of the units from which they originate or to which they are destined (1). Different land use and socioeconomic characteristics, or indicators, have been used in different studies, and in some cases the models obtained were unreliable because of the interdependence of these characteristics (2). On the other hand, trip-generation analysis has been determining the functional relations based on relatively large geographic units (census tracts or traffic zones). These large geographic units led to unreliable results in small-area studies such as a central business district transportation study (3, 4, 5).

This paper attempts to analyze quantitatively the interrelations among different land use, travel, and socioeconomic indicators and to generate a trip-attraction model for large work centers instead of traffic zones.

METHOD OF STUDY

Because of the absence of socioeconomic and travel indicators by work centers, it was necessary to undertake a survey for the purpose of this study. Twenty work centers, having a large number of employees, were selected from the Atlanta metropolitan area. These employees numbering about 25,000 were surveyed by mail during the winter of 1970. The following information was obtained: age, number of children, occupation or profession, education level, number of years at work, home value or rent, lot size, distance from home to work, travel time and distance, personal and family incomes, and car ownership. Similarly, the following was obtained from the employers: number of trip attractions, floor areas, distance from the central business district, and the assessed value of the work centers.

Before employee and employer variables were coded for statistical analysis, it was imperative to quantify all variables in a scalable manner (6). All collected variables were quantified and coded easily with the exception of occupation or profession and education. Education was quantified by using the number of years the person spent acquiring an education. Occupation or profession was quantified by using the North-Hatt occupational scaling method (7).

The method of analysis used in this investigation was predicated by the nature of the data collected and the objectives of the study. The collected data were statistical in nature and are all random variables. This randomness associated with more than 1 variable, plus the interdependency of the variates brought the problem into the realm of multivariate statistics (8).

Two subfields of multivariate statistical analysis were chosen to achieve the objectives of this investigation: factor analysis and component analysis (9). Factor analysis was used to group the observed variables together in ways that permit one to synthesize new entities called factors, or indicators, and to determine the degree of association among these variables. These factors are independent one from the other; i.e., they are orthogonal vectors and they occur in descending order as far as variances are concerned. That is to say, the first factor explains the largest portion of the total variance of the original variables.

Component analysis was used to generate a multivariate statistical model relating the number of trips attracted to work centers to the socioeconomic and travel factors determined from the factor analysis. This analysis starts by determining statistically independent groupings called components. These components are tested for their significance by using multivariate statistical tests, and only the significant components were used in the regression operation to generate the multivariate model.

IDENTIFICATION OF EMPLOYEE AND EMPLOYER INDICATORS

This analysis consisted of identifying the socioeconomic and travel indicators pertinent to the employee and to the employer.

Identification of Employee Indicators

The employee socioeconomic and travel indicators were identified after the employee-connected variables were factor-analyzed. This analysis was performed in 2 steps. The first step analyzed the interrelation among the employee-related variables of all the work centers. This analysis indicated that the 13 variables collapsed into the following 6 factors: 2 socioeconomic indicators having the variables occupation, education, personal and family incomes, age, and years at work strongly interrelated among themselves; 1 travel indicator having the variables travel time and distance strongly associated with it; 1 home indicator having the variables home value, rent, and lot size strongly associated with it; 1 car-ownership indicator having the number of cars in a family related to it; and 1 family size indicator having the number of children related to it (10). The second step was to choose 1 variable from each factor given above, to determine the averages of these variables by work center, and then to factor-analyze them. The variables chosen were number of children, occupation level, home value, number of cars, family income, and distance of travel. Their choice was made depending on their degree of association with their corresponding factors. The factor analysis of the averages of the 6 employee variables of the 20 work centers surveyed showed that these variables were grouped into 4 independent factors. The first factor indicates that the averages of the variables occupation, home value, and family income were strong interrelated under a single factor. The second factor is a single variable factor having the average number of children strongly associated with it. Similarly, the third factor is also a 1-variable factor having the average distance of travel strongly related to it. Conversely, the fourth factor indicates that the averages of the 2 variables, number of cars and family income, are highly interrelated under it. These 4 independent factors show that these employee average variables could be represented by the 4 factors, in other words, by choosing 1 strongly associated variable from each factor. The chosen average variables were number of children, occupation, number of cars, and distance of travel.

Identification of Employer Indicators

The employer indicators were identified by using factor analysis on the 4 collected variables: floor space, distance from the Atlanta central business district, assessed value of work center, and number of work trip attractions. This analysis grouped these variables into 3 independent factors. The first indicates that the number of work trip attractions variable is strongly associated with the floor space variable. The second factor is a single-variable factor having the distance from the central business district variable. Conversely, the third factor indicates that the 2 variables, number of work trip attractions and the assessed value, are interrelated and are strongly associated with this factor.

Because the objective is to generate a relation of the number of work trip attractions variables to the significantly related employer and employee variables, it is imperative to retain the 2 variables, floor space and assessed value, and to reject the distance from central business district variable from later consideration.

RELATIONS BETWEEN EMPLOYEE AND EMPLOYER INDICATORS

When the 2 sets of indicators and the significant variables of the employer and employee had been determined, a factor analysis was performed to determine the interrelations between these 2 sets. The 7 variables analyzed were number of children, occupation level, number of cars in family, distance of travel, floor space, assessed value, and number of work trip attractions. This analysis resulted in the collapse of the 7 variables into 5 independent factors. The first factor indicates that the variables floor space, assessed value, and number of work trip attractions are strongly interrelated under this factor. The second factor shows that the variables occupation, assessed value, and, to a lesser extent, work trip attractions are associated under a single independent factor. The third factor is a single-variable factor and has the number of children variable strongly associated with it. Conversely, the fourth factor indicates that the variables distance of travel, and, to a lesser extent, the number of work trip attractions are strongly associated with it. The fifth factor is a single-variable factor and has the number of cars associated with it. Therefore, the number of work trip attractions variable appears to be associated with factors 1, 2, and 4. So, choosing 1 significant variable from each of these independent factors will determine the independent variables to be used in the component regression model relating the number of work trips to the employee and employer characteristics. The independent variables chosen to relate to the number of work trip attractions variable are floor space, distance of travel, and occupation level.

TRIP ATTRACTIONS MODEL

The work trip attraction relation with the significant employee and employer variables was determined by using component analysis multivariate statistics. This component analysis starts by a principal component analysis on the independent variables floor space, distance of travel, and occupation. This principal component analysis resulted in the grouping of these variables into 3 components. The first component is strongly expressed by the variables occupation and distance of travel, and the second one is expressed by the floor space variable. Similarly, the third component is expressed by the variables occupation and distance of travel.

This analysis indicates that the variance explained by the first component contains the largest amount of variance and the variance explained by the third one contains the least amount. The significance of the amount of variance explained by these components is determined by using the Bartlett test of significance on each residual. This test indicates that, when the third component is alone as a residual, the test is not significant at 0.1 percent level. Conversely, when the first 2 components are tested at the same level, the test is significant.

Having determined the significant components, the component analysis proceeds by using regression analysis on these orthogonal components. This component regression analysis generates the following multivariate statistical model:

$$y = -240.37\, x_1 + 163.12\, x_2 + 2.10\, x_3 - 515.48$$

where

x_1 = average occupation level,
x_2 = average distance of travel between home and place of work,
x_3 = floor space, and
y = number of trips attracted to work centers.

The model can be expressed in its standardized coefficients form as

$$y = -0.142\, x_1 + 0.338\, x_2 + 0.530\, x_3$$

where the terms are defined as before. These standardized coefficients correspond to the variables expressed with 0 means and unit variances. These coefficients show that the variable floor space contributes the most to the model and the variable occupation level contributes the least to the model.

The F-ratio test of significance performed on this model indicates that the model is significant at the 0.001 percent level. Conversely, its coefficient of multiple determination is 0.378 implying that only about 38 percent of the variation in the number of trip attractions to the work centers is explained by the model. Also, the efficiency of the model is about 14 percent, which is relatively low for predictive uses. However, it is worth noting that the structure of this model expresses a rational relation among the variables involved. The model confirms previous findings on the strong relation between the number of work trip attractions and the floor space variables (3). The model also shows that the variables average occupation level and average distance of travel affect the number of work trip attractions to the work centers. The model implies that the work centers that have a great number of work trip attractions are the ones that employ a large number of blue-collar workers. Conversely, it suggests that the large work centers tend to attract workers from a great distance from the center in order to satisfy their large demand of skills.

ACKNOWLEDGMENTS

This research was supported by the Urban Mass Transportation Administration under the Research and Training Grant to the Georgia Institute of Technology.

REFERENCES

1. Guidelines for Trip Generation Analysis. U. S. Govt. Printing Office, Washington, D. C., June 1967, p. 7.
2. Walker, J. R. Social Status of Head of Household and Trip Generation From Home. Highway Research Record 114, 1966, pp. 141-151.
3. Wright, P. H. Relationships of Traffic and Floor Space Use in Central Business District. Highway Research Record 114, 1966, pp. 152-168.
4. Manning, J. M. Urban Travel Data in Rhode Island 1961-1971. In Urban and Regional Information Systems: Past, Present and Future, Eighth Annual Conf. of Urban and Regional Information Systems Assn., 1970, p. 136.
5. Black, A. Comparison of Three Parameters of Nonresidential Trip Generation. Highway Research Record 114, 1966, pp. 1-7.
6. Kendall, M. G. The Advanced Theory of Statistics, 2nd Ed. Charles Griffin and Co., London, Vol. 1, 1945, p. 388.
7. Reiss, A. J., et al. Occupations and Social Status. Free Press, New York, 1961.
8. Mouchahoir, G. E. The Relationsip of Work Trips to Employment Connected Social and Economic Factors. Georgia Institute of Technology, Georgia Institute of Technology, PhD thesis, 1970.
9. Kendall, M. G. A Course in Multivariate Analysis. Charles Griffin and Co., London, Griffin's Statistical Monographs and Courses No. 2, 1965, p. 6.
10. Mouchahoir, G. E. Economy in Data Collection for Planning Studies. Traffic Quarterly.

ACCESS TO JOBS AND WILLINGNESS TO TRAVEL

Charles B. Notess, Civil Engineering Department,
State University of New York at Buffalo

ABRIDGMENT

•DURING the past few years, a number of studies have examined differences in access to jobs from different residential areas in the city and by different modes of travel (1, 2, 3, 4, 5, 6, 7, 8, 9). Generally these studies compared travel time by car and travel time by public transit to show that the worker who had to travel by bus, on the average, traveled as much as 3 times longer door-to-door as the worker who traveled by automobile. By so doing, he had access to almost the same number of jobs as his neighbor with a car had in a city like Buffalo, New York (12). To some extent, the bus rider is excluded from a significant proportion of jobs that are not served by bus lines.

Studies that examined the performance of bus experiments to improve access to suburban jobs, in general, concluded that conventional bus service, as a means to transport inner-city workers to suburban jobs, was not a reasonable solution. The average cost per rider was too high, suburban job locations were too dispersed to be served efficiently by public transit, and the off-peak demand was very small (7, 8).

Important factors, not examined in these studies, are the worker's perceptions of prospects for stable career development in the suburban job and how perceptions of commuting by bus fit into the life style and self image of the worker. A framework for analyzing these factors is presented in another report (10). Published studies by Wachs and Schofer (9) and by Gustafson et al. (11) have examined attitudinal responses to transit characteristics but have not gone beyond to examine aggregate measures of willingness to travel.

This paper summarizes a study in which an aggregate measure of the willingness to travel to work was developed by using, as a basis, the friction factor incorporated in the gravity model for trip distribution. The willingness of automobile users and bus riders were compared for 4 different occupational groups.

DERIVATION OF A WILLINGNESS MEASURE

An ideal measure of willingness should be independent of the particular spatial distribution of jobs and should be independent of the attractiveness of particular jobs. It should be a function of travel time (door-to-door) and mode of travel only. The concept of friction factor as used in the gravity model for trip distribution is ideally suited as a basis for deriving a willingness measure.

The gravity model can be viewed as an equation for estimating the travel activity patterns of persons where trips originate in a zone i and are destined to a zone j; the number of person trips from i to j is T_{ij}. The equation is written in the following form:

$$T_{ij} = P_i \, (A_j F_{ij}) \Big/ \left(\sum_{j=1}^{n} A_j F_{ij} \right) \tag{1}$$

where P_i is the total number of trips produced in zone i for a particular trip purpose, say, work or shopping trips, and A_j is the total number of trips attracted to zone j for that same purpose. The term F_{ij} is commonly called a friction factor and is defined subsequently.

The friction factor F_{ij} decreases with increasing travel time between zones i and j. For a pair of zones, i-j, F_{ij} may be interpreted as a measure of the decreases in attractiveness of a zone or of the willingness to travel there as the travel time to that zone increases. In other words, the factor F_{ij} can be interpreted as a measure of the strength of the willingness to travel as a function of travel time.

Analyses that used the following equation to represent the variation of F_{ij} with travel time are presented in another report (13):

$$F_{ij} = [\alpha\beta/\gamma(\alpha)]\, t^{\alpha-1.0} e^{-\beta t} \tag{2}$$

This equation is the so-called gamma distribution with $\gamma(\alpha)$ equal to $(\alpha - 1)$. The 2-parameter equation provides an efficient way to describe the friction factor with a shape parameter α and a scale parameter β.

CALIBRATION OF GRAVITY EQUATION

The gravity equation, Eq. 1, was programmed for calibration on the CDC 6400 computer at the State University of New York at Buffalo. Origin-destination data from a 1968 survey of work trips in the black ghetto of Buffalo were used, and the particular computer program was developed and converged within 4 iterations to obtain averaged F_{ij} values that replicated averages of the observed T_{ij} values to within 2 percent. Three hundred destination zones were used for trips from each of 7 origin zones. The averaged F_{ij} and T_{ij} values represented time averages of all F_{ij} or T_{ij} within 5- or 10-min time bands. The 10-min time bands were used for bus trips, and the 5-min bands were used for automobile trips (12).

The most important conclusions obtained from examining the F_{ij} curves were as follows: (a) The friction factor curves for bus trips were more erratic than those for automobile trips and did not follow the analytical equation as closely as did the automobile curves; (b) much of the erratic nature in the bus curves was due to the small sample size (fewer than 10 trips for a particular origin zone); and (c) with the aid of families of curves for different values of the parameters α and β, the values $\alpha = 1.2$ and $\beta = 0.08$ were selected as providing a good fit to the friction factor for automobile commuters, whereas $\alpha = 1.7$ and $\beta = 0.04$ were obtained for bus riders (4, 12).

UNITY MEASURE OF WILLINGNESS

The sampling variability was reduced by having 1 numerical measure of friction rather than several points on a curve. Therefore, the friction factor curves were reduced to 1 numerical measure by forming a ratio of average travel time to work by workers residing in a zone divided by the ideal travel time to work if there were no attenuating effect for long trips. This ratio was called the friction index.

In mathematical terms, the friction index was derived as follows. The average of travel time for observed trips from zone i is

$$\overline{t_i} = \left(\sum_{j=1}^{j=n} T_{ij} \times t_{ij}\right) \Bigg/ \left(\sum_{j=1}^{j=n} T_{ij}\right) \tag{3}$$

where

T_{ij} = observed sample of work trips from zone i to zone j,
t_{ij} = travel time from zone i to zone j,
$\overline{t_i}$ = average travel time of workers from zone i,
A_j = total number of workers employed in zone j, and
n = total number of work zones considered.

$$\overline{\text{Ideal } t_i} = \left(\sum_{j=1}^{j=n} A_j \times t_{ij}\right) \Bigg/ \left(\sum_{j=1}^{j=n} A_j\right) \tag{4}$$

$$\text{Friction index} = F_i/\text{ideal } t_i \tag{5}$$

Table 1. Comparison of friction indexes for occupational groups.

Mode	Occupational Group	Friction Indexes Using Total A_j	Friction Indexes Manufacturing A_j	Friction Indexes Nonmanufacturing A_j
Automobile drivers and riders	Male operatives	0.837	0.816	
	Male laborers	0.822	0.802	
	Male others	0.695		0.646
	Female workers	0.453		0.423
Bus riders	Male operatives	0.662	0.646	
	Male laborers	0.743	0.725	
	Male others	0.635		0.592
	Female workers	0.587		0.548
Ratio of automobile to bus indexes	Male operatives	1.26	0.646	
	Male laborers	1.093	0.725	
	Male others	1.093		0.592
	Female workers	0.772		0.548

Table 2. Comparison of friction indexes.

Ratio of Friction Indexes[a]	Automobile Users	Bus Riders
Male operatives and male laborers	0.993	0.87
Male operatives and male others	1.175	1.019
Male laborers and male others	1.182	1.08
Male others and female workers	—	1.08

[a] Male operatives friction index based on manufacturing A_j. Male laborer, male others, and female worker friction indexes based on total A_j.

Typical values for the friction index are given in Table 1 for bus and automobile modes of travel for 4 occupational groups. The friction indexes were calculated for the job distribution A_j (all types of occupations grouped together), for a distribution of manufactoring jobs only, and for a distribution of nonmanufacturing jobs derived as the difference between the latter 2 distributions. (Information on the approximate number of manufacturing jobs in each zone was available from New York State Department of Transportation data.)

Ratios of the friction index for automobile to that for bus are given at the bottom of Table 1. Dividing the automobile value by the bus value removed (canceled out) the effects of the particular A_j distribution on the friction index, and thus it can be shown that the ratios of friction index can also be obtained from a ratio of average travel time by car to that by bus for workers in a particular occupation (12). The ratios indicated that male operatives with a ratio of 1.26 were less inhibited by the friction of space when traveling by car than when traveling by bus. Male laborers and other male workers did not have so different a response between car and bus users because, for these latter 2 groups, the ratios were close to unity, 1.10 and 1.093 respectively.

Let us assume, for lack of more detailed data, that the job distributions of bus riding and automobile riding operatives were best represented by the manufacturing A_j distribution and that male laborers and other males using both modes of travel were best represented by the total A_j distribution. Then the friction indexes are useful for comparing the willingness to travel for these occupational groups.

Ratios of these friction indexes are given in Table 2 and indicate that male operatives were less willing to overcome the friction of space than male laborers, for the ratios of friction indexes for these 2 occupations were less than 1.0. This conclusion was valid, however, only if the A_j distributions used, as given in Table 2, were appropriate. Unfortunately no better data were available.

Similarly, male operatives were more willing to travel than other males, automobile users more so than the bus riders, and male laborers more so than other males.

CONCLUSION

This brief summary shows that it is possible to develop useful measures of willingness to travel as a measure of travel time only if one has specific data on the spatial distribution of jobs for the particular occupational groups under consideration. These indexes effectively separate the attractiveness of particular jobs from the effect of travel time to the job.

REFERENCES

1. Meyer, J. R., Kain, J F., and Wohl, M. The Urban Transportation Problem. Harvard Univ. Press, Cambridge, 1965.
2. U. S. Riot Commission Report of the National Advisory Commission on Civil Disorders (also known as the Kerner Report). Bantam Books, New York, March 1968.
3. Kalachek, E. Ghetto Dwellers, Transportation and Employment. Paper presented at Conf. on Poverty and Transportation, Cambridge, Mass., Amer. Academy of Arts and Sciences, June 1968.
4. Notess, C. B., and Paaswell, R. E. The Mobility of Inner-City Residents. Dept. of Civ. Eng., State Univ. of New York at Buffalo, Prel. Rept., July 1969, and Final Rept., Dec. 1970.
5. Transportation Accessibility From the Model Cities Area—Nashville, Tenn. A. M. Voorhees and Associates, Inc., March 1969.
6. Breuer, R., and Shafer, J. Transit Evaluation: Demand and Non-Demand Aspects. New York State Department of Transportation, Publication TR080036, Oct. 1968.
7. Transportation-Employment Project, South Central and East Los Angeles, to Determine and Test the Relationship Between a Public Transportation System and Job and Other Opportunities of Low Income Groups. Calif. Business and Transportation Agency, Jan. 1970.
8. Pignataro, L. J., Falcocchio, J. C., and Roess, R. P. An Analysis of Selected Bus Demonstration Projects. Paper presented at ASCE National Transportation Eng. Meeting, Washington, D. C., July 1969.
9. Wachs, M., and Schofer, J. L. Evaluating the Transit System and Access to Jobs in Chicago. Paper presented at ASCE-ASME National Transportation Eng. Meeting, Seattle, July 1971.
10. Notess, C., and Wilder, E. Factors Behind Modal Choice in the Ghetto. Paper presented at ASCE National Transportation Eng. Meeting, Boston, July 1970.
11. Gustafson, R. L., Curd, H. N., and Golob, T. F. User Preferences for a Demand Responsive Transportation System. General Motors Res. Laboratories, Warren, Mich., Res. Publ. GMR-1047, Jan. 1971.
12. Notess, C. B., Access to Jobs and Journeys to Work From the Black Ghetto. State Univ. of New York at Buffalo, PhD dissertation, Feb. 1971.
13. Factors and Trends in Trip Length. NCHRP Rept. 48, 1968, 70 pp.

STABILITY OF RECREATIONAL DEMAND MODEL

D. C. Robinson and W. L. Grecco, Joint Highway Research Project, Purdue University

The principal objective of this study was to determine the stability of previously developed recreational demand models. Models were of the form $Y = A e^{-BX}$ and utilized easily obtainable and predictable variables. The study illustrated how the model can be used to predict future attendance and traffic volumes. Three parks were used in the study, and data were collected in interviews with 25 percent of arriving trips at the park entrances. Almost 12,000 interviews were conducted during the period from 1967 through 1969. The new reservoir model was developed by nonlinear regression analysis utilizing distance, population, and influence of other similar facilities. Two equations constituted the prediction model: one for when there is no other similar facility closer to a county than the reservoir under study, and one for when there is another such facility closer to the county than the reservoir under study. A comparison showed that, while parameter B remained fairly constant over time, there was an increase in parameter A.

•IN 1936, a national policy on flood control was established by the Congress of the United States. This policy provided that the federal government would cooperate with states and their political subdivisions on flood control projects; that flood problems would be tackled jointly by the U.S. Department of the Army and the U.S. Department of Agriculture; that project benefits must exceed project costs; and that projects recommended would not be constructed unless specifically authorized by law. Since 1936, more than 40 reservoirs have been constructed in the Ohio River Basin alone; 30 more are in the planning or construction stage.

Flood control, irrigation, and hydroelectric power were the 3 purposes originally considered in the benefit-cost analysis for justification of the construction of dams and their resulting reservoirs. There is, however, an added dividend of flood control dams that only in recent years has been recognized and included in the economic analysis. This dividend is the recreational lakes that are created by such dams.

Recreation is now recognized as an important business in this country. A substantial portion of the gross national product is derived from recreational pursuits in all areas of the nation.

The development of the future highway network must take into account the traffic-generating abilities of a recreational park or reservoir. A recreational facility is of little value unless it has adequate access. Recreational highways exhibit such unique traffic patterns that it is not enough simply to use techniques that have been found valid for the analysis of traffic flows on urban streets and nonrecreational rural highways. The multipurpose reservoirs are natural recreational attractions and consequently recreational traffic generators. It is essential for the full utilization of the recreational potential of a reservoir that transportation planning coincide with reservoir development plans. Little factual information is available at present that can be used by planners to estimate the recreational demand.

SCOPE

The area of water available for recreational purposes within the state of Indiana is in the process of being substantially increased under flood control programs of the U.S. Army Corps of Engineers. Six multipurpose reservoirs have been completed to date; 8 more are authorized and many more are planned (Fig. 1). The 6 completed reservoirs have added a total of more than 20,000 acres of water at summer levels. The Indiana State Department of Natural Resources is responsible for the development and operation of recreational facilities at such reservoirs.

Not until 1967 (1) was any information made available to the Indiana Department of Natural Resources for the planning of recreational facilities at reservoirs (1). The result of this initial research, in which Cagles Mill and Mansfield reservoirs were studied, was a model (referred to as the "previous model") for the prediction of recreational trips to new reservoir areas in Indiana. The model utilized road distances, county population, and influence of other similar facilities as the parameters affecting attendance. The technique developed and reported in the literature illustrates how the model can be used to predict future attendance and traffic volumes to recreational areas at multipurpose reservoirs (2). During the initial work, insufficient data were collected at Monroe Reservoir (which was then in the process of being developed) to be incorporated in the prediction model.

The rapid growth of recreational travel that is expected in the next decade requires that all demand models be under continual surveillance. The scope of this work was to check the stability of the previous model.

The recreation facilities at each of the reservoirs are similar in type; however, the amount of facilities varies among the 3 reservoirs. For example, Cagles Mill and Mansfield each have 1 beach several hundred feet in length; Monroe has 3 beaches (one of which is operated by the U. S. Forest Service). Boat-launching ramps are provided at various locations around each reservoir: 5 at Mansfield, 2 at Cagles Mill, and 9 at Monroe. Within the main recreational areas at each reservoir are located the campgrounds, beaches, concession stands, boat rentals, picnic areas, hiking trails, and bathhouses. In general, each park is well kept by personnel who know and take pride in their work.

Proper utilization of these facilities requires an adequate highway system, ranging from local access roads to state highways. The main objective of this research was to provide a simplified method for estimating future traffic volumes for new facilities of this type.

PROCEDURE

Data collection for this continuing study was carried out between June 1967 and August 1969 at each of the 3 parks. The primary source of data was a 25 percent interview survey of vehicular trips arriving at the parks. The 25 percent sample was chosen because it was considered adequate for analytical purposes and it did not create delays to arriving visitors. During the 3-year period, interviews were performed at the following locations within each park:

Park	Location
Mansfield	Main gate; dam and Hollandsburg boat ramps
Cagles Mill	Main gate; Cunot dock boat ramp
Monroe	Paynetown, Fairfax, and Hardin Ridge gatehouses; Cutright, Dam, and Moores Creek boat ramps

In the interviews, the driver was asked from which county the trip had originated and the purpose of the trip, the interviewer recorded the license number (the prefix of which, on Indiana passenger cars, is a code number relating to the county in which the car was licensed); the number of adults and children (persons under 12 years of age); equipment carried such as a boat, house trailer, or camping trailer; and time of day, date, park, and location (main gate or boat ramp). The number of adults and children was of greater importance prior to 1967 because the fee charged was dependent on the number of adults in each car. However, in 1967 the state introduced a fixed rate for each vehicle; and in 1968, an optional season pass was available.

All the interviews in 1967, 1968, and 1969 were conducted during weekends from Friday afternoon to Sunday afternoon during the months of June, July, and August. Weekends were assigned at random. In 1967, each park was visited on 3 weekends.

In 1968, Mansfield and Cagles Mill were each visited on 3 weekends, and Monroe was visited on 4 weekends. In 1969, each park was visited on 4 weekends.

The general procedure adopted in 1965 and 1966 was maintained during 1967, 1968, and 1969. Interviewing took place on Fridays from 2 p.m until 9 p.m., Saturdays from 9 a.m. until 8 p.m., and Sundays from 9 a.m. until 5 p.m. These hours were selected on the basis of a pilot study made at Mansfield in 1965. After about 9 p.m. on Fridays and before 9 a.m. on any day of the week, few arrivals were noted. The parks were open 24 hours a day throughout the summer, but interviews were conducted only during the stated hours. The park records on attendance showed that on weekends the arrivals during the interview period usually accounted for about 90 percent of the total visitors on Saturdays and Sundays and for about 75 percent on Fridays.

During the 3-year period (1967 through 1969), 11,800 samples were collected by the interviewers and, of this number, 11,400 were usable. The data obtained from the interviews were coded for the summation program that was used primarily to determine the number of annual trips to each park from each county in Indiana and Illinois and from other states.

It was not an unusual occurrence for a visitor to report multiple purposes when asked the reason for visiting a particular reservoir. It is probable that most trips to a reservoir are made with more than one purpose in mind. However, in this study, only the purposes reported were recorded because these were considered to be the purposes that inspired the trip. Also, no effect was made to determine whether, in fact, the stated purposes were actually accomplished. The trip purposes considered were boating, camping, fishing, picnicking, hiking, swimming, looking, and other.

It was apparent, once county trip totals were determined, that more than 90 percent of all trips originated from within 125 miles of a reservoir. Thus, for the purpose of this analysis, no counties beyond 125 miles of each reservoir were considered. The observed trips per county beyond this range were so sparse as to be insignificant. To standardize the trip rate from any particular county required a unit of measure. The previous model used trips per 1,000 population; this was adhered to in this phase. The official attendance (vehicles) for each year at Mansfield and Cagles Mill was obtained from attendance records maintained by the Department of Natural Resources; Monroe attendance figures were obtained from the park superintendent at Monroe Reservoir and the U. S. Forest Service.

The official total attendance figure for each reservoir was divided by the appropriate total attendance expansion factor, which is the ratio of samples interviewed at boat ramps to samples interviewed at main entrances (Table 1). In the case of Monroe, the expansion factors were applied only to the official attendance figures of the State Recreaction Area; the attendance at Hardin Ridge (U. S. Forest Service) Recreation Area was included later. The estimated total attendance (vehicles) at each reservoir for each year is given in Table 1.

The observed trips from a county were divided by the appropriate county trip expansion factor, which is the proportion of the estimated total park trips that were sampled in a year. County trip expansion factors are given in Table 1.

The Indiana county population estimates for 1967, 1968, and 1969 were linear interpolations of projections developed by the Indiana University, Graduate School of Business (3). The Illinois county population estimates were linear projections of 1960 census data and U. S. Bureau of Census estimates for 1966 (4). The distance figures were developed from the center of each county to the center of each reservoir. Road miles of the primary highway system were measured.

It became apparent, when Illinois and Indiana county trip rates were compared for equivalent distances from a reservoir, that Illinois county trip rates were significantly lower. It was necessary that a state-line penalty equal to 30 miles be added to all Illinois counties. This has the effect of including in the analysis only those Illinois counties within 95 miles of a reservoir.

ANALYSIS

Model Development

For each reservoir, a plot of the county trip rates (calculated from 1967, 1968, and 1969 data) versus distance from the reservoir indicated an exponential relation. This supported Matthias' choice of an exponential model to describe the 1965 and 1966 data. It should be added that this result was not entirely unexpected because previous research (5) showed an exponential relation between trip length and distance.

The form of the function used by Matthias and subsequently in this research is

$$Y = A\ e^{-BX}$$

where

Y = annual trips/1,000 population from a county to a reservoir;
A = Y intercept of nonlinear regression curve;
B = rate of change of nonlinear regression curve; and
X = distance from a county to a reservoir, in tens of miles.

There are 2 approaches by which parameters A and B in the equation may be estimated. First, it is possible to use the method of least squares after the function is transformed into

$$\ln Y = \ln A - BX$$

Second, a nonlinear regression analysis that estimates the parameters in an iterative manner may be applied. The first approach assumes that the errors in the transformed function are additive, which necessitates that the errors in the original be multiplicative (an assumption that has no physical basis). The second approach assumes that additive errors are in the original function and, because errors of an additive nature are more probable, this method was adopted. (An added benefit from the use of the second approach was only apparent later; this was when certain distant counties were found to have 0 trip rates. A logarithmic transformation would not have been able to deal with this situation.)

The nonlinear regression analysis utilized was NONLIN, a revised version of SHARE 3094 (6); this was described in some detail by Matthias (1). Basically the program finds the estimates of parameters A and B in the function $Y = A\ e^{-BX} + \epsilon$ by minimizing

$$\sum \epsilon^2 = \sum (Y - \hat{Y})^2$$

where ϵ is the residual error, and $\hat{Y}$ is the estimate of Y. It is an iterative technique that requires an initial estimate of the parameters A and B.

The previous model was made up of 2 regression equations. One equation was to be used for counties that are closest to the specified reservoir, and the other for counties that are closer to one or more other reservoirs than to the specified reservoir. The decision was made to arrange the data into 9 subgroups. Six subgroups were for a combination of Cagles Mill and Mansfield (3 years by closest and intervening categories); and 3 subgroups were for Monroe (3 years by 1 group containing all counties). There are 2 reasons for isolating Monroe data: (a) It is apparent from the total attendance figures that Monroe is still in its initial growth period (in contrast to Cagles Mill and Mansfield, which are older reservoirs), and (b) Monroe is a much larger reservoir than either of the other two (10,750 acres compared to Cagles Mill's 1,400 and Mansfield's 2,100), and is, in a sense, unique because it will remain the largest single body of water in the state for many years. The second reason is essentially the reason for the closest and intervening county groups being combined for Monroe. It is felt that Monroe is such a large trip attractor that intervening opportunities are not really applicable. Most of the reservoirs planned by the state are more nearly the size of Cagles Mill and Mansfield; therefore, for predictive purposes, a model based on data from these 2 reservoirs should be more reliable than one that either includes Monroe data or is based on Monroe data alone.

Monroe Model—The idea underlying the following analysis is that, if it can be shown that the parameter B_i (for $i = 1$ to 3) does not vary significantly among the 3 years, it might be possible to derive a prediction equation (with a pooled estimate of parameter B) by extrapolating the parameter A to the design year.

The first step in the analysis was to test for homogeneity of variances of the trip-rate data during the 3-year period. This assumes that the regression equations were reasonable predictors to the data (which they were). Under these conditions, testing from homogeneity of variances in the data is approximately equivalent to testing for homogeneity of the error estimates of the regression equations. Homogeneity of the error estimates of the regression equations is necessary in order to test the significance of B_i.

Two tests were applied to the data: (a) Bartlett's test (7), in which a chi-square statistic is computed (assuming that there are normal populations), and (b) Foster-Burr's test (8), in which a Q-statistic is computed that is a monotone function of the coefficient of variation of the sample variances. The fact that the populations are not normal reduces the inferences possible from Bartlett's test; however, less research has been directed toward non-normal populations than to normal populations, so the test was applied bearing the limitations in mind.

Both Bartlett's test and Foster-Burr's test produced highly significant statistics for the raw data, Y, and transformed data, ln(Y + constant), and this led to the rejection of the hypothesis of equal population variances (the constant was added to enable the logarithmic transformation to be made). On the basis of this result, it was decided to delete from the data those counties having trip rates of less than 1.0. It was hoped that homogeneous variances would result from this action. This reduced the sample sizes from 64 for each year to 52, 46, 51 for 1967, 1968, 1969 respectively. The nonlinear regression program was rerun with the smaller data sets, and the parameters produced are given in Table 2. Most of the parameters have been only slightly reduced by excluding trip rates less than 1.0.

Bartlett's test and Foster-Burr's test were applied to these data; the chi-square statistic from Bartlett's test was 0.393, and the Q-statistic from Foster-Burr's test was 0.335, both of which are insignificant at an α-level of 0.01. In this case the hypothesis of homogeneity of variances cannot be rejected.

It was then possible to test the hypothesis that the parameters B_i are equal. The procedure, explaned by Ostle (9), is to first test the hypothesis that all the observations can be described by 1 regression equation. If the F-statistic computed is significant (leading to the rejection of the hypothesis), the hypothesis of equal parameters B_i can be tested by another F-test. F-values of 8.63 and 0.496 respectively were obtained from the 2 tests; thus, the hypothesis that all the observations can be described by 1 regression equation is rejected at an α-level of 0.25.

The pooled estimate of parameter B for inclusion in the equation for each year was established when the nonlinear regression program was run for 1967, 1968, and 1969 data combined. The value of B was calculated to be 0.558. As a last step, the nonlinear regression program was rerun for each year, a regression line with parameter B = 0.558 was forced through the data, and parameter A was obtained in the equation

$$Y - A\ e^{-0.558X}$$

The 3 equations that resulted were as follows:

$$Y = 217\ e^{-0.558X} \text{ for } 1967$$

$$Y = 355\ e^{-0.558X} \text{ for } 1968$$

$$Y = 634\ e^{-0.558X} \text{ for } 1969$$

Figure 2 shows how parameter A varies from 1967 to 1969. The sharp increase that has occurred is a combination of the growth of Monroe in terms of facilities, reputation, and popularity and an increase in recreational trip-making in general. The former is by far the largest component of the growth.

From the explanation given above, an extrapolation of the present trend of parameter A (line A) is likely to overestimate the design-year parameter A. What is more likely to happen is a leveling off as indicated by lines B, C, and D. Unless there is knowledge of other factors, however, there is no basis for choosing any one line over the others. It was, therefore, decided to use the value of parameter A as obtained from the 1969 data and to acknowledge that it is a conservative estimator of the total annual trips to Monroe in some future year.

Figure 1. Major reservoirs in Indiana.

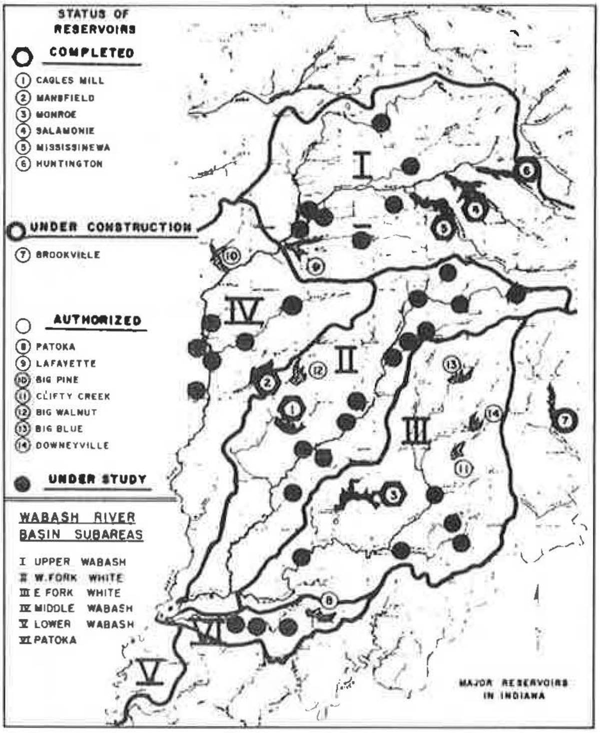


Figure 2. Changes in parameter A in Monroe model from 1967 to 1975.

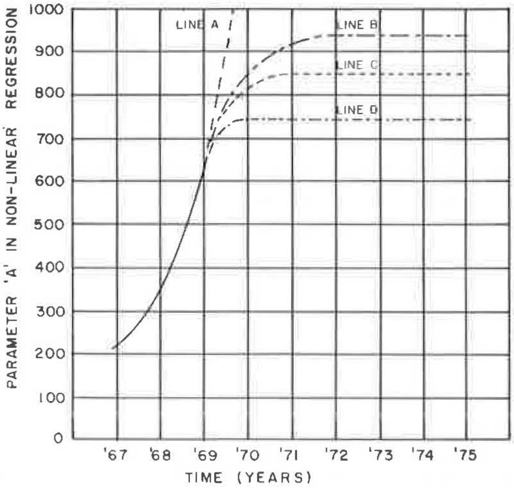


Table 1. Attendance and trip expansion factors.

Facility	Year	Total Attendance Expansion Factor	Estimated Total Attendance	County Trip Expansion Factors
Cagles Mill	1967	0.814	42,713	0.015
	1968	0.781	50,570	0.012
	1969	0.760	43,149	0.022
Mansfield	1967	0.739	60,486	0.018
	1968	0.808	63,592	0.014
	1969	0.855	41,477	0.030
Monroe	1967	0.915	39,269	0.051
	1968	0.897	77,758	0.012
	1969	0.870	108,646	0.013

Table 2. Nonlinear regression parameters.

Trip Rates <1.0	Year	Monroe		Closest		Intervening	
		A	B	A	B	A	B
Included	1967	228	0.606	517	0.571	387	0.715
	1968	342	0.530	554	0.736	105	0.354
	1969	656	0.588	363	0.523	202	0.548
Excluded	1967	222	0.576	516	0.570	243	0.453
	1968	340	0.525	520	0.638	107	0.305
	1969	648	0.575	362	0.521	127	0.511

The choice of a 1969 equation as the prediction equation was not based on the fact that no significant differences were found between the B_1 parameters. The equation adopted as the Monroe model,

$$Y = 634\ e^{-0.558X}$$

however, does include a contribution from the data of each year (the pooled estimate of B), and so the previous analysis has not been ignored. The Monroe model is shown in Figure 3.

New Reservoirs Model—The new reservoirs model is to consist of 2 equations (one from each of the closest and intervening groups of equations) that are considered to be the best for prediction purposes.

Exactly the same procedure that was employed in the development of the Monroe model was employed to develop the equations for the new reservoirs model. The parameters of the 3 equations for the initial run of the nonlinear regression program are given in Table 2. Once again, the variances of each year's data were not homogeneous until counties with trip rates of less than 1.0 were excluded from the analysis. The program was rerun, and the parameters of the 3 new equations obtained are also given in Table 2. The results from Bartlett's test and Foster-Burr's test for those data were such that the hypothesis of equal variances could not be rejected.

The hypothesis of equal B_1 parameters for each year for closest and intervening was tested next, and in both cases it was found that the hypothesis could not be rejected. The data for each year were combined within closest and intervening, and the nonlinear regression program was rerun to find a pooled estimate of parameter B for each group. The pooled estimates of B were 0.573 and 0.407 for closest and intervening respectively. The result of forcing these B values through the data for each year is the following equations for closest:

$$Y = 520\ e^{-0.573X} \text{ for } 1967$$

$$Y = 465\ e^{-0.573X} \text{ for } 1968$$

$$Y = 398\ e^{-0.573X} \text{ for } 1969$$

and for intervening:

$$Y = 212\ e^{-0.407X} \text{ for } 1967$$

$$Y = 151\ e^{-0.407X} \text{ for } 1968$$

$$Y = 136\ e^{-0.407X} \text{ for } 1969$$

It is immediately apparent that parameter A is decreasing in both cases (while in the case of Monroe, parameter A was increasing yearly). To understand why this is the case requires that the location of Mansfield and Cagles Mill with respect to Monroe be considered. All 3 reservoirs are within 60 miles of each other; because of this, it would be naive to think that the attendance at Mansfield and Cagles Mill should remain unaffected during the growth period of Monroe. It is considered likely that this downward trend in parameter A is no more than a transient response to the appearance of Monroe and that it will not continue for more than a few years. For this reason and for the reason that the future recreational reservoirs (for which the new reservoirs model is intended) will not be close to such a large facility as Monroe, it was decided to use the equations that were developed from 1967 data for closest and intervening.

The actual equations adopted to constitute the new reservoirs model are

$$Y = 520\ e^{-0.573X} \text{ for closest}$$

$$Y = 212\ e^{-0.407X} \text{ for intervening}$$

Both equations, which are shown in Figures 4 and 5 respectively, use the pooled estimate of parameter B.

Trip-Making Characteristics

Total Annual Trips—It is not enough for the planner to know how many trips (as predicted by the new reservoirs model) will be made to a particular reservoir in any year. The additional information that he requires is the distribution of those trips during the year, the week, and the day so that he can provide for adequate park facilities, seasonal hiring of park staff, and easy and adequate access. Because the planner is interested in the maximum volumes, it is in terms of these that the following analysis is performed.

Approximately 95 percent of all trips to a reservoir are made between the beginning of April and the end of September. This is based on the earlier study, for no out-of-season interviews were performed during this phase. The maximum volume week was determined for each reservoir for each year from official attendance figures, and the average ratio of maximum volume week to total annual trips was calculated to be approximately 10 percent.

In the earlier work, it was found that, on the average, 25 percent of all weekly trips arrived at the reservoir during the period from Monday through Friday morning, assuming that weather conditions were similar. This means that, on the maximum volume weekend, 75 percent of the 10 percent of the total annual trips to the reservoir can be expected, which amounts to 7.5 percent.

Approximately 50 percent of all weekend trips arrived on Sunday. It is, therefore, concluded that on the maximum volume weekend the reservoir attendance will amount to 7.5 percent of the total annual trips and that the highest daily volume (3.75 percent of the total annual trips) will occur on Sunday.

Trip Distribution—A further breakdown may be made on the basis of hourly arrivals that were recorded for each reservoir in the initial phase. It can be seen that, on the average, 62 percent of all Sunday arrivals come in the 4-hour period between 11 a.m. and 3 p.m. This information can be used to calculate the capacity required on reservoir access roads.

Besides the vehicular trips that can be expected on the maximum volume weekend, it is of importance to know how many people are associated with those trips. During this study, it was found that the average number of persons per trip was 3.75 and the average number of children per trip was 1.02.

Figure 6 shows that 90 percent of the sample trips originated within the 125-mile radius adopted for this analysis. This median distance traveled is 52 miles, and the associated travel time is 62 minutes.

RESULTS

The primary objective of this phase was to evaluate the growth trends of recreational usage of multipurpose reservoirs with reference to the model developed earlier. The choice of an exponential model, $Y = A e^{-BX}$, to relate trip rates and distances in the earlier phase was substantiated by the data collected during this study. Three equations of the same form were developed. Of these, 2 equations (developed from data collected at Mansfield and Cagles Mill reservoirs) constituted the new reservoirs model. The third equation (the Monroe model) is to be used to predict annual trips to Monroe reservoir only.

The 2 equations developed in the initial phase for the closest and intervening categories were $Y = 338\ e^{-0.479x}$ and $Y = 129\ e^{-0.488}$ respectively. Comparing the equations from both phases shows that, although an increase in the value of parameter A (by factors of 1.54 and 1.64 for the closest and intervening categories respectively) has occurred over time, there has been little change in the value of parameter B (almost none in the case of the closest counties). This is an important result, for it implies that a growth in the trip rates (which was being investigated in this phase) is best measured by changes in the value of parameter A. Furthermore, if continued study indicates even higher trip rates, only the parameter A in each of the 2 equations need be adjusted. It is not known by how much or in what manner parameter A of the 2 equations is likely to change during a period of 1 or 2 decades. The data collected in this recreational study rendered any prediction of the future behavior of parameter A unwise; only the fact that A did increase over time was observed.

Figure 3. Annual trips to Monroe.

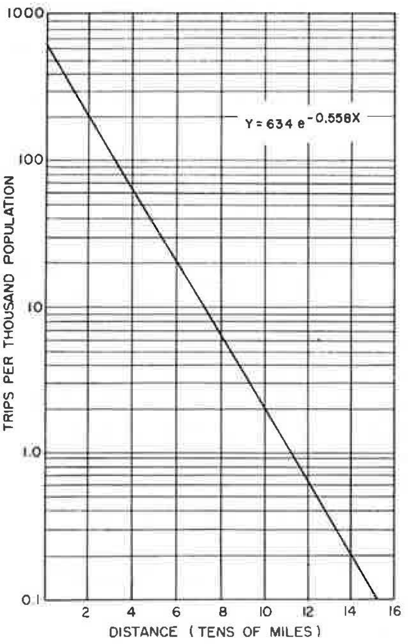


Figure 4. Annual trips to closest park.

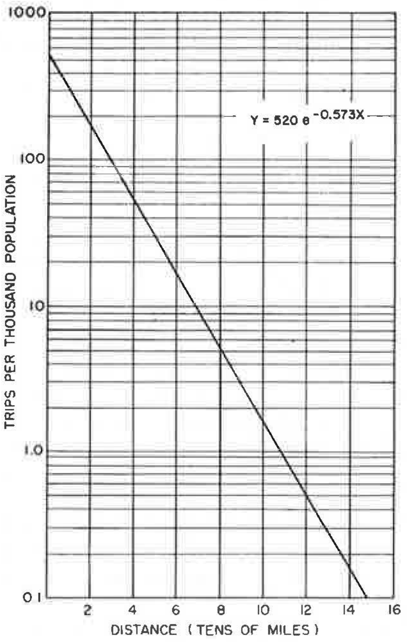


Figure 5. Annual trips to intervening park.

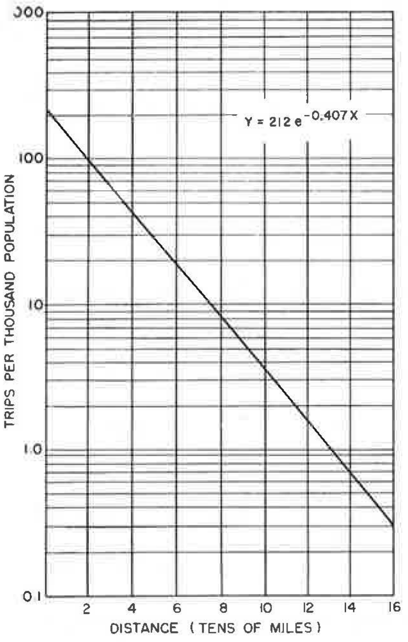
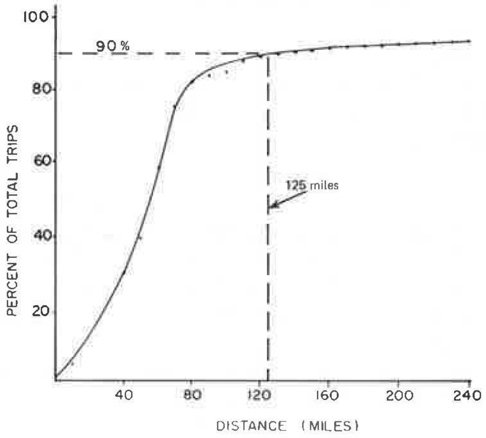


Figure 6. Cumulative distribution of trips.



The single equation that constitutes the Monroe model is $Y = 634\ e^{-0.558x}$. There is no way of comparing this equation with those of the new reservoirs model because in its development all counties within 125 miles of Monroe were used in the same category. No Monroe data were used in the initial phase, so there is no way to make a comparison between the 2 phases of the study. This equation is a conservative estimator for the total annual trips to Monroe because Monroe can still be considered to be in its initial growth period.

It is concluded that the new reservoirs model, which is based on easily understood and readily obtainable variables (distance, population, and influence of similar facilities), is able to predict future attendance at new reservoirs with reasonable accuracy. In contrast to other previously developed models, which require many socioeconomic and park characteristics variables (often difficult to measure and evaluate and extremely difficult to project), the new reservoirs model is probably as accurate and much simpler to use. The new reservoirs model is adequate for advanced-planning purposes and can be used to predict reservoir attendance and traffic volume estimates.

The objectives of this study were to check the previously developed models for stability over time and to present a simplified procedure that could be easily implemented by the highway department. One can conclude that models of this type must be under constant surveillance, for the demand function is obviously changing. The simplified prediction procedure can be summarized in the following manner:

1. Determine the location of the reservoir;
2. Locate other similar recreational facilities;
3. Determine the road distance (miles) to the reservoir from counties within 125 miles;
4. Obtain county population predictions for the design year;
5. Determine which of the counties are closer to the reservoir under study than to any other similar facility;
6. Determine the trip rates for each county closest to the reservoir (Fig. 4);
7. Determine the trip rates for the remaining counties (Fig. 5);
8. Calculate for each county the total annual trips by multiplying the trip rate by the population prediction; and
9. Sum the total annual trips for all counties, and divide by 0.9 to account for trips originating farther than 125 miles away and to obtain the estimated total trips for the design year.

REFERENCES

1. Matthias, J. S. Recreational Impact of Multi-Purpose Reservoirs. Joint Highway Research Project, Purdue Univ., Aug. 1967.
2. Matthias, J. S., and Grecco, W. L. Simplified Procedure for Estimating Recreational Travel to Multi-Purpose Reservoirs. Highway Research Record 250, 1968, pp. 54-69.
3. Indiana Population Projections: 1965-1986. Graduate School of Business, Indiana Univ., Res. Rept. 3, Sept. 1966.
4. U. S. Bureau of the Census. Estimates of the Population of Counties and Metropolitan Areas, July 1, 1966: A Summary Report. U. S. Govt. Printing Office, Washington, D. C., Current Population Repts., Series P-25, No. 427, 1969.
5. Data Projections. Chicago Area Transportation Study, Final Rept., Vol. 2, July 1960.
6. Marquardt, D. W. Least Squares Estimation of Non-Linear Parameters, A Computer Program in FORTRAN IV Language. IBM, Share Library Distribution 3094, March 1964.
7. Bartlett, M. S. Some Examples of Statistical Methods of Research in Agriculture and Applied Biology. Jour. the Royal Statistical Soc. (suppl.), Vol. 4, No. 137, 1937.
8. Foster, L. A. Testing for Equality of Variances. Purdue Univ., PhD thesis, 1964.
9. Ostle, B. The Problem of Several Samples of Groups. In Statistics in Research, Iowa State Univ. Press, 1963, Section 8.26.

RELIABILITY ANALYSIS IN LAND USE-TRANSPORTATION PLANNING

Kumares C. Sinha, Department of Civil Engineering, Marquette University

ABRIDGMENT

•IN THE preparation of land use-transportation system plans for any area, the input data involve several demographic and economic variables. For example, the future transportation demand of an area can be predicted on the basis of variables such as population, automobile ownership, and employment. Once the estimates of the input variables for a future design year are prepared, the aggregate information about the transportation demand for that design year can be obtained, and the required transportation facilities can be planned accordingly. However, the estimation of each of the causal variables is associated with uncertainties. These uncertainties are associated to a certain extent with the theory and technique used in predictive models that determine functional relations of the transportation demand with the causal variables. Similar statements can be made for most of the planning process, for land use-transportation planning entails a complex system acted on by numerous variables whose behavior is not well understood. In the area of planning methodology, the emphasis has been thus far on the development of approximate descriptions of unpredictable, socioeconomic phenomena. In planning practice, however, the use of any analysis is to aid the planner in making decisions. These decisions might involve the planning and design of public facilities such as transportation systems, sewerage and water supply systems, educational facilities, and recreational areas. Because of the uncertainties associated with the predictions of future year demands of such systems, the planner is faced with a difficult task of choosing the most accurate design level of demands. For example, if the design demand is expressed in terms of population figures and if the estimated future population figures are associated with a wide range of uncertainty, no amount of sophistication and refinement of the planning models will cause any substantial improvement in the probability of success of the facilities planned on the basis of these models.

For an efficient and effective planning process, a planner must make decisions as prudently as possible in the face of uncertainty. Before a particular element in plan design is included, the planner must search for the best design suited to the requirement involving the probability of success of the particular design element. In most situations the planner will obviously depend on his judgment as well as his experience. However, Bayesian decision statistics provide an excellent tool that can aid the planner in arriving at an optimal decision by combining his subjective judgment with objective information.

A planned system may be defined as one that is conceived, designed, and implemented to a specified level of reliability. A level of reliability is a measure of assurance that the system planned will serve its intended function successfully. Planning reliability can then be defined as a quantitative measure that the system planned will achieve successfully some presented level of performance.

In case of a transportation system plan, the component elements that constitute the plan design must be systematically examined in terms of their adequacy for the required demand during their intended life. For example, one aspect of the planning reliability of a transit system can be conceived as the measure of success in the prediction of ridership demand. Because the design level of ridership demand will determine the extent of the facilities to be provided, it is essential that the associated reliability in ridership prediction be carefully examined. The reliability of a highway system plan must also be examined in terms of its adequacy for the intended period of its design life.

Both the future ridership demand for a proposed transit system and the expected traffic volume on a planned highway network can only be estimated as a prediction. Because of the nature of the forecasting process in transportation system planning, the adequacy of a plan design can only be examined in the face of uncertainty. In other words, the evaluation of planning reliability is mostly concerned with the analysis of uncertainties involved with the forecasting process. With the application of a probability theory, a rational procedure can be developed to examine the extent of uncertainties involved in a particular demand forecast. Then, a decision theoretic approach can be adopted to determine the optimal level of a design parameter that is to be considered in systems planning.

As an example, the problem of forecasting the total population in the 7 counties of the southeastern Wisconsin region was considered. The analysis was performed in 2 steps. In the first step, the reliability of population projections obtained from the application of various techniques was tested. In the second step, estimates were prepared to evaluate the reliability of forecasting in a particular county. Subsequently, an optimal decision was made regarding the level of future design-year population for each county. The reliability of population projections was measured by the probability of predicted population figures falling within a given tolerance range. The reliability values for the different forecasting techniques considered ranged from 0.660 to 0.977 for a tolerance range of ± 5 percent. On the other hand, the reliability of forecasting for individual counties ranged from 0.868 to 0.969 for the same tolerance range. The upper bound of the optimal size of the sample to be tested was found to be 6 for given prior information and for assumed values of severity constant and unit sampling cost.

The example discussed here is a simple one; in reality, however, the planning and design decisions are associated with multiple variables, and a reliability analysis of the entire system should be made by optimizing the combined expected utility. Futhermore, more reliable cost data are required to develop associated utility functions. These data would include the cost items involved with the overdesigns and underdesigns of a planned system such as a transportation facility. However, these cost data can be developed with reasonable accuracy on the basis of the records maintained by the various public and private agencies.

The procedure outlined in the paper provides an opportunity for a planner to arrive at an optimum level of design-year population. Such a procedure can be an extremely useful tool in the overall land use-transportation planning process.

ACKNOWLEDGMENT

The work was in part sponsored by a research and training grant from the Urban Mass Transportation Administration.

DISTRIBUTION PATTERNS OF AUTOMOBILE TRAVEL IN THE NEW YORK METROPOLITAN AREA

J. David Jordan, Tri-State Regional Planning Commission

ABRIDGMENT

•PLANS for highway improvements are dependent on traffic growth. Even aside from automobile ownership, this growth in traffic can be studied at the household, the basic trip-producing unit. To this end, home interview survey data for selected communities in the New York metropolitan area were processed to obtain the geographic distribution of vehicle-miles traveled in relation to the location of the households owning cars. Of particular interest were variations in the amount and distribution of travel due to the effect of density and supply of expressways.

STUDY DESIGN

Four areas were selected for analysis: 2 in New York City and 2 in the suburbs. Each was 4 square miles in size and was identified by the name of the community that covered most of its area. The areas were chosen in such a manner as to form 2 matched pairs. Within each pair, the 2 constituent communities had similar population densities and roughly equivalent transit service but differed in the supply of expressways.

Area	Community		Density Range (persons/ sq mi)	Expressway Access
	Number	Name		
New York City	1	Flatbush, Kings County	35,000	Poor
	2	Fordham, Bronx County	40,000	Good
Suburbs	3	Westwood, Bergen County, New Jersey	5,000 to 6,000	Poor
	4	Albertson, Nassau County, New York	5,000 to 6,000	Good

Automobile driver trips originating from or terminating at households within each selected area were summarized according to their airline trip lengths from the area's geometrical center. The vehicle-miles of travel generated by these trips were also summarized according to distance from the center.

RESULTS

The suburban areas generated about twice as many automobile driver trips per automobile as those in the city. Within each of the matched pairs, however, the differences in the supply of expressways had an insignificant effect on the travel generated per vehicle.

Community	Automobile Driver Trips/Automobile
1	2.1
2	1.8
3	4.4
4	3.8

Because there was no significant variation in average trip length among the 4 areas, the suburban areas also generated about twice as many vehicle-miles of travel per automobile as the city pair, with variations in expressway supply causing negligible differences.

Community	Vehicle-Miles of Travel/Automobile
1	7.9
2	8.0
3	17.1
4	16.7

The automobile driver trip distribution pattern displayed minor differences; the city areas had a larger percentage of trips in the 2- to 5-mile range, and the suburban areas had a higher percentage beyond 14 miles. This tendency was more apparent in the distribution of vehicle-miles of travel where the suburban areas had about a quarter of their travel beyond 10 miles, compared to 11 to 16 percent for the city pair.

Community	VMT Within 5 Miles of Area Centroid (percent)	VMT Beyond 10 Miles From Area Centroid (percent)
1	65.1	16.1
2	65.7	11.2
3	56.7	23.4
4	51.5	26.2

Because the 2 pairs of locations represented extremes of density, a third pair of communities was selected for similar analysis. Both locations were in the intermediate density range of 15,000 to 20,000 persons/square mile, but one was better served by limited-access facilities. The amount of vehicle travel generated per automobile was quite similar for both and was between the values obtained earlier for the high density pair and the suburban pair.

CONCLUSION

An analysis of home interview survey travel data in 4 selected communities indicated that the supply of expressways had a minimal effect on the generation and distribution of automobile trips. Residents of the suburban communities generated about twice as many vehicle-miles per automobile as did those who lived in the denser urban areas. Generally, as density increased, the week-day mileage generated per car decreased, no doubt because of an improved level of transit service and the greater costs associated with moving and parking automobiles.

ACKNOWLEDGMENTS

Invaluable aid was received in the preparation of the complete version of this paper from several persons: general guidance and direction from Lawrence V. Hammel and Edward F. Sullivan; technical assistance from William H. Coelln, Thea Edelstein, and Elizabeth Phifer; clerical assistance from Diane Williams; and data processing by Widge Connolly. The preparation of the report was financed in part through funds made available by the Federal Highway Administration and an urban planning grant from the U. S. Department of Housing and Urban Development, under the provisions of Section 701 of the Housing Act of 1954, as amended, and in cooperation with the states of Connecticut, New Jersey, and New York.

SPONSORSHIP OF THIS RECORD

GROUP 1—TRANSPORTATION SYSTEMS PLANNING AND ADMINISTRATION
Charles V. Wootan, Texas A&M University, chairman

Committee on Transportation Forecasting
George V. Wickstrom, Metropolitan Washington Council of Governments, chairman
Mark M. Akins, Daniel Brand, Austin E. Brant, Jr., Henry W. Bruck, Francis E. Coleman, Thomas B. Deen, William L. Grecco, Charles H. Groves, John R. Hamburg, Lawrence V. Hammel, Donald M. Hill, Wolfgang S. Homburger, Frank E. Horton, Gerald H. Johnston, Louis E. Keefer, Robert Kochanowski, Herbert S. Levinson, Brian V. Martin, James J. McDonnell, Eugene R. Mills, Kenneth W. Shiatte, Paul W. Shuldiner, Vergil G. Stover, Anthony R. Tomazinis, Robert T. Wood, Richard D. Worrall

James A. Scott, Highway Research Board staff