TRANSPORTATION RESEARCH RECORD 673

# Transportation Forecasting and Travel Behavior

TRANSPORTATION RESEARCH BOARD

COMMISSION ON SOCIOTECHNICAL SYSTEMS
NATIONAL RESEARCH COUNCIL

NATIONAL ACADEMY OF SCIENCES
WASHINGTON, D.C. 1978

**Transportation Research Record 673**
Price $11.00

modes
1 highway transportation
2 public transit
4 air transportation

subject areas
12 planning
13 forecasting

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

**Notice**

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

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

To eliminate a backlog of publications and to make possible earlier, more timely publication of reports given at its meetings, the Transportation Research Board has, for a trial period, adopted less stringent editorial standards for certain classes of published material. The new standards apply only to papers and reports that are clearly attributed to specific authors and that have been accepted for publication after committee review for technical content. Within broad limits, the syntax and style of the published version of these reports are those of the author(s).

The papers in this Record were treated according to the new standards.

**Library of Congress Cataloging in Publication Data**
National Research Council. Transportation Research Board.
Transportation forecasting and travel behavior.

(Transportation research record; 673)
Includes bibliographical references.
1. Transportation planning—Congresses. 2. Choice of transportation—Mathematical models—Congresses. 3. Urban transportation—Congresses. I. Title. II. Series.
TE7.H5 no. 673 [HE193] 380.5'08s [380.5'2] 79-12520
ISBN 0-309-02813-2

**Sponsorship of the Papers in This Transportation Research Record**

**GROUP 1—TRANSPORTATION SYSTEMS PLANNING AND ADMINISTRATION**
*E. Wilson Campbell, New York State Department of Transportation, chairman*

**Transportation Forecasting Section**
*John R. Hamburg, John R. Hamburg and Associates, chairman*

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

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

Committee on Aviation Demand Forecasting
*David W. Bluestone, Consultant, Silver Spring, Maryland, chairman*
*Michael R. Armellino, G. R. Besse, Cecil O. Brown, Arthur S. De Vany, George Howard, Adib Kanafani, Gene S. Mercer, William R. Nesbit, Terrence Lee Parker, David E. Raphael, Robert H. Rollins II, George N. Sarames, Spyridon N. Sideris, Laurel A. Smith, Edward C. Spry, Nawal K. Taneja, Kenneth R. Velten*

James A. Scott and Herbert J. Guth, Transportation Research Board staff

Sponsorship is indicated by a footnote at the end of each report. The organizational units and officers and members are as of December 31, 1977.

# Contents

# Functional Analysis of Mode Choice

Robert J. Meyer, Department of Geography, and Irwin P. Levin, Institute of Urban and Regional Research, University of Iowa
Jordan J. Louviere, Center for Behavioral Studies of the Institute for Policy Research, University of Wyoming

**This paper develops a relatively new paradigm for mode-choice behavior modeling. The paradigm emphasizes functional form by establishing functions that relate subjective evaluations of transportation system attributes to objective levels of these attributes and functions that relate observed mode choices and preferences to combinations of subjective impressions. These functions are derived from a theory of decision making and behavior that views the decision maker as an integrator of information. According to this theory, the overall evaluation of a given transportation system can be represented as an algebraic combination of the traveler's evaluations of the various attributes of the system. Two experiments were conducted to evaluate this paradigm. Questionnaires were distributed to respondents, who were asked to indicate their degree of preference for car or bus for each of a series of hypothetical mode-choice situations. These situations were generated by combining varying levels of time difference favoring car over bus, cost difference favoring bus over car, and number of riders in the car. Each judgment thus required a trade-off of cost, time, and interpersonal factors. Cluster analysis was used to separate respondents into distinct subgroups of homogeneous decision makers. These subgroups differed in terms of overall preference for car or bus and the relative weighting or trade-off of factors. Actual mode choice for work trips was then predicted on the basis of preference responses to the hypothetical mode-choice situations, estimates of cost and time factors for individual respondents, and transportation availability constraints. A high level of predictive validity was attained in each experiment. It is suggested that the present paradigm may be useful for analyzing traveler decision processes, for estimating latent demand for alternative transportation opportunities, and for predicting responses to altered or new transportation systems.**

The purpose of this paper is to report an empirical evaluation of a new paradigm in behavioral transportation modeling. The paradigm is largely an outgrowth of a recent series of investigations into the processes of transportation mode choice that used the information integration (or functional measurement) theory of human judgment research (1). The general emphasis of the paradigm is on functional form—establishing functions that relate objective attributes to subjective impressions and functions that relate subjective impressions to observed choice behavior. While its general structure has previously been outlined in some detail (2), the paradigm has yet to be fully evaluated empirically. It is toward the end of achieving such an evaluation that this research is directed.

The first part of this paper reviews the paradigm in light of the current state of the art in disaggregate transportation modeling, and the second part reports the results of two investigations designed to relate abstract (hypothetical) mode preferences to actual mode choices for a sample of consumers. The paper concludes with a discussion of the advantages and limitations of the approach in predicting mode choice vis-à-vis existing capabilities.

## STATE OF THE ART: RANDOM UTILITY MODELS

The literature on disaggregate modeling approaches to the study of mode-choice behavior is an extensive one. As a number of good reviews of this literature are available elsewhere (3), we shall provide only a rough sketch of the most salient directions evident in this literature in order to compare features of the proposed paradigm with those of the more traditional approaches.

The most prevalent modeling paradigm to date in the analysis of mode-choice behavior has been the stochastic (random utility) model (3). In this model, individuals are thought to hold independent utilities for each of N alternatives, and the probability that any alternative i will be chosen is the probability that the utility of alternative i is greater than the utility of any of the other N - 1 alternatives. These utilities are thought to comprise two independent elements: a vector of strict (nonstochastic) utilities reflecting observed characteristics of an alternative and an associated random component reflecting unobserved characteristics. The distribution function that best describes the random component is assumed on an a priori basis by a given investigator. Various model forms, such as the multinomial logit and probit models, are then generated by differing distributional assumptions.

The vector of strict utilities of an alternative may be characterized in terms of a set of salient attributes (dimensions) of the alternative and a transformation or composition rule by which the multidimensional vector is mapped into a unidimensional overall utility.

Until now there have existed few theoretical guidelines to assist the researcher in specifying either the components of the utility vector or the composition rule. Relevant attributes have commonly been assumed to be a set of objective measures of modes and users, such as observed mode travel time and user income (4), and the composition rule has traditionally been assumed to be linear-additive (5).

Given the arbitrary nature of these assumptions, stochastic choice models have been greatly weakened as theoretical tools for the analysis of individual traveler behavior. As a result, traditional models have served mainly as static, descriptive devices.

In recent years, a considerable amount of research has been directed toward providing a firmer behavioral basis upon which to construct models of traveler mode choice. Specifically, this research has been characterized by (a) attempts to identify relevant attributes in mode choices (6), (b) attempts to relate various attitudinal measures to mode choices (7), and (c) attempts to provide alternate conceptual frameworks for the analysis of traveler mode choice (8). While this research has served to introduce sets of methodologies and constructs that may be useful in the analysis of traveler behavior, it has fallen somewhat short of the goal of providing sets of firm behavioral postulates from which models of traveler choice behavior might be derived. Many basic questions relating to the characteristics and composition rules of individual utility functions remain unanswered.

### An Alternate Paradigm

We shall advance an alternate paradigm to serve as a framework for analyzing travel behavior. Its most significant departure from other conceptualizations (8) is its emphasis on functional form—i.e., the relationships between decision attributes and observed choice behavior. Such functions permit better understanding of observed behavior within existing transportation systems and better prediction of likely responses to changes in such systems.

Louviere and others (2) outlined the general form of a paradigm from which a behavior-based theory of travel behavior might emerge. It consists of a series of relationships thought to reflect how measurable attributes of travel modes are translated into individual choice behavior. Specifically, they defined $S_{ik}$ as the objective value of the ith attribute of mode k, $S_k$ as a vector of such attributes ($S_{1k}, S_{2k}, \ldots, S_{ik}$), $x_{ikl}$ as the subjective (perceived) value of the ith attribute of mode k for trip purpose 1, and $x_{kl}$ as a vector of perceived attributes of mode k for purpose 1 ($x_{1kl}, x_{2kl}, \ldots, x_{ikl}$). Further, they defined $R_{kl}$ as the unidimensional subjective value (utility) of mode k for trip purpose 1 and $T_{kl}$ as the observed patronage of mode k for trip purpose 1. They then established the following recursive system.

$$x_{ikl} = F(S_{ik}) \qquad (1)$$

$$R_{kl} = G(x_{kl}) \qquad (2)$$

$$T_{kl} = H(R_{kl}) \qquad (3)$$

$$T_{kl} = I(S_k) \qquad (4)$$

In other words, Louviere and others established a formal framework for examining the following relationships:

1. Function relating subjective perceptions of mode attributes to objective magnitudes;
2. Function by which an i-dimensional vector of perceived attributes is transformed into a unidimensional subjective response space;
3. Function relating such overall subjective responses to observed travel behavior; and
4. Composite rule relating the original objective attribute values to observed choice behavior.

Much of the traditional research in mode-choice modeling might be characterized as attempts to directly examine the relationship expressed in Equation 4. We, however, argue that such a relationship becomes meaningful only when it is established in the context of a recursive system (such as outlined above). In addition, recent attitudinal research might also fit this framework. Studies that have attempted to derive measures of the perceived quality of modal attributes (6) involve Equation 1, while studies that deal with the relationship between attitudes toward modal attributes and mode choices (7) involve Equation 3.

The following section describes one approach for simultaneously establishing the functional forms expressed in Equations 1-4.

## Functional Measurement

Functional measurement is a method of describing the judgment and decision processes underlying behavioral data. If our data are derived from observations of human choices and preferences, then the processes or functions describing these choices and preferences can be investigated within this framework. While other approaches such as conjoint measurement have been suggested for deriving such functions, functional measurement appears to provide the most flexible analytic tool because of its ability to diagnose alternative functional forms (combination rules). Reviews of functional measurement and its applications to modeling choice and decision rules are available (1, 8).

In this approach, each stimulus object is considered to be a combination of attributes. Algebraic rules or utility functions are used to describe the ways in which individuals trade off these combinations of attributes. The general form of the algebraic expression relating the overall evaluation or utility of a stimulus object (e.g., a transportation system) to the subjective values of its various attributes can be stated as $R_j = f(x_{1j}, x_{2j}, \ldots, x_{kj})$, where $R_j$ is the overall evaluation of stimulus combination j and $x_{ij}$ is the subjective value of attribute i on stimulus j.

In many applications, $R_j$ is a rating of the desirability of stimulus j. The function f is estimated from goodness-of-fit tests of alternative model forms. The parameters $x_{ij}$ are estimated from responses to various stimulus combinations and can then be related to objective stimulus attribute levels, $S_{ij}$. This relationship between $x_{ij}$ and $S_{ij}$ corresponds to Equation 1 above.

Equation 2 relating overall response to a combination of perceived attributes is typically obtained by an analysis of variance of responses in a factorial experiment where each dimension of a multiattribute stimulus is varied over several levels. The crucial design feature of such an experiment is that respondents make a single evaluation of a complex system rather than separately evaluate single attributes in unspecified contexts. Analysis of variance provides a goodness-of-fit test for alternative models. The reader is referred to Anderson (8) for a complete discussion of various model forms and how they are treated.

The functional measurement technique has been applied to the study of mode-choice decisions in a number of instances. For example, studies by Levin (1) employed functional measurement in the analysis of student mode preferences and generally uncovered decision rules that were nonlinear in form. These findings were significant in that they cast doubt upon the assumption of linearity in utility functions common in applications of existing mode-split models such as the multinomial logit. Although these studies were primarily concerned with diagnosing combination rules used in simulated mode-choice situations rather than with describing actual mode choices, some pilot work has tried to relate laboratory-derived models to actual mode choices. The results of these pilot studies suggested that habitual car drivers and bus riders have different trade-offs (combination rules) in evaluating car and bus attributes. Such results are encouraging, because they help to empirically define the relationships expressed in Equations 3 and 4 of our recursive system—that is, the relationships between subjective responses in a laboratory simulation setting and actual mode-choice behavior.

The experiments described below follow this latter trend and employ the functional measurement approach to further explore the relationships given in Equations 1-4. In addition, they expand the simple car-bus mode choice to include shared rides as well as solo car driving. The purpose of the experiments was to uncover the form of the algebraic utility or decision model underlying mode-choice trade-offs and to relate this model to actual mode-choice proportions. Also of interest were (a) whether variations in utility functions across groups of consumers can be related to socioeconomic, demographic, and situation characteristics, and (b) whether actual mode choices for work trips can be predicted on the basis of responses to hypothetical mode-choice situations.

## EXPERIMENT 1

Responses to a questionnaire were obtained from a sample of 99 employees of the University of Iowa. The questionnaire included sections designed to assess the worker's personal background, work schedule, distance from work, present mode split and satisfaction level, estimates of transportation costs, and estimates of the importance of a variety of factors related to transportation. The most important section of the questionnaire was a series of mode-choice responses to various trade-off situations. The specifics of this key series are described below. Finally, in an effort to gain information about constraints that may have influenced actual mode choices, the following open-ended question was inserted at the end of the questionnaire: "What are the most compelling reasons why you personally choose the method of travel you use to get to and from work?"

## Specifics of Mode-Choice Questions

Respondents were presented with descriptions of 27 hypothetical trade-off situations described by these factors: (a) time difference (0, 15, or 45 min/d longer for bus than for car), (b) cost difference (0, 25, or 75¢/d more for car than for bus), and (c) number of riders in the car with the driver (0, 1, or 3). This latter factor thus includes both ride sharing and solo driving as mode-choice alternatives. Each given situation was described by one level of each of two factors, (a) and (b), (a) and (c), or (b) and (c). One complete factorial design was formed for each pair of factors; three 3-×-3 factorial designs were formed overall.

In the instructions, respondents were told to assume that both a car and a bus were available to them in each hypothetical situation and that, assuming these availabilities, they were to respond on the basis of the information presented.

The purpose of this instruction was to elicit a car-bus propensity abstract from an availability constraint. For each hypothetical situation, the respondent was asked to rate the relative likelihood of taking the bus or car. A 20-point rating scale was used, where 0 represented "certain to take car" and 20 represented "certain to take bus." Respondents were to use numbers between 0 and 20 to represent varying degrees of preference for car or bus. This car-bus preference scale was used previously (1) and provides information about degree of preference as well as binary mode choice.

## Results

Description of the results will be divided into three phases. In phase 1, analyses of responses to the hypothetical mode-choice situations will be presented. Phase 2 will explore the relationships between decision processes identified in phase 1 and group differences in socioeconomic, behavioral, and situational characteristics. In phase 3, responses in the experimental task will be related to actual mode choices.

### Phase 1

Because each respondent completed only one replication of the experiment, decision models could not be tested at the level of the individual respondent. However, by grouping respondents who exhibited similar arrays of responses into segments, inferences about individual decision-making processes could be made at a minimum risk of fallacy.

In order to derive homogeneous decision-making segments, the raw responses of each of the 99 respondents in the experimental task were subjected to a cluster analysis. This appeared to provide the most reasonable grouping tool for this purpose by virtue of its ability to differentiate respondents based on both pattern and magnitude of response. Q-mode factor analysis, a possible alternative, would differentiate purely on pattern. For example, an individual who would take the car under all trade-off situations could, quite conceivably, be grouped with one who would ride the bus under all situations. Because we were interested in relating the groups to actual mode ridership, this would clearly not be a desirable result.

The results of the analysis suggested that the data set comprised three salient clusters of respondents, one with 30 members, one with 51 members, and one with 18 members. The next stage of analysis was to identify each group in terms of differences in their revealed decision-making processes.

To permit this identification, separate analyses of variance were performed for the responses of each group. Three two-way repeated measures analyses, corresponding to the three 3-×-3 designs contained in the experiment, were conducted for each segment. An examination of the grand mean across all cells for each analysis for each group provided a clear interpretation of the groupings.

Group 1 (30 members) had a grand mean of 6.3; group 2 (51 members) had a grand mean of 10.5; group 3 (18 members) had a grand mean of 15.3. Recalling that responses were recorded on a 20-point rating scale, we see clearly that group 1 was a car-biased group and group 3 was a bus-biased group. Group 2 was in the middle and, for reasons that will be made clear later, was defined as an unbiased group.

Plots of mean values for each cell of the cost difference-time difference subdesign for each group are shown in Figure 1. Several things should be kept in mind when examining the three panels. Parallel or nearly parallel lines show that the two factors being plotted combine in an additive fashion to determine car-bus preference ratings for that group. The slopes and separations of the lines reflect the relative degrees of importance or weights of the two factors.

The comparative spacing of the lines in a given panel and the shape of each line (straight, negatively accelerated, or positively accelerated) provide information about the psychophysical functions, that is, the relationships between objective attribute values and their subjective counterparts (Equation 1). If the lines in a given panel converge at a particular level of the variable plotted on the abscissa, this shows differential weighting (nonadditivity), with that particular level having greater weight than other levels of that variable.

For all three groups, preference for the car increased (approached the low end of the scale) as time savings for car over bus increased and preference for the car decreased as cost savings for bus over car increased.

For the car-biased group (Figure 1a) the nonparallelism suggests a nonadditive combination rule for time differences and cost differences in determining car-bus preferences. This was confirmed by a significant interaction in the analysis of variance. Convergence of the lines at a time difference of 45 min indicates that cost difference had less effect at high time differences than at low time differences. This finding replicates and extends the generality of the Levin study (1) and supports the interpretation that car-bus preferences are based on a weighted averaging (a common form of Equation 2) of cost and time factors, where respondents place greatest weight on those pieces of information that support their initial biases. Car-biased individuals thus tended to ignore the cost differences favoring the bus when time savings favoring the car were great.

For the unbiased group (Figure 1b), the curves are nearly parallel, which suggests that weight biases were not present in this group. The sizable spacings and

**Figure 1. Time difference and cost difference interaction plots for three bias groups.**

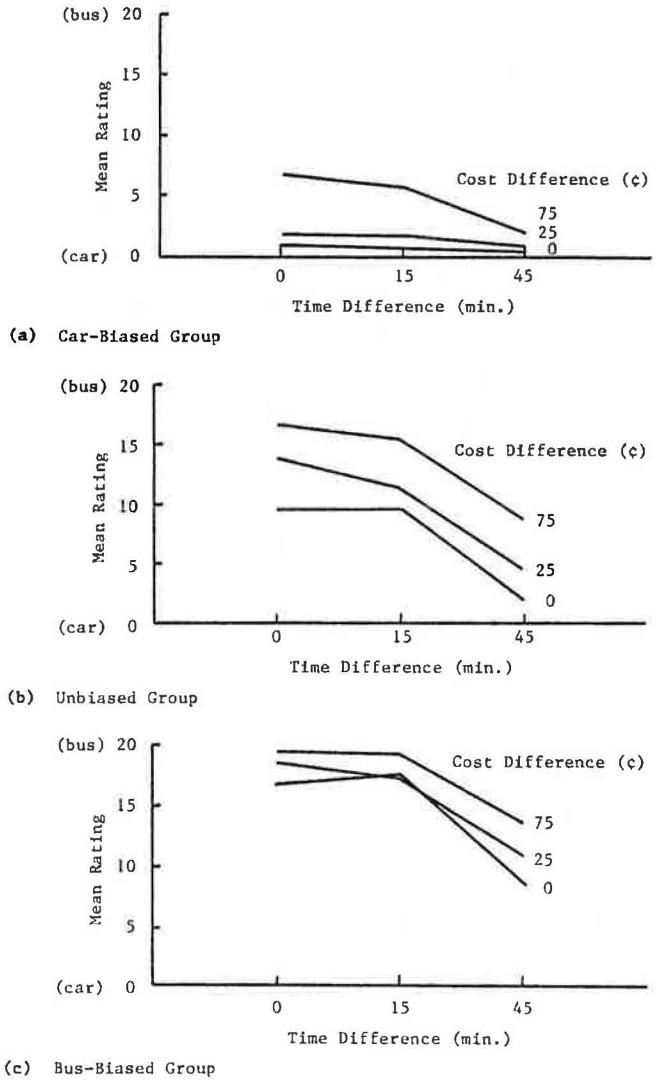



**Table 1. Group means and standard deviations on questionnaire items.**

| Item | Car-Biased Group | | Unbiased Group | | Bus-Biased Group | |
|---|---|---|---|---|---|---|
| | Mean | SD | Mean | SD | Mean | SD |
| Age, years | 37.4 | 10.5 | 36.3 | 13.9 | 39.1 | 13.1 |
| Sex, 0 = female, 1 = male | 0.47 | 0.5 | 0.49 | 0.5 | 0.61 | 0.5 |
| Income, 7 categories | 5.3 | 1.7 | 4.6 | 1.9 | 4.8 | 1.5 |
| Years employed at University of Iowa | 8.5 | 8.3 | 5.5 | 13.7 | 8.9 | 6.5 |
| Home-to-work distance | 3.8 | 3.4 | 7.8 | 9.6 | 2.5 | 3.2 |
| Home-to-bus-stop distance | 0.96 | 2.9 | 4.40 | 9.4 | 0.93 | 3.3 |
| Variability of work shift, 0 = no, 1 = yes | 0.57 | 0.5 | 0.45 | 0.5 | 0.28 | 0.4 |
| Importance ratings (20-point scale) | | | | | | |
| Travel time | 15.3 | 1.3 | 13.0 | 3.6 | 11.5 | 3.5 |
| Cost | 7.9 | 4.2 | 11.8 | 2.7 | 12.0 | 2.7 |
| Amenities | 7.9 | 5.5 | 4.6 | 3.9 | 1.1 | 0.5 |
| Convenience | 13.6 | 3.1 | 9.7 | 4.6 | 6.7 | 5.4 |
| Privacy | 12.6 | 2.9 | 12.6 | 3.3 | 11.1 | 5.4 |
| Energy conservation | 9.1 | 3.5 | 11.2 | 3.8 | 12.6 | 4.5 |
| Satisfaction with current mode | 17.5 | 4.3 | 14.9 | 4.7 | 17.4 | 3.2 |

slopes of the curves suggest that the manipulated factors had large and systematic effects on the car-bus preferences of the unbiased group. Thus, the approximately neutral mean response of this group reflects a balancing or trade-off of factors rather than a lack of responsiveness to the variations.

Finally, the plot for the bus-biased group (Figure 1c) reveals a small (but statistically significant) interaction between cost differences and time differences that is of opposite form to that observed for the car-biased group. Cost differences tended to have less effect at low time differences than at high time differences.

The figure also shows how responses to multiattribute systems can be used to define the functions relating subjective perceptions of mode attributes to objective magnitudes (Equation 1). Car-bus preferences change little as time difference increases from 0 to 15 min, but preference for the car increases greatly as time difference increases from 15 to 45 min. The psychophysical function for time difference is thus positively accelerated for each group.

The three groups identified by the cluster analysis differed in terms of the relative weighting of factors manipulated in the experiment. The car-biased and bus-biased groups differed considerably; the unbiased group showed some of the same effects as each of the other two groups, namely, a large time difference effect as the bus-biased group and a significant rider effect as the car-biased group. Only the bus-biased group was uninfluenced by the number of riders. The other two groups showed a decreased preference for the car as the number of riders increased.

At first thought, it might seem counterintuitive that respondents who preferred the money-saving mode (bus) would be heavily influenced by time factors and that respondents who preferred the time-saving mode (car) would be heavily influenced by cost factors. However, what this means is that degree of preference for the bus by respondents in the bus-biased group was influenced by the amount of extra time involved in taking the bus, and degree of preference for the car by respondents in the car-biased group was influenced by the amount of added cost involved in driving a car. Viewed in terms of these trade-off processes, the results make sense.

The next phase of data analysis examines the relationship between group differences found in phase 1 and differences in other factors measured in the experimental questionnaire.

## Phase 2

Table 1 summarizes results for various parts of the questionnaire when respondents were divided into the groups identified in phase 1. In terms of the socioeconomic variables, the key differences among groups were in terms of income, home to work distance, and work shift variability.

As might be expected, the car-biased group tended to have a higher income and greater variability in working hours. The most pronounced difference was in terms of home to work distance, where the unbiased group lived considerably farther from place of employment than either of the other two groups.

Ratings of importance of various factors differed among groups. As would be expected, amenities, convenience, and privacy were rated more important by car-biased respondents than by bus-biased respondents, while conserving energy was rated more important by bus-biased respondents. The high rating of privacy by the car-biased group is consistent with the large effect of number of riders observed for that group in phase 1. Travel time was rated more important by car-biased respondents, and cost was rated more important by bus-biased respondents. While this seems to be the opposite result of phase 1, these ratings were made in the abstract and did not actually involve trade-offs of specified levels of competing factors.

While car-biased respondents may presently choose car over bus because of time savings, this does not necessarily mean that they would be unresponsive to changes in cost factors. The functional measurement procedure used in phase 1 revealed trade-off relationships that operate in car-bus preferences and would seem to provide more information about decision processes underlying mode choice than would simple importance ratings. In particular, the nonlinear functions obtained in the trade-off analyses show that cost and time factors increase in importance when they reach extreme values.

Ratings for the unbiased group were generally intermediate to the other two groups. However, this group had lower ratings of satisfaction than did the others. This is consistent with responses to the open-ended question at the end of the questionnaire. Respondents in the unbiased group were most apt to indicate that they took their present mode only because it was the only one available. Results for this group suggest that it is composed of many respondents who are captives of their present mode but who would consider switching modes if transportation alternatives were offered.

A regression analysis was then conducted to quantitatively relate the grouping assignments to the following socioeconomic measures, demographic characteristics, and transportation constraints obtained in the questionnaire: home to work distance, home to bus stop distance, work time (day versus night), type of work shift (fixed versus variable), variability of work schedule, business and personal needs (ratings) for car, convenience of parking at work place (rating), work place (code), age, sex, and income.

The ability of a linear combination of these variables to predict group membership was measured by two statistics: the overall proportion of variance ($R^2$) of grouping explained by the linear combination and the proportion of cases correctly assigned to each bias group. The resulting overall regression was significant beyond the 0.01 level, but it explained only 27 percent (corresponding to $R = 0.52$) of the variance in grouping. This corresponded to 67 percent of cases being correctly assigned to bias groups in a discriminant analysis. While this result is disappointing from the point of view of identifying bias groups on an a priori basis, it was not unexpected, since the effect of socioeconomic variables on attitudes is most likely one that operates over time. One might, then, expect the cross-sectional correlation to be low.

### Phase 3

The next phase of analysis was the crucial one of relating responses in the experimental task to actual mode-choice behavior. Ideally, this would be done by taking a model of mode choice derived from hypothetical trade-offs for individuals and substituting in measures of the individuals' real-world transportation environment. Each of these one-point predictions would then be correlated with frequency of patronage of car and bus. In the present case, individual models were not available and a simplified alternative model was tested for the frequency of a given mode choice:

$$F_i = f \tag{5}$$

where f equals bias, home-to-work distance, most expensive mode, and bus availability.

Bias was measured by the individual's mean response over all cells in the experimental design. The second factor, home-to-work distance, was a surrogate for the time difference factor manipulated in the experiment. The third factor, most expensive mode, was a binary surrogate for the cost difference factor, because many respondents could not articulate their estimates of costs of various travel modes, but some respondents indicated that it was actually cheaper to drive a car than to take the bus. The last factor, bus availability, was inserted as a binary variable to reflect constraints on actual mode choice.

It might be helpful at this point to briefly discuss the substantive meaning of the model being tested. It is hypothesized that, in an experimental situation, individuals carry with them and reflect in their responses the factors and attitudes relevant to them in their transportation decisions. These affect their overall bias in the experimental task and their weighting of factors manipulated in the experiment. However, real-world constraints may be removed in the controlled experimental task to simplify analysis of the decision processes, such as by specifying that both car and bus are available in the present case. Thus, the present model for predicting actual mode choice incorporates a measure of response, surrogates for the variables manipulated, and a classification of real-world constraints into the experimental task.

The dependent variable, mode patronage, was measured in terms of the proportion of work trips by bus during the month prior to receipt of the questionnaire. As the questionnaire was administered during the summer, a number of respondents indicated that they either walked or rode a bike. However, in all but one of these cases, respondents added information about their nonsummer mode. This information was used to reassign nonmotor vehicle trips. The one exception was dropped from further analysis, as was one other respondent who indicated proportion of trips on the inner-campus bus system rather than home-to-work trips. Hence, the final sample size used for this analysis was 97.

Specifically, the following regression of the factors defined in Equation 5 was used to predict bus patronage:

$$\text{Prop}_B = [a(\text{mean}) + b(\text{HWD}) + c(\text{min})](\text{avail}) + e \tag{6}$$

where

$\text{Prop}_B$ = proportion of bus trips,
mean = mean response to experimental task (20-point scale),
HWD = home to work distance,
$\text{min} = \begin{cases} 1 \text{ if car rated cheaper than bus or} \\ 0 \text{ if otherwise, and} \end{cases}$
$\text{avail} = \begin{cases} 0 \text{ if no bus available or} \\ 1 \text{ if otherwise.} \end{cases}$

The availability factor was entered as a multiplier in the regression equation because, if it were at a 0 level, then frequency of bus patronage would be 0.

The prediction was good and explained over 78 percent (based on $R = 0.885$) of the variance in the proportion of bus trips for different respondents. An examination of residuals revealed some tendency for overprediction of low values and underprediction of high values, but overall the model appears to provide a reasonable description of the data, especially in light of the crudity of measurement of some of the predictor variables.

The single factor, bias (mean response on experimental task), accounted for over 70 percent of the variance in the number of bus trips. It is quite likely that, had actual travel time and cost difference measures been available, the overall predictive ability of the model would have been even higher.

In order for the predictive ability of the present model to be compared to more traditional models such as the logit, respondents were divided into two groups, car riders and bus riders, according to the most frequently used mode. The present model was then applied in a discriminant analysis to determine its classificatory ability. The result was that 94 percent of the cases were correctly classified. This result is comparable to those obtained in the more successful applications of logit-type analyses that have been reported (4).

For a further comparison with traditional methods, a linear combination of importance ratings of various factors measured in the questionnaire was used in place of the bias measure in Equation 7 to predict bus patronage. The following importance ratings were used: travel time, cost, amenities, convenience, privacy, and energy conservation. The resulting regression model was highly significant, but the predictive ability of 59 percent was clearly less than that obtained with the bias measure from the experimental task.

## EXPERIMENT 2

In an effort to provide further substantiation of the results obtained in the first experiment, a second questionnaire was designed and distributed to a random sample of 150 residents of Iowa City.

### Questions

The form was similar to the first, with two major modifications: (a) the experiment included the factors travel time difference, cost difference, and number of riders in a 3-×-3-×-3 factorial design where each evaluation made by the respondent was based on all three factors (as compared to two factors for each evaluation in experiment 1); and (b) direct estimates were obtained for each respondent of actual car-bus travel time and cost differences, as well as number of auto riders.

### Results

Of the original 150 questionnaires distributed, 72 were returned. As the focus of the investigation was on car-bus modal split for work trips, nonworkers and individuals living a few blocks from their work places who indicated walking as the primary mode were deleted from the analysis. This reduced the total sample size to 48.

Following the steps employed in the analysis of the first experiment, raw responses by each of the 48 individuals to the hypothetical mode-choice situations were first cluster analyzed. As in the first experiment, three salient clusters of respondents were identified: a car-biased group (N = 14), a bus-biased group (N = 17), and an unbiased group (N = 17). Responses within each of these groups were then subjected to analysis of variance.

Despite the small sample sizes, results bore a general resemblance to those obtained in the first experiment. For example, plots of the cost difference times time difference interaction revealed a noticeable convergence of responses toward the lower end of the scale (indicating car preference) for the car-biased group at a time difference of 45 min. This, again, suggests a nonadditive combination rule for time and cost differences for this group. The major dissimilarity between the two sets of results was an absence of response concentrations at extreme ends of the scale for the car- and bus-biased groups. This would appear to be related to the nature of the new design—the rider factor was considered simultaneously with time and cost differences. Hence, the observed effects of time and cost reflect the averaging of a third factor that has a moderating influence.

Actual mode-choice behavior was related to experimental response in a fashion similar to that of the first experiment. In the new analysis, however, estimates of time and cost differences as provided by respondents were available. The model tested, therefore, was

$$\text{Prop}_B = [a(\text{mean}) + b(\text{cos dif}) + c(\text{tim dif}) + d(\text{riders})] \times (\text{avail}) + \epsilon \tag{7}$$

where

mean = mean response to experimental task (20-point scale),
cos dif = estimate of actual car-bus cost difference,
tim dif = estimate of actual car-bus time difference,
riders = number of riders who share (or would share) a work trip, and
avail = $\begin{cases} 0 \text{ if no bus available or} \\ 1 \text{ if otherwise.} \end{cases}$

The level of prediction of the proportion of bus trips was similar to that reported for the first experiment. The resulting $R^2$ was 0.77 (based on R = 0.88), which corresponded to 95 percent of respondents being correctly classified into predominantly bus or predominantly car groups in a discriminant analysis.

## DISCUSSION

This paper has advanced and empirically assessed an alternate paradigm in the modeling of transportation mode choice. The approach departs from most traditional modeling paradigms in terms of its emphasis on deriving functional forms that best describe the processes by which individuals arrive at transportation-related judgments.

Results of the reported studies produced two findings with respect to the utility of behavioral models of mode choice. First, they showed that mode-choice models derived in earlier laboratory studies with student populations can be generalized to nonstudent populations. In fact, cluster analysis of behavioral data led to a more meaningful system of classifying respondents than would a priori population subdivisions. Second, they showed that the rating responses to hypothetical mode-choice trade-off situations are related to actual mode choices. Specifically, a model that combined responses to an experimental task with situational constraints yielded high explanatory ability in the prediction of actual mode choices. The levels of prediction obtained with the simple regression models compared favorably with reported successful applications of traditional stochastic mode-choice models.

A series of equations was outlined for developing a behavior-based theory of travel behavior. The experimental task of the present study directly examined Equations 1 and 2, which deal with subjective evaluations of mode attributes and the transformation and integration of subjective evaluations into an overall subjective response. Equations 3 and 4, which deal with the relationships linking objective and subjective attribute values to actual choice behavior, were examined, at least in a preliminary manner, in the current attempt to predict actual mode choice.

The establishment of stronger links to actual choice behavior awaits further study in which trade-off factors manipulated in hypothetical mode-choice situations can be measured accurately for each individual in the sample and decision models can be calibrated for individual consumers.

Once clear links are established between rating responses in abstract settings and actual mode-choice behavior, the stage will be set for the most useful application of the behavior-modeling approach. The derivation of decision-making models through carefully designed experimental tasks can be used to predict future responses to changes in transportation systems. Existing stochastic demand models seem theoretically weak for such purposes. Behavioral models can be used to directly estimate latent demand for alternative transportation opportunities. In the present study, for example, the mode-choice model without the availability constraint might be thought of as measuring such demand.

The relative degree of success of this study in predicting mode choice directly from attitudinal data is noteworthy. The functional measurement approach appears to be a means of side-stepping traditional issues related to the identification and measurement of all the relevant variables in mode-choice decision making. By combining experimental design and demographic analysis, a variety of variables affecting mode choice can be studied directly; other factors can also be shown to exert their influences through individual differences in response bias and the weighting of information.

The general paradigm proposed in this paper is one that seeks to advance our understanding of human travel behavior through a concerted program of laboratory experimentation and real-world verification. If laboratory behavior in simulated transportation environments can be shown to be predictive of that observed outside the laboratory, then such simulation would provide a powerful tool in the testing and development of theories of travel behavior.

This research provided some initial support for such a link. Nevertheless, the research program remains at an embryonic stage, and many basic questions related to the utility of behavioral models in the prediction of travel behavior remain unanswered. We hope the results of this research will encourage additional efforts in this direction.

REFERENCES

1. I. P. Levin. Information Integration in Transportation Decisions. In Human Judgment and Decision Processes in Applied Settings (M. F. and S. Schwartz, eds.), Academic Press, New York, 1977.
2. J. J. Louviere, L. M. Ostresh, D. Henley, and R. J. Meyer. Travel Demand Segmentation: Some Theoretical Considerations Related to Behavioral Modeling. In Behavioral Travel Demand Models (P. R. Stopher and A. H. Meyberg, eds.), Lexington Books, Lexington, MA, 1976.
3. P. R. Stopher and A. H. Meyburg. Urban Transportation Modeling and Planning. Lexington Books, Lexington, MA, 1975.
4. A. Talvitie. Comparison of Probabilistic Modal Choice Models: Estimation Methods and System Inputs. HRB, Highway Research Record 392, 1972, pp. 111-133.
5. M. G. Richards and M. Ben-Akiva. A Simultaneous Destination and Mode Choice Model for Shopping Trips. Transportation, Vol. 3, 1974, pp. 343-356.
6. G. C. Nicolaidis. Quantification of the Comfort Variable. Transportation Research, Vol. 9, 1975, pp. 55-66.
7. D. T. Hartgen. Attitudinal and Situational Variables Influencing Urban Mode Choice: Some Empirical Findings. Transportation, Vol. 3, 1974, pp. 377-392.
8. N. H. Anderson. Functional Measurement and Psychophysical Judgment. Psychological Review, Vol. 77, 1970, pp. 153-170.
9. D. A. Hensher. Perception and Commuter Modal Choice: A Hypothesis. Urban Studies, Vol. 12, 1975, pp. 101-104.

*Publication of this paper sponsored by Committee on Passenger Travel Demand Forecasting and Committee on Traveler Behavior and Values.*

# Incentives and Disincentives of Ride Sharing

Joseph B. Margolin and Marion Ruth Misch, Program of Policy Studies in Science and Technology, George Washington University, Washington, D.C.
Mark Stahr, Urban Planning Division, Federal Highway Administration, U.S. Department of Transportation

**This research examines consumer motivation concerning ride sharing, particularly carpooling, according to a market segmentation approach. A sequential design permitted (a) developing hypotheses about ride-sharing motivation based on qualitative data from intensive discussions in decision analysis panels, (b) testing those hypotheses by means of quantitative data obtained by survey, (c) developing program strategies on the basis of the results and pretesting those strategies with an additional series of decision analysis panels. The major market segmentation involved dividing the sample by commuting mode and pattern and by occupation type, although additional independent variables were also utilized. This paper concentrates on the carpooling attitudes and perceptions of carpoolers versus solo drivers. Illustrative findings are also presented by occupation group, commute pattern, and sex to illustrate the power of the finer market segmentation. The factors discussed include, first, attitudes toward costs or interpersonal aspects of carpooling (including match methods), time variables, carpool routes, parking management and convenience issues and, second, demographic characteristics of the two types of commuters. A special analysis focuses on the attitudes of those solo drivers who stated that**

**they were interested in carpooling versus those who stated that they were not. The purpose is to highlight the motivation of a prime target group toward three carpool strategy areas: carpool match methods, parking management, and dedicated carpool routes. Notable findings include the limited appeal of external efficiency factors such as cost savings, the power of social aspects of carpooling that can act as either barriers or incentives, and the need for personalized carpool programs that also reach out to actively involve the potential pooler. Specific program strategies are offered.**

Despite some examples of successful ride-sharing programs, neither carpooling nor other forms of ride sharing have yet become widely popular. This study, which focused on social as well as economic aspects, provides evidence that a significantly larger portion of the population can be induced to share rides if attention is paid to the different needs, perceptions, lifestyles, resources, and values of various market segments within the population.

The work has had three objectives: (a) to develop and refine methods for assessing the effects of given psychological, social, and economic factors on mode choice, (b) to discover and explain ways in which these factors facilitate or inhibit carpooling, and (c) to devise carpool-promoting strategies based on traveler needs and attitudes and to obtain a traveler-based pretest of the effectiveness of these strategies.

The investigation differs from previous studies, therefore, in terms of the scope of its objectives. It also differs in methodology, which, although it has been reported in detail previously (1), will be reviewed here briefly to provide context for the findings that follow.

## METHODOLOGY

The special problem addressed in the study was the lack of sophisticated data about consumer attitudes. Little systematic, in-depth information existed about which specific and different behavioral incentives and disincentives affect different groups of transportation consumers, and how these might affect ride sharing. A sequential in-depth study design was therefore devised and carried out in the Washington, D.C., metropolitan area.

### Phase 1

The first step involved going to the consumer to listen to his or her preferences, complaints, problems, and the trade-offs he or she might or might not be willing to make in choosing to carpool. This was done in a series of open-ended discussions with small consumer groups (5-9 people), or decision analysis panels. The methodology requires considerable expertise in managing and analyzing group dynamic process and does not focus discussion or limit findings to a predetermined set of issues. It provided a hypothesis-generating phase for the study and yielded qualitative data about the perceptions, attitudes, and transportation behavior of carefully segmented groups of consumers. The 21 decision analysis panels were selected according to age, sex, occupation, commuting pattern, and other lifestyle variables. The findings were used to build a survey questionnaire for the second phase of the study.

### Phase 2

In this hypothesis-testing phase, traveler attitudes discovered in phase 1 were quantified. Five hundred and sixteen people were individually interviewed according to a stratified sampling process that provided the distribution shown in Table 1, where the blue- and white-collar and managerial/executive/professional (MEP) workers are divided equally into three groups (33.3 percent each). Data were analyzed by multivariate analysis of variance in order to assess the significance of differences between groups on intercorrelated measures (2).

### Phase 3

The quantified results were then utilized to build two types of program strategies: (a) those having broad appeal and (b) those having special strategies aimed at particular market segments of the population. Finally, these strategies were pretested with 25 new decision analysis panels.

The selection of commuters for this last series of panels included a disaggregation by occupation and commuting pattern. However, their current mode choice differed from the survey. Four panels were made up of carpoolers, two of carpool rejectors (defined as people who had had an opportunity to carpool but who did not take it), and six of groups with mixed commuting modes.

## MOTIVATIONAL COMPARISON OF SOLO DRIVERS AND CARPOOLERS

### Demography of Carpoolers versus Solo Drivers

Carpoolers tended to be older males with regular work hours from households where more than one adult was employed full time. They commuted a greater distance than the solo drivers, but total trip time averaged nearly the same, about half an hour. This was not necessarily due to greater travel speed.

Panel data had indicated a considerable emphasis by carpoolers on efficient driving and route selection. Women tended to drive alone more than men. The solo drivers tended to be younger, and from lower income households than carpoolers. As expected, more solo drivers worked at jobs with shift changes or with officially irregular work hours.

So far, this picture is much like that of other studies. A new finding in our sample was that the delays of carpool matching services produced a significant negative effect. Among those who had participated in a carpool campaign, the solo drivers reported having had to wait longer for a response than carpoolers.

In general, the most surprising finding was not in the ways that the two kinds of commuters differed, but in the similarity of their commuting resources. They both tended to show the same ratio of licensed drivers to cars in the household and availability of alternative ways to commute. On the average, both had lived at the same address and worked at the same address for the same length of time. Finally, less than a third of either type of commuter could say that they had ever been exposed in the past to a carpool campaign.

### Attitudinal Factors Affecting Mode Choice

#### Cost

Cost is often considered the most important appeal of carpool programs. This is understandable. It is clear that carpooling does cut commuting costs; carpoolers are often eager to talk about the money they save; carpool campaigns emphasize the bargain. Even the solo drivers in our sample perceived carpooling as a financial gain; in fact, they considered it a better bargain than the carpoolers did. However, this research did not bear out the importance of cost as the most influential factor affecting the decision to carpool. This is particularly true for a large part of the solo driving population, with whom we are primarily concerned.

We asked respondents a series of trade-off questions to rate the cost savings of carpooling against such drawbacks as the time it takes to pick up members, not being free to run errands at will on the way home, etc. Results are presented in Table 2 ($p = 0.001$).

Depending on the trade-off, approximately 67-85 percent of the carpoolers were in favor of the cost savings, as expected. The solo drivers were quite different. While half of them did agree that "the fact that carpooling is less expensive than driving alone to work makes it well worth the effort," even this level dropped sharply with the mention of specific drawbacks. Only 38 percent favored the cost saving over the time it takes to pick up members, or over the desire to run errands on the way home. Only 34 percent thought the saving worth having to depend on other people, and a mere 23 percent considered that the money compensated for waiting for

**Table 1. Phase 2 survey sample design.**

| | Mode | | | | | |
|---|---|---|---|---|---|---|
| | Carpoolers (49.6%) | | | Solo Drivers (50.4%) | | |
| Commuting Pattern | Blue Collar (%) | White Collar (%) | MEP (%) | Blue Collar (%) | White Collar (%) | MEP (%) |
| Suburb to central business district (33.3%) | 5.5+ | 5.5+ | 5.5+ | 5.5+ | 5.5+ | 5.5+ |
| Suburb to congested suburb (66.7%) | 11.3- | 11.3- | 11.3- | 11.3- | 11.3- | 11.3- |

**Table 2. Cost savings opinions.**

| | Travelers Favoring Cost Savings | | | | | |
|---|---|---|---|---|---|---|
| | Total Sample | | Solo Drivers | | Carpoolers | |
| Opinion | No. | % | No. | % | No. | % |
| Carpooling is cheaper and worth the effort (agreement) | 515 | 68.0 | 257 | 50.6 | 254 | 85.8 |
| Insurance problems are a drawback (disagreement) | 514 | 56.6 | 258 | 44.5 | 252 | 69.0 |
| Pickup time is not worth the saving (disagreement) | 512 | 54.9 | 258 | 37.6 | 250 | 74.8 |
| Errands are more important than the saving (disagreement) | 512 | 53.5 | 257 | 37.8 | 251 | 70.1 |
| Depending on others is not worth the saving (disagreement) | 513 | 51.9 | 257 | 34.2 | 252 | 69.4 |
| Carpooling is cheaper but not worth waiting for late members (disagreement) | 514 | 46.7 | 258 | 22.7 | 252 | 66.6 |

**Table 3. Cost savings and shifting to carpooling.**

| | Response | | |
|---|---|---|---|
| Question Choices | No. | % | Difference |
| How likely would you be to carpool if that way your share of the parking costs were | | | |
| ½ of what you now pay | 191 | 46.3 | 6.6 (NS) |
| ¼ of what you now pay | 191 | 45.7 | |
| Free | 190 | 52.9 | |
| Would you change to carpooling or the bus if your parking costs increased over what you pay now by | | | |
| \$5/month | 235 | 28.8 | 38.5 ($p = 0.001$) |
| \$10/month | 234 | 43.5 | |
| \$20/month | 234 | 67.3 | |
| Would you shift to carpooling or the bus if the price of gas went to | | | |
| \$0.90/gal | 257 | 56.7 | 20.6 ($p = 0.001$) |
| \$1.30/gal | 255 | 71.0 | |
| \$2.00/gal | 254 | 77.3 | |

**Table 4. Results of multivariate analysis of agreement on the significance of social factors in carpooling.**

| | | Mode | | | Occupation | | | | Combinations of Mode, Occupation, and Work Site | | | Mode × Sex | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | Mode × Site | | | Solo Driver | | | |
| Opinion Statement | Total Sample | p | Solo Driver (%) | Car-pooler (%) | p | Blue Collar (%) | White Collar (%) | MEP (%) | p | CBD (%) | Suburbs (%) | p | Men (%) | Women (%) | Car-poolers (p) |
| Socializing is pleasant | 73.2 | <0.04 | 68.2 | 78.3 | | 82.2 | 73.7 | 63.2 | NS | | | <0.04 | 70.9 | 63.7 | NS |
| Talking shop is pleasant | 52.0 | <0.04 | 38.9 | 65.2 | NS | | | | NS | | | NS | | | |
| Carpooling is worth personal disagreements | 32.0 | <0.04 | 23.8 | 40.5 | <0.04 | 42.3 | 28.8 | 25.3 | | 18.7 | 26.0 | <0.004 | 28.4 | 24.1 | NS |
| Dislike smoking | 46.7 | NS | | | <0.003 | 39.0 | 45.3 | 55.5 | NS | | | NS | | | |
| Dislike rule making | 61.0 | <0.04 | 64.3 | 57.3 | | 76.3 | 60.1 | 47.0 | NS | | | NS | | | |

**Table 5. Results of multivariate analysis of agreement on the significance of choosing carpoolers.**

| | | | Occupation | | | | Sex | | |
|---|---|---|---|---|---|---|---|---|---|
| Opinion Statement | Total Sample (%) | Modes Combined (p) | p | Blue Collar (%) (N = 169) | White Collar (%) (N = 179) | MEP (%) (N = 166) | Men (%) (N = 321) | Women (%) (N = 192) | Sex Within Mode Combined[a] (p) |
| If I were joining a carpool now, it would be important to be able to meet the people at least once before making definite arrangements | 86.8 | NS | NS | | | | NT[b] | NT[b] | NS |
| I would only carpool with people I already know | 38.7 | NS | <0.01 | 42.2 | 42.2 | 32.0 | 36.8[c] | 42.5[c] | NS |

[a] Numbers are 150 men and 103 women solo drivers and 168 men and 89 women carpoolers.
[b] No test of statistical significance is yet available; differences do not appear significant.
[c] Although no test of statistical significance is yet available, differences appear to be significant.

**Table 6. Results of multivariate analysis of agreement on the significance of carpool match system preferences.**

| | | Mode | | | Occupation | | | |
|---|---|---|---|---|---|---|---|---|
| Opinion Preference | Total Sample (%) | p | Solo Drivers (%) | Carpoolers (%) | p | Blue Collar (%) | White Collar (%) | MEP (%) |
| Be contacted by a carpooler | 71.4 | NS | | | <0.02 | 72.8 | 68.0 | 73.0 |
| Helped by a neighborhood coordinator | 68.8 | NS | | | NS | | | |
| Have no help | 68.2 | <0.04 | 64.8 | 71.7 | <0.01 | 76.9 | 62.9 | 64.6 |
| Use a carpool locator list | 56.3 | NS | | | <0.04 | 57.4 | 49.4 | 63.5 |
| Computerized system | 53.1 | NS | | | <0.01 | 70.6 | 64.8 | 64.8 |

late carpool members. Before leaving this chart, we should note that even the carpooler's enthusiasm about cost as the prime factor cooled under the pressure of trade-off conditions.

However, the disincentive of greatly increased commuting costs for solo drivers had more effect. As Table 3 indicates, substantial numbers of solo drivers (67 percent) said that they might switch from solo driving if parking costs increased by \$20 a month, and 77 percent said the same should gas prices rise to \$2.00/gal.

It is important to note here that such predictions of future behavior change cannot be taken literally. People consistently overpredict their future behavior in the face of a hypothetical change of nearly any type. Despite this, these solo drivers did not predict a change in their commuting method for the reward of lower parking costs. Also, much less prediction of change occurred at lower levels of increased parking and gas costs—if gas went to \$0.90 or \$1.10/gal, for instance. During the decision analysis panels, cost had emerged as a more significant factor for blue-collar workers than for the other groups. The survey confirmed this. Among solo drivers, the blue-collar workers (regardless of income level) were more concerned with costs and more willing to put up with possible carpool problems in order to save money than were other occupation groups. With this exception, however, the fact must be faced that, in this largely affluent society, cost savings alone cannot be relied on to make most people carpool. To confirm the finding, when people were asked what their major commuting problem is, only 5 percent mentioned any kind of economic factor.

## Interpersonal Aspects of Carpooling

If the most commonly emphasized factor, cost, emerged to be of relatively low concern to many solo drivers, the most neglected aspect of carpooling programs was found to be the most important: the personal and social.

Helping a neighbor or having company were common motivations to carpool. Disagreements or personal incompatibilities frequently caused carpools to break up. The very situation of being in a carpool is perceived as a combined business and personal arrangement, and one for which we do not have well-established social customs. Both carpoolers and solo drivers found the socializing that a carpool offers to be pleasant, but the solo drivers had more misgivings about handling specific problems. Because 65 percent of these solo drivers had been in a carpool at some time in the past, their misgivings cannot simply be due to lack of experience.

The results for a sampling of 516 people are presented in Table 4 ($p = 0.001$). Only 39 percent of the solo drivers thought talking shop during the ride would be pleasant, versus 66 percent of the carpoolers. Even fewer, only 24 percent, could say that the chance to socialize was worth personal disagreements that might have to be ironed out (versus 41 percent of the carpoolers), and a strong 64 percent disliked the idea of rule making for the carpool (versus 57 percent of the carpoolers). The results for work site and sex were not significant.

As an illustration of the power of analyzing these data by more finely drawn groups, among solo drivers, women were a little more wary than men about how pleasant the socializing might be and about having to handle disagreements. Solo drivers who work in the suburbs are somewhat more willing to take on the disagreements than those commuting into town. There may be an urban "hassle factor" that adds tension to the trip.

## Choosing Carpoolers

Choosing carpool mates emerged as an extremely personal matter. Both types of commuters tended to react the same way, as can be seen in Table 5. Some 87 percent wanted to meet prospective members once before making any arrangements, and 39 percent felt they would actually have to know the people first, a concern felt by more blue- and white-collar workers than MEPs.

Carpoolers were more relaxed than solo drivers about riding with people who do not work at similar jobs, but this area is a sensitive one, because people tend to be reluctant to admit what might seem to be a prejudice. On the whole, male solo drivers were most vocal about wanting to ride with people at similar job levels; 43 percent of the men versus 19 percent of the women solo drivers expressed this. There were, however, no significant differences by occupation group.

When 516 respondents were also asked to rate different matching systems for carpooling, both carpoolers and solo drivers rank ordered the match methods in the same way (Table 6), favoring personal methods most. It is evident that the two most common systems used—locator lists and computerized match systems—have the least appeal. Interest in (a) being contacted by a carpooler, (b) having the help of a neighborhood coordinator, or (c) having no help at all is significantly greater (by chi-square analysis) than the response to either locator lists or computer matches. Expressed preferences for the latter two were not significantly different from chance. This is true at least as these methods now operate, that is, impersonally. The strongest preference was to be contacted by another commuter forming a pool (71 percent of the total sample); the next strongest was to be helped by a local neighborhood carpool coordinator (69 percent); and the third choice was to have no help at all, which was the only match method where solo drivers and carpoolers differed. The solo drivers felt a little less confident about handling things alone; only 65 percent of them endorsed this, versus 72 percent of the carpoolers.

It is clear that people do not want to become involved with others they know nothing about. For several population segments, it does not appear from the decision analysis panels to be the computerized matching in itself so much as the fact that the system does not include a person who could be contacted in the neighborhood or at work and who would know something about the specific individuals on the matching list or could offer advice about handling the combined personal and business situation of a carpool. However, there are also privacy issues arising from computerization that have been dealt with in the full report (3).

## Commuting Time and Carpools

Time was of great importance to all commuters interviewed, but the big barrier carpooling time represents to solo drivers appears to be the perception that carpools make one late. Both types of commuters were actually getting to work in about the same amount of trip time. When the misperceptions about carpooling trip time were analyzed, one after another related to uncertainty about possible trip delays, rather than speed of travel. Carpooling is seen as fraught with potential time problems that are unnecessary if there are clear rules about how long to wait for late members, what to do when a driver cannot meet the pool, etc.

## Carpool Lanes and Parking Management

These two important issues are presented together, because what the commuters in this study want is a smooth, hassle-free trip to work—a clear and open road and a place to park at the end of it.

The single greatest complaint, and the only one initiated by a significant portion of the people sampled, was traffic congestion, the primary problem to 37 percent of those interviewed. Dedicated roadways open to multi-occupancy vehicles (vanpools, carpools, and buses) received a resounding endorsement by solo drivers, provided that the roadway is available for significant portions of the trip (see the table below).

| Express Lane Proportion of Trip | Solo Drivers Who Would Carpool if Offered an Express Lane: Unlikely to Carpool (%) | Likely to Carpool (%) |
|---|---|---|
| One-quarter | 51 | 37.1 |
| One-half | 41.3 | 47.0 |
| Three-quarters | 33.9 | 57.9 |

From the table, only 37 percent predicted they would carpool if a lane were available for a quarter of the trip, while 58 percent would do so for a lane available for three-fourths of the trip. Where dedicated routes were rated against carpool drawbacks, the pervasive unwillingness of the solo driver to leave work at a fixed time each day was much less powerful; 56 percent of the solo drivers still rated the lane positively. It must be noted, however, that those commuting to the urban downtown still remained reluctant indeed to leave work at a fixed time each day, even for the incentive of a dedicated lane for "much of" the trip. Only 21 percent agreed. This probably relates to the congested trip time left before they get to their destination. The difference may suggest that dedicated lanes on highways encircling cities may be even more powerful incentives than those on highways radiating from them.

The strongest disincentive for using the dedicated routes was the need to depend on others, although even here 48 percent of the solo drivers thought the lane worth it. The finding may relate to the experience of the Washington, D.C., sample, who tend to think of carpool lanes in terms of the stringent requirement for four riders on the local Shirley Highway lane; that may be too many people to wait for.

Parking at work was, of course, extremely important to both types of commuters. The effectiveness of possible employer bans on parking for solo occupant autos was widely admitted, but private employers interviewed in panels considered it impossible to implement because of the risk of losing the competition to hire and retain quality employees.

People in the sample took parking for granted. Over 70 percent parked free in employer lots. Few turned to commercial lots. Few worked at organizations where any incentive for carpooling was offered—either guaranteed parking, cheaper parking, or parking closer to the work entrance.

In general, the solo drivers were very reluctant to accept carpooling, even if that were the only way to get a guaranteed parking space at work, when they considered it in relation to having to leave work at a fixed time each day (the single greatest time barrier to carpooling).

However, the disincentives of the alternatives to guaranteed parking were more powerful, as shown in the following table. Only 21 percent of the solo drivers would be willing to hunt for a parking space for 20 min and only 10 percent for half an hour. A mere 6 percent would risk a ticket five times a month, and only 2 percent would say they would run a daily risk of getting a ticket in order to park at work.

| Disincentive Question for Solo Drivers | Sample Size | Percentage Agreeing |
|---|---|---|
| To continue driving alone to work, would you be willing to look for a parking space, if you had to, for | 251 | |
| 10 min | | 45.0 |
| 20 min | | 20.8 |
| 30 min | | 10.4 |
| To continue driving alone to work, would you be willing to risk getting a parking ticket | 246 | |
| Once a month | | 31.7 |
| 5 times a month | | 5.7 |
| Every day | | 2.0 |

### Things Called "Convenience"

It was necessary to ask highly specific questions under the heading of "convenience," because the panels had made clear that it has multiple meanings and tends to be used as shorthand or euphemism for any difficulty that bothers a commuter.

Solo drivers were altogether more concerned about convenience factors than carpoolers. A large majority—76 percent—thought rearranging and adjusting schedules to suit the needs of other people a serious problem in carpooling, and 51 percent felt that there would not be enough room to carry and store packages. This concerns the tool-carrying blue-collar and errand-running white-collar workers more than the MEPs. When carpoolers and the 65 percent of solo drivers who had carpooled in the past were asked why they had carpooled, 22 percent of the solo drivers versus only 14 percent of the carpoolers mentioned convenience factors.

Solo drivers but not carpoolers perceived carpools as physically crowded. They thought the carpool ride less relaxing than carpoolers, and they worried about not being able to start the trip when they wanted to. They hesitated more than carpoolers in feeling they could rely on a carpool or have door-to-door service.

## MOTIVATING THE SOLO DRIVER TO CARPOOL

The purpose of this section is to highlight the differences between those solo drivers who expressed some interest in carpooling and those who stated they had none. When solo drivers were asked "Are you interested in carpooling to work?" 17 percent responded "Yes, definitely"; 23 percent said "Yes, possibly"; 16 percent replied "Not sure"; and 43 percent said "No." With such a wide variation of interest, one high priority strategy would be to tailor carpool programs to the interested solo drivers. A special analysis was therefore performed concerning solo drivers' attitudes toward three important strategy or policy areas: carpool matching system, parking management, and dedicated carpool routes.

A discriminant function procedure was chosen, since this technique highlights the differences between groups (4). It has been used successfully in transportation research to predict mode choice from beliefs about buses, carpools, and single-occupant autos (5). Solo drivers were classified according to their interest in carpooling as a function of their attitudes toward the three policy areas. A high colinearity that would distort the results emerged among some variables in each area, so variables were first combined by classical factor analysis (6) into composite variables.

The solo drivers were grouped into two categories: those 40 percent interested in carpooling (definitely or possibly) and those 43 percent definitely not interested. The 16 percent who answered "not sure" were excluded from analysis. The range of responses indicated that a statistically necessary assumption that this group was in between the interested and not interested groups could not be made. The unsures, of course, are important to consider in future analyses since many might carpool if their particular needs were met.

### Carpool Matching

The three composite variables were "advance knowledge of carpool mates," "no assistance desired in forming carpools," and "assistance desired." By far the most powerful discriminator was attitudes toward having assistance. Those interested in carpooling desired help—and by personal more than impersonal match systems—although they tended to favor any help (discriminant function coefficient of 1.08 versus 0.20 and 0.21 on "advance knowledge" and "no assistance").

Those not interested in carpooling were far less positive about assistance, should they at some time want to become involved in a pool. Responses to the questions relating to advance knowledge of carpool mates did not delineate the two groups; all solo drivers analyzed wanted to meet carpoolers in advance, although all did not necessarily require people they already knew or job peers in a carpool.

The "no assistance" variable also failed to discriminate. It was a split composite. All solo groups favored forming carpools without help, and all rated computer matching negatively.

### Parking Management

Five composite variables involved either the incentives for carpooling or the disincentives for solo driving that follow.

Incentives for Carpooling
- Carpooling favored for cost savings and convenience
- Guaranteed parking worth carpooling time concessions

Disincentives for Solo Driving
- Carpooling favored over paying parking cost increases
- Solo driving favored despite risk of parking tickets
- Solo driving favored despite parking space hunt

The carpooling incentives distinguished the two groups (coefficients of 0.51 and 0.58, respectively), which appealed to those interested in carpooling. The disincentives did not; all found them onerous.

### Carpool Lanes and Roads

Two composite variables emerged: the "use of express lanes and roads considered in isolation" and "express routes versus dependence on others." The first was an excellent discriminator (discriminant function 0.86), which indicates that express routes considered by themselves appeal strongly to solo drivers interested in carpooling. However, the trading off of the use of such routes versus being tied to a fixed departure time and depending on others discriminated poorly (coefficient 0.21). Concerns about departure times and about relying on others during one's commute appeared pervasive among all solo drivers.

### Combined Variables

One last discriminant analysis combined all composite variables from all three policy areas. The two best discriminators were

1. Guaranteed parking for carpoolers being worth fixed departure times, extra commuting time, and waiting for late members (0.41) and
2. Likelihood of carpooling if an express lane or road were available for a sizeable portion of the trip (0.39).

Those interested in carpooling responded to these incentives and were best distinguished from those not interested in carpooling by their responsiveness.

Summary

Clearly, in aiming programs at the most likely to carpool solo drivers, the incentives of guaranteed parking—where parking is otherwise a problem—and express routes will tend to be high-payoff strategies, as will assistance with personal match methods. It will also, of course, be vital for transportation planners to consider, in addition, the social dynamics and other elements discussed in previous sections.

## RECOMMENDATIONS

The recommendations offered are based on the finding from panels that privately owned auto transportation is likely to continue to dominate personal travel in the foreseeable future. Carpooling can be a viable private means of dealing with transportation and energy problems, but, if it is to succeed, a long-range, well-planned, high-priority effort is required. In addition, a blend of economic and behavioral incentives will be required. Any effort to replace a highly valued activity has to employ equally powerful motivations. Disincentives or purely economic incentives are insufficient in a nation of affluent and independent individuals who already have investments in the automobile.

Several demographic characteristics have been confirmed as predisposing individuals to carpooling and so defining some preferred target populations:

1. Those with long commutes,
2. Drivers over 30 years old,
3. Commuters living in multiperson households,
4. Those with regular hours in the case of MEP and white-collar commuters to the central business district and blue-collar workers to congested suburban sites, and
5. Availability of parking incentives.

However, the strategies that follow are recommended for more resistant groups. Each has value with a broad range of population segments; each may also provide different incentives to different subgroups.

In order to address the needs of different transportation program personnel, the full report deals with all the critical variables found, such as parking, cost, social factors, and convenience, and relates strategies concerning each to the needs of each specific population segment. We will abbreviate the process here by describing the major recommendations and how they would motivate the particular subgroups to share rides. These recommendations emerge from an integration of findings from the decision analysis panels of phases 1 and 3 and from the survey.

### Strategy 1: A Personalized System

The most important recommendation is that the carpooling formation system be personalized and that it offer a quick response and an active outreach program. Passive systems, such as lobby locator lists, place the burden of understanding the system on the individual. They require the highly motivated to go through the various steps and to overcome obstacles required to form a carpool and to have the perseverance, understanding, and resources to achieve and maintain a compatible carpool. For example, there are large segments of the population who are interested in carpooling but who hesitate to telephone a stranger. The reasons range from sheer timidity to fear of becoming caught in an unpleasant situation that one could not comfortably terminate. This is a socially awkward situation for both women and men.

### Strategy 2: Appeal to Population Subsegments

It is critical that a carpool program concentrate on appeals that are appropriate to particular subsegments of the population. We have described how motivation for and problems with carpooling vary according to occupation type and conditions, experience with carpooling, size of employment site, age, income, sex, length of commute, and so on. Therefore, it is vital that the program planners know and be governed by the characteristics of the people in the community as well as its geography.

### Strategy 3: Local Carpool Coordinator

One method for achieving these criteria is the use of the local carpool coordinator. Located at the work site or at the home end, whichever provides the best leverage, the local coordinator would serve the functions of

1. Learning about his or her population's needs and nature,
2. Providing information about carpooling,
3. Initiating and coordinating personalizing strategies,
4. Assisting in forming new carpools via various substrategies including bringing people together or employing existing groups,
5. Assisting in enlarging current carpools,
6. Providing early warning of trouble and helping to deal with problems in existing carpools, and
7. Providing emergency service when carpools break down temporarily.

The coordinators may eventually move into another transportation energy-saving activity, such as buspools to athletic or other events.

It is important that the local coordinator understand the factors that motivate people to carpool and that contribute to a good carpool. He or she must be taught that no single factor is preeminent but that an introduction of multiple factors such as distance, time, and especially social and personal dynamic factors produce an effective program. It is also important that the local coordinator become a resource to the community rather than a burden. He or she must also learn to enlist existing enthusiasts or community leaders to achieve his or her objective rather than do things alone. In doing so, he or she obtains community consensus and pressure in behalf of ride sharing. A handbook for local coordinators would be needed.

A local coordinator is someone at the work site who relates well to people and has their confidence and can handle the concepts described. The benefit to the employer of assigning a good person will derive from increased promptness, morale, and productivity (interest in workers by management), and possibly from land economics as less is needed for parking.

At the home end, local coordinators may be retired or disabled persons with the need to continue productivity, housewives with limited mobility during the day but time to use the phone, or civic leaders with a desire to make a contribution and time to do so.

The local coordinator can make use of computer printouts where computerized matching systems exist. These can be strengthened through his or her capacity for personalizing the match and taking the initiative.

He or she can arrange special carpools, such as no smoking, "silent in the morning," all one sex, or whatever is required by the constituency.

He or she can eliminate much of the hassle and doubt that interferes with carpool formation and can help solve the problems that result in carpools breaking up.

The local coordinator's strategy would be directed to all potential carpool targets. However, it would be most effective with those populations where hesitancy, need for guidance, and assistance with problem solving are important factors—in short, for the marginal carpoolers who constitute a significant part of the population. Further, the presence of such an energy-related agent in the community will provide direct evidence of the reality of the national program.

### Strategy 4: Preparation of a Handbook

Strategy 4 involves the preparation of a "how to carpool" handbook to be used as part of the enrollment campaign and to contribute to a higher level of carpool continuation. This concept was very strongly endorsed by all phase 3 panels and would provide the information about carpooling, such as how to deal with problems and gripes, that this research and other sources have determined to be important. Once again, the target population for this strategy is broad. It includes those who might be resistant to or puzzled by some aspects of the carpooling situation that they do not understand or have misconceptions about.

### Strategy 5: Parking Strategies

Parking is not a problem for a large part of the population examined. However, effective strategies can be addressed to blue-collar workers who have difficulty finding parking spaces and are least able to risk parking tickets. Here parking for carpools will have value. For white-collar populations (largely women in this sample), the avoidance of frightening walks on dark winter evenings may motivate carpool formations. Close-in, covered, guaranteed parking is also likely to motivate the status-conscious executive to carpool.

### Strategy 6: Personal Safety

The survey and decision analysis panels suggest that blue-collar workers and female white-collar workers were both concerned about waiting for carpools at the work end on streets or roads that were often unsafe. In other cases, such rendezvous points as fringe parking lots were frightening. Secure meeting points will be valuable.

### Strategy 7: Cost Incentives

This is not the critical variable it was assumed to be, but it remains important to some segments of the population. Carpooling as a money saver will appeal to the lowest end of the income distribution. However, these are largely blue-collar and low-level white-collar workers for whom carpooling also has a special drawback: they are frequently docked in pay or may soon lose their jobs if they are late. Special adjustments to meet this problem will have to be made in cooperation with employers.

### Strategy 8: Familiarization Methods

In conjunction with the activities of the local coordinator, neighborhood or work meetings before forming carpools are particularly important to blue- and white-collar workers. This is particularly true of women, who want to know more about potential carpool mates before committing themselves.

### Strategy 9: Carpool Lanes as Powerful Incentives

Despite the fact that the MEP group was the least responsive of the three occupation types, this incentive is strong enough to lure many managerial personnel away from their "unpredictable hours" excuse for not carpooling. Blue-collar workers are the strongest advocates of the carpool lane, probably because of their fear of being made late by congestion with subsequent loss of wages. There is reason to believe that carpool lanes on ring roads would be even more powerful incentives and would address the majority of commuters.

## OCCUPATION GROUPS

Blue-collar workers have high potential for ride sharing, with sociability factors at the fore. They have little need to achieve mastery by controlling their own transportation mode. There are two large subgroups, one concerned with cost and another involved in long commutes. Particular care is needed in legitimizing the setting of rules for a carpool. Because blue-collar workers tend to drive the more unreliable, older autos, the availability of standby cars via the carpool becomes an asset.

White-collar workers are a high potential group for carpooling but have many special needs. These include knowing a great deal in advance about whom one may carpool with and arrangements for shopping at lunch hour and on the way home. Particular care must be taken in making it legitimate to set rules for a carpool. White-collar workers will respond well to outreach programs, to meetings where difficulties can be aired, and, for the large proportion who are women, to the opportunity to talk with women currently carpooling in order to understand in what ways it can be workable.

MEPs form a high dominance group with many status and mastery needs, and desires for flexibility in terms of departure time at the end of the work day. They would be responsive as a group to carpools with staggered hours, to the opportunity to be part of an advisory board that sets up the carpooling program at a work site or in a neighborhood civic association, and to carpools limited to medium-sized and larger cars.

## ACKNOWLEDGMENTS

The research reported in this paper was supported by the Federal Highway Administration of the U.S. Department of Transportation. The findings and conclusions are ours and are not necessarily endorsed by the Federal Highway Administration. We are indebted to the following people, who provided assistance either in the behavioral investigation or with the preparation of the current paper: Ricardo Dobson, the original project monitor, who made conceptual contributions to the research beyond the usual input of a monitor, and Carl Shea, Bruce Spear, and Mary Lynn Tischer for their assistance with the special analysis of solo drivers according to their interest in carpooling.

## REFERENCES

1. J. Margolin, M. Misch, and R. Dobson. Incentives and Disincentives to Ridesharing Behavior: A Progress Report. Paper presented at the 55th Annual Meeting of the Transportation Research Board, Jan. 22, 1976.

2. D. Morrison. Multivariate Statistical Methods. McGraw-Hill, New York, 2nd ed., 1976.
3. J. Margolin and M. Misch. Incentives and Disincentives for Ridesharing: A Behavioral Study. Federal Highway Administration, U.S. Department of Transportation, March 1977.
4. N. Nie, C. Hull, J. Jenkins, K. Steinbrenner, and D. Bent. Statistical Package for the Social Sciences. McGraw-Hill, New York, 2nd ed., 1975, pp. 434-467.
5. R. Dobson and M. L. Tischer. Beliefs About Buses, Carpools, and Single Occupant Autos: A Market Segmentation Approach. Transportation Research Forum Proceedings, Vol. 17: Indians, 1976, pp. 200-209.
6. H. Herman. Modern Factor Analysis. Univ. of Chicago Press, Chicago, 1967.

*Publication of this paper sponsored by Committee on Passenger Travel Demand Forecasting and Committee on Traveler Behavior and Values.*

# Attribute Importance in Multiattribute Transportation Decisions

Michael A. Johnson, Institute of Transportation Studies, University of California, Berkeley

**This report describes a study of the relative importance of various travel attributes as influences on commuters' choices among car, bus, and Bay Area Rapid Transit (BART) for traveling to work in the San Francisco Bay area. A sample of commuters were interviewed, and each was asked to rate his or her satisfaction with car, bus, and BART on each of the attributes studied. The relative importance of the attributes was inferred by examining these ratings and the relationships between the ratings and the usual choice of travel mode. The study differed from previous similar research in that attribute importance was measured with a statistic that estimated how much each attribute contributed to differences in utility among the choice alternatives. Most previous research failed to consider an essential component of the quantity measured by this statistic, namely, average differences in utility among alternatives caused by average differences among alternatives in the levels of each attribute. Among the attributes judged to be most important were safety from crime, seat availability, and dependable arrival, which are ordinarily not included in quantitative planning procedures such as travel demand forecasting and cost-benefit analysis.**

To a large extent, the experience of urban travel by any method can be described in the abstract as a composite of varying travel attributes. This paper describes a study of ten different travel attributes and their relative importance as influences on commuters' choices among car, bus, and Bay Area Rapid Transit (BART) for traveling to work in the San Francisco Bay area. The attributes were (a) cost, (b) total travel time, (c) dependability, (d) relaxation, (e) safety from accidents, (f) time use while traveling, (g) flexibility, (h) seat availability, (i) safety from crime, and (j) waiting time.

A sample of commuters were interviewed and each was asked to rate his or her satisfaction with car, bus, and BART on each of the ten attributes. The relative importance of the attributes was inferred by examining these ratings and the relationships between the ratings and the usual choice of travel mode.

The research was intended to have some immediate applications as a general diagnostic tool in transportation planning for evaluating the relative importance of various attributes that might otherwise be misjudged or overlooked. Primarily, however, the research was considered exploratory, the first stage in a multistage research strategy. Applications to quantitatively detailed planning procedures—such as travel demand forecasting or cost-benefit analysis—require additional research to identify policy variables that underlie the attributes identified as important and to determine how these policy variables are related to utility and behavior (1).

In basic objectives and methodology, this research was similar to a number of recent studies (2, 3, 4, 5, 6, 7, 8). The study differed from previous research, however, in that attribute importance was measured with a statistic that estimated how much each attribute contributed to differences in utility among the choice alternatives, for the people in the study sample. Most previous research has failed to consider an essential component of the quantity measured by this statistic, namely, average differences in utility among alternatives caused by average differences among alternatives in the levels of each attribute.

To demonstrate the importance of this difference, one must consider some theoretical and methodological issues in detail. This is done in the following section of the report. Readers interested primarily in the substantive conclusions of the research could skip to the section on data collection without loss of continuity.

## MEANING AND MEASUREMENT OF ATTRIBUTE IMPORTANCE

The theoretical concepts underlying this study can be summarized in the form of a linear utility model. For a detailed discussion of linear utility models and their applications to research on travel behavior see Domencich and McFadden (9). The model is

$$U_{mk} = \sum_j B_j X_{jmk} + e_{mk} \qquad (1)$$

where

$U_{mk}$ = utility of travel mode m for person k,
$X_{jmk}$ = measured value of attribute j for mode m and person k,
$B_j$ = coefficient representing the influence on utility of attribute j as measured with variable $X_{jm}$, and

$e_{mk}$ = stochastic error term influencing the utility of mode m for person k.

The fundamental, virtually tautological axiom underlying the use of this model is that, given a choice among two or more alternatives, an individual will select the one with the greatest utility.

Given the assumptions embodied in the linear utility model and having estimated the utility coefficients $B_j$, it is possible to calculate one or more importance statistic for each attribute variable. These statistics would indicate the extent to which each attribute influences the utilities of and consequently choices among the alternatives in the choice set (10).

A simple importance statistic can be defined, in terms of a choice made by a particular individual between two alternatives, as

$$I_{jk}(m,n) = B_j \left| X_{jmk} - X_{jnk} \right| \tag{2}$$

This importance statistic reflects the extent to which the phenomena measured by variable j contribute to differences in utility between alternatives m and n for person k.

Note that the importance statistic in Equation 2 is calculated as the product of two factors: first, the utility coefficient, which indicates the extent to which one unit of the variable is related to utility, and, second, the number of units of the variable by which the two alternatives differ. As the value of either of these two factors increases, the value of the importance coefficient increases. If either factor has a zero value—i.e., either the variable has no influence on utility or the alternatives do not differ on the variable—the value of the importance coefficient is zero.

## A General Importance Coefficient

The simple importance statistic (Equation 2) can be generalized to apply to choices made by any number of people among any number of alternatives. To generalize the statistic to samples of more than one person, the average value of the statistic can be calculated over all people in the sample. Similarly, to generalize the statistic to choices among more than two alternatives, the average of the two-alternative statistics can be calculated over all possible pairs of alternatives in the choice set. Finally, to simplify the calculations, averages of absolute value terms can be approximated with root mean squares. Thus an aggregate importance statistic for variable j for choices made by k individuals among P alternatives is:

$$I_j = B_j \left( \left\{ \sum_{k=1}^{K} \sum_{m=1}^{P} \sum_{\substack{n=1 \\ (m>n)}}^{P} [(X_{jmk} - X_{jnk})^2] \right\} \bigg/ K\binom{P}{2} \right)^{1/2} \tag{3}$$

Comparisons of this statistic for different attributes indicate, for the study sample, the relative extent to which each attribute contributes to differences in utility among the set of alternatives investigated (car, bus, and BART).

## Components of the Importance Statistic

The aggregate importance statistic (Equation 3) can be partitioned into two components, $I_j = C_{j1} + C_{j2}$, where

$$C_{j1} = B_j \left( \left\{ \sum_{m=1}^{P} \sum_{\substack{n=1 \\ (m>n)}}^{P} [(\bar{X}_{jm} - \bar{X}_{jn})^2] \right\} \bigg/ \binom{P}{2} \right)^{1/2} \tag{4}$$

$$C_{j2} = B_j \left[ \left( \sum_{k=1}^{K} \sum_{m=1}^{P} \sum_{\substack{n=1 \\ (m>n)}}^{P} \left\{ [(X_{jmk} - X_{jnk}) - (\bar{X}_{jm} - \bar{X}_{jn})]^2 \right\} \right) \bigg/ K\binom{P}{2} \right]^{1/2} \tag{5}$$

Thus

$$\bar{X}_{jm} = \left[ \sum_{k=1}^{K} (X_{jmk}) \right] \bigg/ K$$

and

$$X_{jn} = \left[ \sum_{k=1}^{K} (X_{jnk}) \right] \bigg/ K$$

The second of the two importance components, $C_{j2}$, is a standardized utility coefficient, which, for any variable, equals the utility coefficient that would be estimated if the variable were transformed such that the mean variance of the differences in variable values between pairs of alternatives was 1.0.

The standardized utility coefficient can be interpreted as a measure of partial attribute importance. It reflects the extent to which a change of one standard deviation in the attribute difference variable causes a change in the utility difference between two alternatives and, consequently, a change in the choice probabilities for the two alternatives. It indicates how much the attribute contributes to variations over the sample in the utility differences between alternatives. However, unlike the total importance coefficient (Equation 3), the standardized utility coefficient is not sensitive to the average utility differences between alternatives caused by the attribute. For example, the standardized utility coefficient would not reflect the extent to which choices among car, bus, and BART were influenced by average differences in travel time among the three modes.

This failure to reflect average differences also holds for two other statistics—t-statistics and correlation coefficients—that are commonly interpreted as measures of importance for variables in linear utility models.

Johnson (10, 11) discusses the properties of standardized utility coefficients, t-coefficients, and correlation coefficients, and derived their relationships to the coefficient of total importance for a simple case.

# DATA COLLECTION

## The Survey

The study was based on data obtained in the spring of 1975 from 258 people in the San Francisco-Oakland area. The sample was designed to consist of potential transit commuters living and working in areas well served by bus and BART. Operationally, this meant people who lived in areas accessible to bus and BART service and who worked in San Francisco, Oakland, or Berkeley—all cities well served by bus and BART.

At the time of the survey the BART system was operating on all lines of the system during the day on weekdays but had no evening or weekend service. The people in the study sample were interviewed by telephone. The sample was selected by using random telephone dialing (12).

As indicated in the following table, the characteristics of the study sample were generally comparable to census statistics (1970 census) for workers in the San Francisco-Oakland area. However, as expected from the sample design, the sample had a higher proportion of transit commuters than did the metropolitan area as a whole (24 percent versus 15 percent).

| Variable | Study Sample (%) | Workers in San Francisco-Oakland Area (%) |
|---|---|---|
| Sex | | |
| Male | 56 | 61 |
| Race | | |
| White | 74 | 84 |
| Age | | |
| Under 45 | 69 | 63 |
| 45-64 | 29 | 35 |
| Over 64 | 2 | 3 |
| Income | | |
| Under $8000 | 23 | 23 |
| $8000-$14 999 | 36 | 40 |
| Over $14 999 | 41 | 36 |
| Autos in household | | |
| 0 | 13 | 8 |
| 1 | 39 | 42 |
| 2 or more | 48 | 50 |
| Usual mode to work | | |
| Drive auto | 61 | 65 |
| Ride auto | 5 | 9 |
| Transit | 24 | 15 |

### Attribute Ratings

Survey respondents were asked to indicate their perceptions of commuting by car, bus, and BART by rating each of the modes available for their work trip on the ten attributes of interest. The rating categories were good, fair, and poor.

The wordings used to describe the attributes were

1. Cost: "the cost,"
2. Total travel time: "the total travel time door to door,"
3. Dependability: "knowing you can get to work on time,"
4. Relaxation: "how much you can relax,"
5. Safety from accidents: "safety from accidents,"
6. Time use: "the chance to do useful or pleasant things while traveling,"
7. Flexibility: "being able to travel when and where you want to,"
8. Seat availability: "your chances of getting a seat,"
9. Safety from crime: "safety from crime and being annoyed by the unpleasant behavior of other people," and
10. Waiting time: "the time you spend waiting."

Respondents were not asked to rate modes that they reported to be impossible to use in commuting.

The ratings were ordered by modes within attributes; i.e., all available modes were rated on cost, and then all available modes were rated on total travel time, etc.

Three of the attributes—seat availability, safety from crime, and waiting time—were rated only for bus and BART. Because of a misunderstanding of the interview instructions, ratings of these three attributes were made only if both bus and BART were reported to be possible for the respondent's trip. Consequently, the rating data on these attributes were available for a sample smaller than the one for data on the other attributes.

## DATA ANALYSIS

### Interrelationships Among Attribute Variables

The first step in the data analysis was to examine the intercorrelations among the attribute rating variables in order to identify any groups of highly intercorrelated variables. Matrices of Pearson correlation coefficients were calculated separately for the set of attribute rating variables for each of the three modes. Each matrix was then analyzed with a factor analysis procedure consisting of an image analysis (13) followed by an oblique Oblimin rotation (14).

The correlation matrices and factor analysis results were very similar for each of the three modes and identified three groups of highly intercorrelated variables: (a) time, dependability, waiting time, and, to a lesser extent, flexibility; (b) relaxation, time use, and, to a lesser extent, safety from accidents and seat availability; and (c) safety from crime and waiting time.

The correlation of transit ratings to safety from crime and waiting time is interesting. It suggests either a coincidental correlation of different underlying determinants—i.e., bus waits may tend to be longer in more dangerous areas—or the possibility that the ratings of both attributes reflect the common influence of perceived danger, meaning that a more dangerous situation may make waiting seem longer. For these three groups of variables and for selected subsets of these groups the average intercorrelations were calculated. The results are presented in the table below (attributes in the car column marked with a hyphen were not rated).

| Attribute Groups | Rated Mode | | |
|---|---|---|---|
| | Car | Bus | BART |
| Time, dependability | 0.44 | 0.54 | 0.56 |
| Time, waiting time | - | 0.63 | 0.36 |
| Time, dependability, flexibility | 0.37 | 0.49 | 0.42 |
| Time, dependability, waiting time | - | 0.54 | 0.42 |
| Time, dependability, waiting time, flexibility | - | 0.48 | 0.38 |
| Relaxation, time use | 0.50 | 0.54 | 0.28 |
| Relaxation, time use, seat availability | - | 0.50 | 0.23 |
| Relaxation, time use, safety from accidents | 0.42 | 0.45 | 0.30 |
| Relaxation, time use, seat availability, safety from accidents | - | 0.42 | 0.25 |
| Safety from crime, waiting time | - | 0.45 | 0.34 |
| First seven attributes | 0.26 | 0.35 | 0.21 |
| All attributes | - | 0.35 | 0.23 |

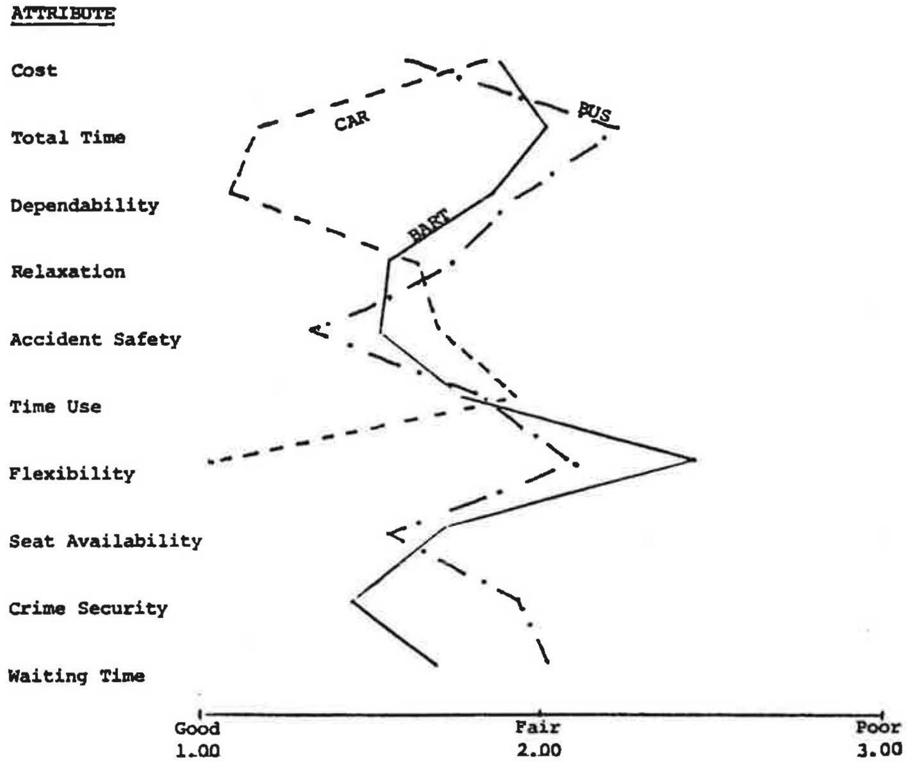
None of the groups of variables was sufficiently intercorrelated to suggest that the variables measured entirely the same phenomena. Nevertheless, the interrelationships among the attribute variables should be kept in mind when evaluating the results of subsequent analyses. It is possible that for intercorrelated variables the relationships to behavior may reflect the influence of a common set of underlying policy variables. Johnson (11) has discussed the problem of evaluating attribute importance when attribute variables are intercorrelated and has considered the advantages and disadvantages of several alternative methods of analysis.

### Average Attribute Ratings

The next step in the data analysis was to compute the average rating of each attribute for car, bus, and BART. Other things being equal, the more alternatives differ, on the average, with respect to an attribute, the more influence the attribute has on preferences among the alternatives. The average ratings are illustrated in Figure 1.

On the average, the car was rated as far superior to bus and BART on total travel time, dependability, and flexibility. On the other attributes, car commuting was rated as slightly inferior to transit travel, especially with respect to safety from accidents. The average ratings for bus and BART commuting were generally similar, the major differences being that BART commuting was rated as slightly better in terms of safety

Figure 1. Average attribute ratings of car, bus, and BART for commuting to work.




from crime, waiting time, and relaxation.

Seat availability, crime safety, and waiting time were not rated for car travel. However, assuming, as seems reasonable, that car would have been rated good on these attributes, the differences in evaluations between the car and the two transit modes are substantial.

## Relationships Between Attributes and Behavior

To evaluate the extent to which the average differences in ratings reflected average differences in utility and to estimate the other components of attribute importance, it was necessary to analyze the relationships, over the study sample, between the attribute ratings and preferences among the rated modes.

### Graphs

As a preliminary step in the analyses and as a convenient means of visualizing the relationships between the attribute ratings and behavior, graphs were constructed by relating the probability of choosing among alternative modes to differences in attribute ratings for the modes. A separate graph was calculated for each attribute.

To simplify the graphical presentation and to increase the size of the sample reflected in each graph, the information on bus and BART modes was condensed to create a single "preferred" transit mode for each individual. If the person regularly commuted by one of the transit modes or if only one mode was possible, it was considered the preferred mode. Otherwise, the preferred mode was determined by a question in the interview on which mode the person would prefer to use if he or she did not drive to work. The graph for each attribute related the probability of choosing transit over auto to differences in attribute ratings for the two alternatives.

The graphs are presented in Figure 2. They indicate that the ratings for all the attributes were strongly related to reported behavior. For most of the attributes the sample proportions using transit ranged from between 0 and 10 percent when the differences value was minus two (auto rated good, transit rated poor) to about 50 percent when the difference value was plus two (auto rated poor, transit rated good). The relationships were somewhat weaker for the attributes of relaxation, time use, and safety from accidents, however.

### Logit Analyses

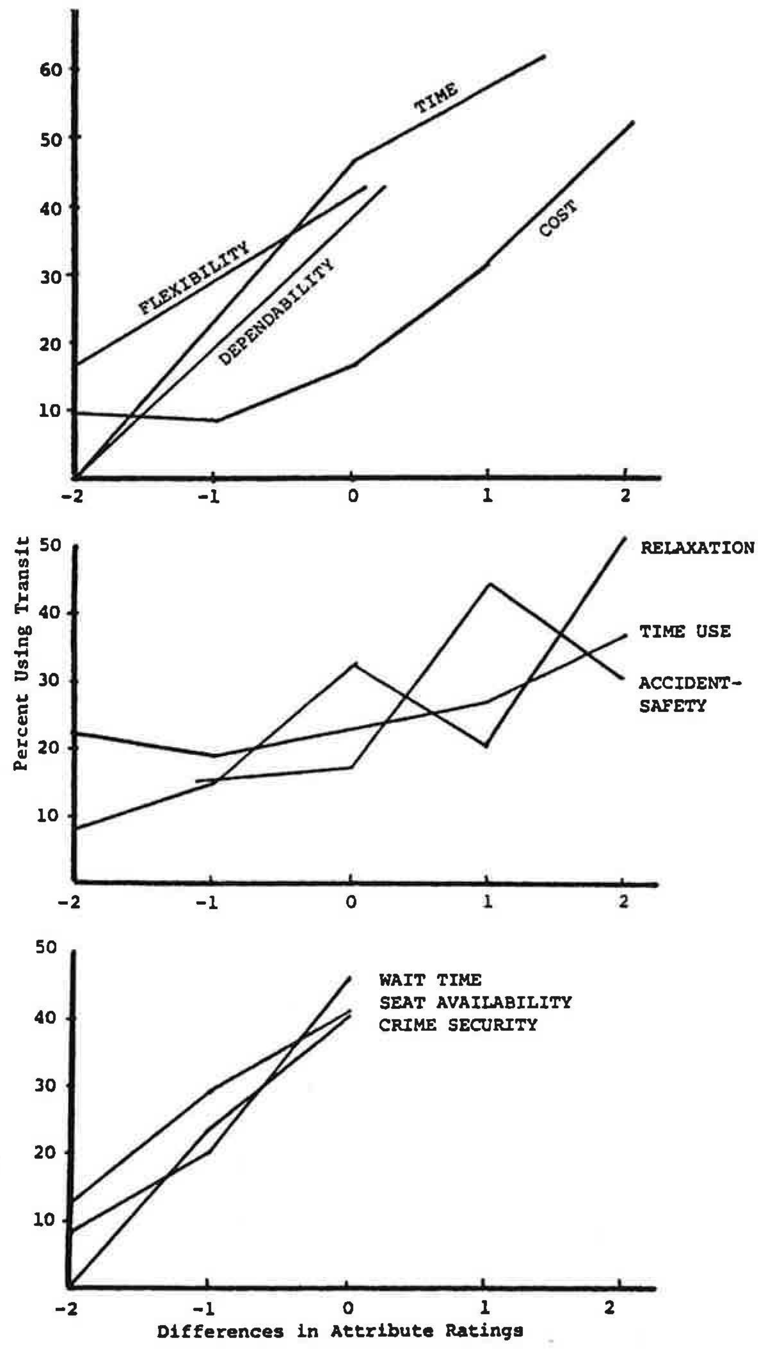
To provide a more sensitive and theoretically appropriate analysis of the relationships between the attribute ratings and behavior, maximum likelihood logit analyses (15) were done that related the attribute ratings to the choices among car, bus, and BART over the study sample. The analyses were carried out on the QUAIL system of computer programs (16).

The results are shown in the following table, where $B_z$ is the standardized utility coefficient and LRI is the likelihood ratio index of "pseudo $r^2$."

| | Statistic | | | | |
|---|---|---|---|---|---|
| Attribute | $B_z$ | t | LRI | Correct (%) | df |
| Cost | 0.89 | 4.22 | 0.34 | 75 | 218 |
| Time | 1.26 | 5.34 | 0.40 | 78 | 218 |
| Dependability | 1.37 | 4.97 | 0.40 | 78 | 218 |
| Relaxation | 0.63 | 3.25 | 0.30 | 76 | 218 |
| Safety from accidents | 0.60 | 3.09 | 0.30 | 74 | 218 |
| Time use | 0.35 | 1.91 | 0.27 | 76 | 218 |
| Flexibility | 0.49 | 2.54 | 0.28 | 74 | 218 |
| Seat availability | 0.79 | 2.35 | 0.28 | 68 | 138 |
| Safety from crime | 1.19 | 3.39 | 0.34 | 70 | 138 |
| Waiting time | 1.51 | 3.84 | 0.38 | 71 | 138 |
| All attributes | – | – | 0.57 | 79 | 129 |

Although the primary purpose of the logit analyses was to obtain a utility coefficient for each attribute, as an input to calculating attribute importance coefficients, the logit results also included values of the LRI, which reflected the strength of the relationship between the

Figure 2. Relationships between travel mode choice and attribute ratings.




attribute measures and travel mode choice.

For the individual attributes, the values of the LRI ranged from 0.27 to 0.40. For a multiple logit analysis using the entire set of variables simultaneously, the value of the index was 0.57. These LRI values are equal to or larger than most values that have been reported for similar research on travel mode choice, using either subjective or objective data. The logit results thus corroborate the evidence, shown in the graphs, that the attribute ratings were substantially related to reported behavior.

## Importance Coefficients

For each attribute, the estimated utility coefficient was combined with values of the attribute ratings, over the sample, into an importance coefficient, based on Equation 3 above, that reflected the extent to which the attribute contributed to differences in utility among the alternative travel modes. Waiting time, safety from crime, and seat availability were not rated for the car alternative, so calculation of the importance coefficients was based on the assumption that car would have been rated good on these attributes by all respondents.

Three sets of importance coefficients were calculated in order to see how the importance of each attribute as an influence on choice is reflected among the three pairs of modes (car-bus, car-BART, and bus-BART). An additional set of coefficients was calculated to reflect the overall importance of each attribute, for choices among all three modes. The importance coefficients are presented below.

| Attribute | Sets of Travel Modes: Car-Bus | Car-BART | Bus-BART | Car-Bus-BART |
|---|---|---|---|---|
| Cost | 0.95 | 0.98 | 0.74 | 0.91 |
| Time | 1.78 | 1.74 | 1.44 | 1.69 |
| Dependability | 1.83 | 1.82 | 1.56 | 1.76 |
| Relaxation | 0.67 | 0.68 | 0.51 | 0.64 |
| Safety from accidents | 0.61 | 0.71 | 0.57 | 0.63 |
| Time use | 0.37 | 0.40 | 0.25 | 0.36 |
| Flexibility | 0.97 | 1.10 | 0.64 | 0.94 |
| Seat availability | 0.89 | 1.01 | 0.95 | 0.95 |
| Safety from crime | 2.01 | 1.27 | 1.53 | 1.64 |
| Waiting time | 2.30 | 1.78 | 1.92 | 2.02 |

In terms of overall importance and considering choices among all three modes, the attributes seemed to cluster into several groups having roughly equal importance statistics. Waiting time, dependability, total time, and safety from crime appeared to be the most important attributes. Cost, seat availability, and flexibility appeared to be next in importance, followed by relaxation and safety from accidents. Time use appeared to be the least important attribute.

For choices among the different pairs of alternatives, the relative importance of the attributes appeared to be about the same as for choices among all three modes. The major differences were that waiting time and safety from crime appeared relatively less important for choices between car and BART, that flexibility appeared relatively less important for choices between bus and BART, and that seat availability appeared relatively less important for choices between car and BART.

## CONCLUSIONS

Most of the attributes investigated in this study appeared to be important influences on travel mode choice. Respondents tended to rate the modes differently for most attributes, and the differences were strongly related to reported behavior. Among the attributes judged to be important were safety from crime, dependability, and seat availability, which are not typically included in quantitative planning procedures, such as travel demand forecasting or cost-benefit analysis.

The results suggest that these attributes should be taken more into account in transportation policy decisions. However, the conclusions must be qualified by the uncertainties discussed above regarding the extent to which the observed relationships to behavior of the different attribute variables actually reflected the influence of different underlying policy variables.

As discussed at the beginning of this paper, the conclusions from this research may have some immediate policy implications as general diagnoses. For example, safety from crime appears to be an important influence on choices between car and bus travel and may have some direct policy implications to transit managers serving the studied population (or similar populations elsewhere). Primarily, however, the research should be viewed as the first stage in a multistage research strategy. Subsequent research stages should identify policy variables that underlie the attributes identified as important and should determine how these policy variables are related to utility and behavior. The results of these later stages can be applied to more quantitively detailed planning procedures.

The major benefit of first-stage research on attribute importance is that it allows the relatively expensive and time-consuming research on objective policy variables to be focused on the most essential attributes.

For some travel attributes—such as safety from crime or social status—it may not be possible to identify a manageable set of policy variables underlying the attribute ratings. To evaluate the consequences of policies with respect to such attributes, subjective methods could be used in which people were asked to indicate the influence of contemplated changes on their attribute ratings. The changes in ratings could then be used to evaluate consequent changes in utility and behavior, using previously determined utility coefficients for the rating variables. A disadvantage of this procedure is that it requires a special research effort to estimate rating changes for every contemplated policy change.

## ACKNOWLEDGMENTS

This research was supported by the National Science Foundation through grants from the Research Applied to Nation Needs Program to the University of California, Berkeley.

## REFERENCES

1. D. A. Hensher and P. B. McLeod. Towards an Integrated Approach to the Identification and Evaluation of the Transport Determinants of Travel Choices. Transportation Research (in press).
2. R. Dobson and M. L. Tischer. Comparative Analysis of Determinants of Modal Choices of Central Business District Workers. TRB, Transportation Research Record 649, 1977, pp. 7-14.
3. D. T. Hartgen. Attitudinal and Situational Variables Influencing Urban Mode Choice: Some Empirical Findings. Transportation, Vol. 3, 1974, pp. 377-392.
4. G. C. Nicolaidis. Quantification of the Comfort Variable. Transportation Research, Vol. 9, 1975, pp. 55-66.
5. F. T. Paine, A. N. Nash, S. J. Hille, and G. A. Brunner. Consumer Attitudes Toward Auto Versus Public Transport Alternatives. Journal of Applied Psychology, Vol. 53, No. 6, 1969, pp. 472-480.
6. W. W. Recker and R. F. Stevens. Attitudinal Models of Modal Choice: The Multinomial Case For Selected Nonwork Trips. Transportation, Vol. 5, 1976, pp. 355-375.
7. W. W. Recker and T. F. Golob. An Attitudinal Modal Choice Model. Transportation Research, Vol. 10, 1977, pp. 299-310.
8. B. D. Spear. Generalized Attribute Variable for Models of Mode Choice Behavior. TRB, Transportation Research Record 592, 1976, pp. 6-11.
9. T. A. Domencich and D. McFadden. Urban Travel Demand: A Behavioral Analysis. Elsevier, New York, 1975.
10. M. A. Johnson. Defining and Measuring Attribute Importance in Choices Among Multi-Attribute Alternatives. Travel Demand Forecasting Project, Institute of Transportation Studies, Univ. of California, Berkeley, Working Paper 7702, 1977.
11. M. A. Johnson. Attitudes, Beliefs, and Transportation Behavior. Urban Travel Demand Forecasting Project, Institute of Transportation Studies, Univ. of California, Berkeley, Phase I Final Rept. Series, Vol. 6, 1977.
12. M. A. Johnson and D. G. McFadden. Field Materials for the 1975 Attitude Pilot Study Survey. Travel Demand Forecasting Project, Institute of Transportation Studies, Univ. of California, Berkeley, Working Paper 7610, 1976.
13. Henry F. Kaiser. Image Analysis. In Problems in Measuring Change (C. W. Harris, ed.), Univ. of Wisconsin Press, Madison, 1963.

14. H. H. Harman. Modern Factor Analysis. Univ. of Chicago Press, 2nd Ed., 1967.
15. D. McFadden. Conditional Logit Analysis of Qualitative Choice Behavior. In Frontiers in Econometrics (P. Zarembka, ed.), Academic Press, New York, 1973.
16. J. Berkman, D. Brownstone, G. Duncan, and D. McFadden. QUAIL User's Manual. Urban Travel Demand Forecasting Project, Institute of Transportation Studies, Univ. of California, Berkeley, Working Paper 7402, 1974.

*Publication of this paper sponsored by Committee on Passenger Travel Demand Forecasting and Committee on Traveler Behavior and Values.*

# Intercity Rail Travel Models

Gerald S. Cohen, Nathan S. Erlbaum, and David T. Hartgen, Planning Research Unit, New York State Department of Transportation

**Using a 1975 aggregate data base of 31 pairs of cities, forecasts are made of 1975-1980 rail patronage in the New York City-Buffalo corridor. A two-stage modeling process is used to estimate total city-pair volume by purpose, using gravity formulations relating annual volume to city size, government employment, and hotel and motel sales receipts. Binary logit models are then developed in which rail competes differentially with air, auto, and bus in order to avoid independent irrelevant alternatives assumptions. Rail service and terminal quality variables are included with time, cost, and frequency. The total rail share is then determined algebraically from the binary models. Pivot-point analysis is used to increase the accuracy of the forecasts. Results show that rail competes differently with each mode. Against air, the key variables are frequency and time ratios; against auto, the frequency, cost, and time ratios and terminal quality are important; against bus, train service quality, frequency ratio, and time difference are important. Elasticities of demand vary considerably by mode and distance. Forecasts show that if train, track, service, and terminal improvements are implemented as planned in the corridor over the next 5 years, 1980 link volumes will increase 58-105 percent over 1975 levels, with most diversion coming from short-distance auto trips. The net effect of this diversion will be to reduce 1980 corridor energy requirements by 9 percent over 1975.**

The Planning Research Unit of the New York State Department of Transportation (NYSDOT) recently cooperated with Union College to study the energy efficiency of train service in the New York City (NYC)-Buffalo corridor. NYSDOT's role was to develop a workable model of train passenger demand and to analyze energy and passenger-kilometer efficiencies of alternative train, track, and service improvements in the corridor. This report summarizes the rail passenger demand models developed in the study. It briefly describes the data used, the developed models, and the pivot-point and normalization procedures used to increase the accuracy of forecasts. Rail demand forecasts and the potential for modal energy savings in the corridor are also discussed.

## DATA

The data collection effort (1) concentrated on cities within the NYC-Buffalo corridor (Figure 1), with selected other city pairs included for continuity or availability or both of data. In total, 31 city pairs were included in the data base. Each city pair is described by a wide range of data elements (1) that include city size and spatial separation variables and modal service variables such as travel times, costs, and frequencies. In addition, the following quality-of-service data describing train and terminal service characteristics were also included:

Quality of rail service—snack car availability, sleeper car availability, lounge car availability, baggage service, package express, on-time performance, schedule match, dining car availability, and car type.

Terminal quality—parking availability, number of spaces, parking fee, parking lot lighting, terminal snack bar, local transportation, distance to downtown, and modernness of terminal.

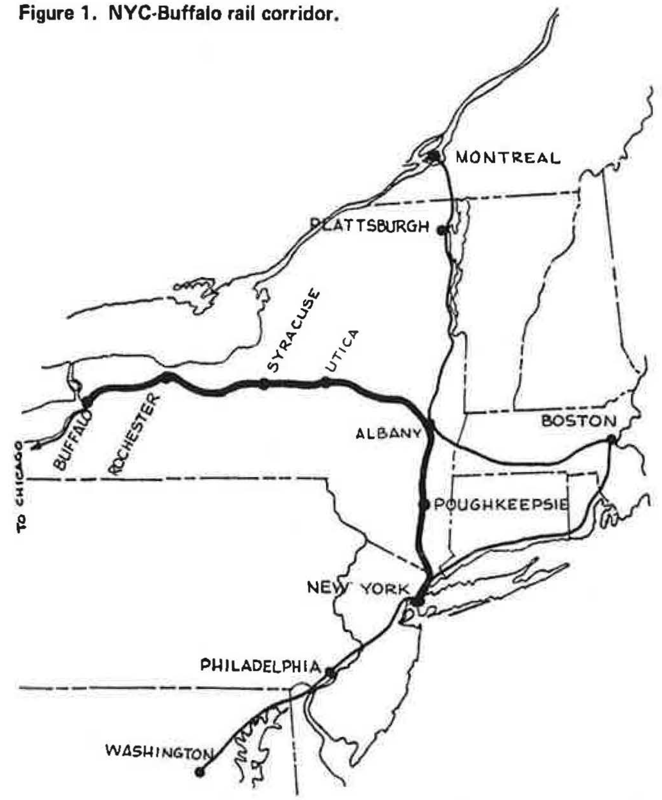
Three central findings of the preliminary research conducted for the study (1,2) were as follows. First, when all factors are considered, a hybrid modeling approach—forecasting total intercity volume and then separate modal shares—appears to be the most productive. Second, care must be taken to avoid formulations that contain the so-called independence of irrelevant alternatives assumptions (IIA assumptions). Third, since quality of service influences modal choice, models developed should consider quality-oriented data on the rail system, including rail terminals. Ideally, such data should also be available for the other modes.

These principles led to the development of a new approach to intercity rail passenger demand modeling, which has been fully documented (3). The approach is to use two total travel models that forecast travel via all four modes between each city pair. These were developed for business and nonbusiness travel. The models have a simple gravity format. Estimates of total volume, however, will not replicate observed total volume because of residual errors caused by incomplete model specification. The slippage between estimated and true total volume in the base year (1975) can be eliminated for future forecasts, using pivot-point procedures.

Within each trip purpose, three separate competition models are developed for rail versus air, auto, and bus. The model form is binary logit. Each model includes only those variables relevant to the binary choice, for example, rail versus bus. These models are then used to derive a consistent estimate of the rail share. These rail shares, however, will not replicate the observed rail shares because of residual errors caused by incomplete model specification. The slippage between estimated and true rail shares in the base year (1975) can also be eliminated using pivot-point analysis.

The future modal volumes, in particular the rail volumes, are then obtained by multiplying the total volumes (as computed by the total models) by the modal share (as computed by the share models). Pivot-point analysis, described later, is used to reduce forecasting

Figure 1. NYC-Buffalo rail corridor.




from the residual error between the estimated and actual observations, which is contained in the calibrated models.

When forecasts of rail volume are made, the future modal shares, estimated with pivot-point, will no longer total 1.0. To ensure that they do, a normalization factor must be applied to all forecasts.

One reason for using this step-down approach, rather than the alternative approach of constructing direct-demand modal models such as the Kraft-Sarc or Baumol-Quandt, is that, although city attractiveness and city-pair impedance measures are generally the variables that influence the total travel volume, they are not the primary variables influencing mode choice. The construction of a sequence of models thus enables the analyst to better isolate the effects of changes in many variables on both total travel and mode usage.

## TOTAL TRAVEL

Gravity-like models were constructed separately for business and nonbusiness travel. These models are of the form

$$\text{Total annual trips}_{ij} = [K(\text{attractiveness of i and j})^a] \div (\text{travel impedance of i and j})^b$$

Some of the several different measures tested are

- $P_iP_j$ = population product;
- $In_iIn_j$ = median income product;
- Gov $P_iP_j$ = (Gov$_i$ $P_i$)(Gov$_j$ $P_j$); that is, population product weighted by percentage of government employment;
- 25+ $P_iP_j$ = (Inc 25+$_i$ $P_i$)(Inc 25+$_j$ $P_j$); that is, population product weighted by the percentage of families earning $25 000+;
- Au $P_iP_j$ = (Auto$_i$ $P_i$)(Auto$_j$ $P_j$); that is, population product weighted by the percentage of families owning an auto;
- Hot $P_iP_j$ = (Hot$_i$ $P_i$)(Hot$_j$ $P_j$); that is, population product weighted by the percentage of total receipts that are from hotels and motels;
- D = distance;
- ATT = air travel time;
- AUT = auto travel time;
- BTT = bus travel time;
- RTT = rail travel time;
- Avg BT = average business travel time; and
- Avg NBT = average nonbusiness travel time.

The last two variables are volume-weighted averages of the travel time by all modes.

All models were calibrated by using the aggregate data documented in (1). Stepwise linear regression techniques (BMD02R) were used for estimating coefficients. Results are more fully documented elsewhere (3).

The use of the variables (Hot $P_iP_j$) and (Gov $P_iP_j$) generally resulted in models with a slightly (0.05) higher $R^2$ than the other attractiveness measures. In addition, they permit additional analysis of the effect of changes in variables other than merely city size. In most cases there was little difference in model strength for different travel impedance measures. Generally the strongest models contained bus travel time, but only slightly weaker models were obtained using auto time, rail time, or over-the-road distance. Average business time produced significantly weaker models than average nonbusiness time, and the models using air travel time as the measure of travel impedance were far weaker than the other models. This is because of the combined effects of (a) a high proportion of air traffic in the business market and (b) the much faster air speeds for longer trips, which results in an anomaly that long-distance interchanges (e.g., Buffalo-Albany) may actually have a shorter average business travel time than shorter distance interchanges (e.g., Albany-Rochester). Average nonbusiness travel time, however, is generally monotonic with distance, is a reasonable proxy of intercity spatial separation, and permits policy analysis. The models finally selected were

$$T = 37.15\,(\text{Gov } P_iP_j)^{0.70}(\text{Avg NBT})^{-2.57}$$
$$F = \quad (58.97) \qquad (64.77)$$
$$R^2 = 0.758$$

for business trips and

$$T = 12.88\,(\text{Hot } P_iP_j)^{0.94}(\text{Avg NBT})^{-2.69}$$
$$F = \quad (55.92) \qquad (59.36)$$
$$R^2 = 0.746$$

for nonbusiness trips. Here, T is in hundreds of trips, Gov $P_iP_j$ and Hot $P_iP_j$ are in thousands, and Avg NBT is in hours. These models were chosen for several reasons.

The attractiveness measures have intuitive appeal. It seems reasonable that business trip volume should be influenced by both the populations of the cities and the proportion of government workers. For example, Albany and Washington have a much greater volume of business travel between them than might be expected if one merely considered their population and distance. Similarly, the percentage of hotel and motel receipts partially reflects a high proportion of tourist trips in nonbusiness travel.

The travel impedance measure used, average nonbusiness travel time, also had several virtues. Unlike distance, it is policy sensitive, since travel-time changes in a given mode will influence the overall city separation. The implied elasticities are more reasonable; for example, if one considers the best business model using rail time as an impedance measure, it appears that a 1 percent decrease in rail travel time will result in an increase in total volume of 1.8 percent. This is clearly an unrealistic level of induced travel. In contrast, if one considers the best business model using average nonbusiness travel time, a 1 percent decrease in rail travel time will result in an increased volume equal to 2.57 times the rail share (mean value ~0.02), or 0.051 percent.

Although the $R^2$ values of the recommended models are not the largest obtained, the proportion of variation explained is only slightly lower. The $R^2$ values of most of the models obtained have approximately the same magnitude as those of the recommended models.

## MODE CHOICE

There are several reasons for the use of binary logit competition models. By developing binary competition models the planner can obtain additional insight into the variables that influence particular mode choices. For example, the results suggest that rail can best compete with bus by concentrating on improving the amenities it offers but can best compete with auto by improving travel time.

The approach appears to reduce the problems caused by the independence of irrelevant alternatives axiom.

The models can be readily calibrated by using a standard stepwise linear regression program (BMD02R). Thus, a greater insight into the relative importance of the variables in the model can be obtained. Present multilogit models do not offer stepwise selection of variables.

Direct demand models such as Kraft-Sarc are less suitable when 0-1 dummy variables representing availability of service items are used.

Although the model is not a constant elasticity model, the formula for determining the elasticity of demand with respect to a particular independent variable does have a simple intuitive form and is analytically tractable.

The general form of the models is a binary logit model describing the mode choice between rail and one other mode. For example, using the typical logit form, the competition of rail versus air may be expressed as

$$\text{LN}\left\{[\text{rail volume}/(\text{rail} + \text{air volume})] \div [(1 - \text{rail volume})/(\text{rail} + \text{air volume})]\right\} = G_{air} = a_0 + a_1 x_1 + \ldots \quad (1)$$

where $G_{air}$ is calibrated by using linear regression techniques and the x's are variables describing the rail-air trade-off. The rail versus auto and rail versus bus competitions are similarly calibrated. These three equations can then be solved simultaneously, resulting in the principal equation used for forecasting the rail share.

$$P_{rail} = [\text{rail}/(\text{rail} + \text{air} + \text{auto} + \text{bus})] = [1/(e^{-G_{air}} + e^{-G_{auto}} + e^{-G_{bus}} + 1)] \quad (2)$$

where the total rail modal volumes can then be estimated by

$$\text{rail volume} = [\text{prob (rail)}]\ (\text{total volume, from total models}) \quad (3)$$

Many variables were considered as potential predictors of mode choice, including relative travel times, costs, and frequencies for all four modes. Various forms of these variables were considered, including costs per kilometer, relative times, costs and frequencies, and time differences. The only successful combination for the cost variables was relative cost. Generally, the cost variables proved unsatisfactory, since the cost and cost per kilometer variables for all four modes are generally highly positively correlated with trip length and hence negatively correlated with rail share.

Other variables considered were rail quality variables. These variables included terminal descriptors such as the nature of restaurant and parking facilities and train or system descriptors such as on-time performance and proportion of trains with snack cars. As was the case with cost variables, the use of cross-sectional data resulted in the need to discard certain potential variables because their correlations with the dependent variable had the wrong (illogical) sign. For example, sleeping cars are generally used on particularly long routes. It is on these routes that rail is least competitive with air. Thus, the data show that, as the percentage of sleeping cars increases, the proportion of rail use to air use decreases. Therefore, a variable such as percentage of sleeping cars will not be a useful variable for making policy-sensitive forecasts, since its introduction will result in absurd forecasts.

Preliminary calibration efforts showed that, in several cases, variables describing rail terminal characteristics would provide more insight into choice decisions if they were combined into an index of terminal quality. When possible, indexes were evaluated on the basis of logic, contribution to model strength, and ease of forecasting. The most effective index proved to be

$$\text{index 1} = \text{park} + \text{dine} - \text{dist.} \quad (4)$$

where

park = a 0-1-2 variable describing parking conditions at each terminal [described in (2)],
dine = a 0-1 variable describing dining facilities, and
dist. = distance (kilometers) from rail terminal to downtown.

Table 1 summarizes the models. The air versus rail competition models generally had the highest $R^2$. In both the business and the nonbusiness models, most of the variance was explained by the ratios of air-to-rail time and air-to-rail frequency. Train quality-of-service variables were not important in the business model. We hypothesize that air already holds a very significant competitive edge. For nonbusiness trips, terminal characteristics also appear to be a significant predictor of mode choice, as indicated by the presence of index 1.

The F-statistics suggest that frequency appears to be more significant for business trips, and time for nonbusiness trips. The business and nonbusiness auto-rail competition models were similar. Rail frequency, however, was a strong variable for business trips but did not appear in the nonbusiness model. In contrast, index 1 (terminal quality) was a more significant variable in the nonbusiness model.

The elasticities for relative time are quite large in both models, undoubtedly due to the numerous short-distance interchanges in the corridor. This suggests that a significant increase in rail speeds would greatly increase rail's ability to compete with auto.

The G(x)'s for both business and nonbusiness bus-rail competition models are functions of relative fre-

**Table 1. Modal competition share models for rail versus air, auto, and bus.**

| Model | $R^2$ | G(x) |
|---|---|---|
| Rail versus air | | |
| | 0.79 | -4.14 - 0.39 (air:rail freq) + 5.33 (air:rail time) |
| F | | 29.74 11.13 |
| Elasticity | | -0.46 -0.61 |
| | 0.80 | -4.02 - 0.25 (air:rail freq) + 0.30 (index 1) + 6.46 (air:rail time) |
| F | | 17.77 3.63 23.37 |
| Elasticity | | -0.10 0.69 0.26 |
| Rail versus auto | | |
| | 0.67 | -15.27 + 0.008 (rail freq) + 0.235 (index 1) + 1.99 (auto:rail cost) + 7.67 (auto:rail time) |
| F | | 14.41 2.21 9.0 17.23 |
| Elasticity | | 0.39 0.40 1.92 4.99 |
| | 0.66 | -15.32 + 0.32 (index 1) + 1.60 (auto:rail cost) + 9.56 (auto:rail time) |
| F | | 5.03 7.18 33.77 |
| Elasticity | | 0.69 1.78 7.16 |
| Rail versus bus | | |
| | 0.71 | -9.21 + 10.54 (food) - 0.23 (bus:rail freq) + 0.45 (bus:rail time) |
| F | | 11.74 46.25 15.27 |
| Elasticity | | 0.15 -0.03 0.004 |
| | 0.62 | -11.60 + 11.09 (food) - 0.094 (bus:rail freq) + 0.55 (bus:rail time) |
| F | | 8.45 9.94 13.05 |
| Elasticity | | 0.67 -0.05 0.02 |

quency, travel time differences, and a measure of rail service amenities. This measure of rail amenities is the variable "food service," which is the proportion of trains that have either a snack bar or a dining car. All variables have large F-values, but the $R^2$-values for both models are moderate. This suggests that the unexplained variance is in large part attributable to a specification error; i.e., some of the factors that influence the mode choice between rail and bus are not included in the model.

Some of these factors that were not included in our study are bus service quality variables, relative proportion of advertising expenditures, and bus terminal characteristics.

## PIVOT-POINT

Pivot-point is a procedure for adjusting a forecast so that it does not contain the residual error between the estimated and the actual observation, which is contained in the calibrated model. To illustrate, consider model $Y = a + b(x)$, representing the best-fitting line that can be drawn through the data.

Let $(x_o, y_o)$ be an actual observation used to calibrate the model. Now, $\hat{Y}_o = a + bx_o$ is the estimated value of $Y_o$. In an actual forecast the pivot-point argument reasons that the best estimate for $Y_F$ is

$$Y_F = (Y_o/Y_o)(a + b_1 x_{1F} + \ldots) \tag{5}$$

Note that the pivot-point line is not parallel to the best-fitting line. Rather than preserve the slope, the method preserves the elasticity of Y with respect to x.

## FORECASTS

To determine the future condition of rail service, extensive use was made of the Amtrak 5-year plan, New York's statewide master plan for transportation, and the proposed uses of the funds available form the recent New York State Railroad Bond Issue.

The statewide master plan and the New York State Railroad Bond Issue provide for a gradual improvement in rail service over the next decade. Rail costs (fares) are forecast to increase at a rate of approximately 5 percent per annum, and auto costs are projected using a transportation price index developed by the U.S. Department of Commerce (also projected to grow at about a 5 percent rate).

Generally it is anticipated that there will be a gradual turnover of trains, with Turbotrain and then Amfleet rolling stock being substituted for conventional equipment. Rail frequency is not anticipated to be increased in proportion to ridership increases in the next decade.

Travel times will increase somewhat between 1975 and 1977, chiefly because of slow orders imposed by reconstruction of sections of deteriorating track. After 1977, however, travel times are expected to decrease by 7-15 percent by 1980, with more significant improvements between 1980 and 1985. For example, it is anticipated that the travel time between NYC and Albany will decrease by about 20 min (-10.6 percent) between 1975 and 1980. These improved travel times reflect equipment and track improvements anticipated in the next decade (Table 2).

Forecasts of intercity rail traffic are made by adjusting an initial forecast by the pivot-point and normalization factors. The results of this calculation give city-pair volumes for the future year and are then added to obtain link forecasts. The effect of service improvements on ridership is clearly shown in Table 2 and in Figure 2.

For most links, the lowest rail volume occurred in 1977. The drop ranged from 29 percent for the NYC-Albany link to 2 percent for the Rochester-Buffalo link. By 1978, however, the situation will have improved: rail travel times will have greatly decreased and frequency will have increased or remained the same, but rail costs will be up. The effect of these improvements is to increase rail ridership substantially over 1975 levels, for most links. The trend will then continue in 1979 and 1980.

In terms of energy, the effect of this growth is shown in Table 3. The net effect is to reduce the total 1975 corridor energy requirements by 9 percent in 1980. The bulk of this energy savings comes from a 15 percent reduction in auto energy consumption, as a direct result of auto patronage shifts to rail. The increase in rail energy accounts for only 2 percent of corridor energy, while air and bus account for 31 and 5 percent respectively. Thus, the improvements would generate a ten-to-one energy savings: for every joule added to rail energy, about ten are saved from auto. It is clear that the proposed improvements to rail, which cause only small increases in total rail energy themselves, reap large reductions in total corridor energy consumption. This finding is consistent with most published opinions on modal energy consumption, in that the greatest energy savings can be accrued by those policies or acts that divert auto travel to other modes.

Table 2. 1980 NYC-Buffalo rail corridor percentage of use for 1975.

| Link | Service (%) Frequency of Service | Travel Time | Rail Cost | Patronage Forecasts (%) |
|---|---|---|---|---|
| NYC-Albany | 14.3 | -10.6 | 50.0 | 57.8 |
| Albany-Utica | 25.0 | -7.8 | 42.9 | 79.0 |
| Utica-Syracuse | 25.0 | -11.8 | 63.6 | 84.5 |
| Syracuse-Rochester | 33.3 | -14.8 | 40.0 | 105.6 |
| Rochester-Buffalo | 33.3 | -9.6 | 43.8 | 93.0 |
| Total corridor | | | | 66.6 |

Figure 2. Intrastate link volumes.

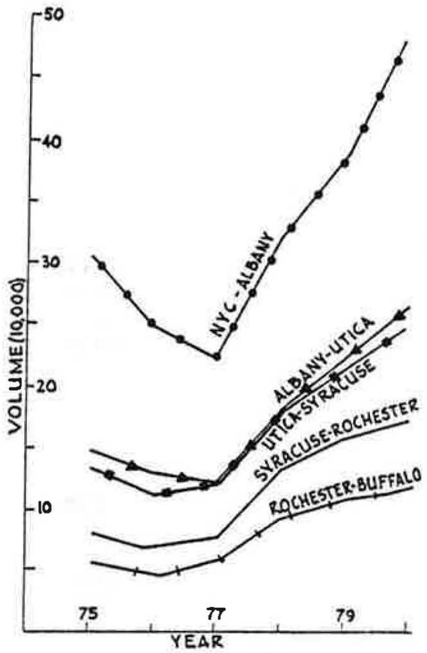



Table 3. Corridor energy consumption.

| Mode | 1975 Rate (GJ) | 1975 Rate (%) | 1980 Rate (GJ) | 1980 Rate (%) | Difference (GJ) | Change (%) |
|---|---|---|---|---|---|---|
| Air | 9 825 | 24.2 | 9 831 | 26.7 | 2.1 | +0.02 |
| Auto | 28 948 | 71.2 | 24 685 | 67.1 | -4263 | -14.73 |
| Bus | 1 646 | 4.1 | 1 647 | 4.5 | 1.0 | +0.06 |
| Rail | 199 | 0.5 | 622 | 1.7 | 423 | +212.11 |
| Total[a] | 40 623 | 100.0 | 36 787 | 100.0 | -3836 | -9.44 |

Note: 1 GJ = 9.47 × 10⁹.

[a]These totals were converted from BTUs (1 BTU = 1.05 kJ) and are therefore not the actual column totals, which were given in billion BTUs.

## CONCLUSION

The analysis shows that planned rail service improvements in the NYC-Buffalo corridor will have a substantial effect on rail patronage in the corridor over the 1977-1980 period. Rail traffic in this corridor is particularly sensitive to relative travel time (rail versus the other three modes), because the generally short-distance interchanges now favor auto use. If improvements in rail travel time are made, diversions will come primarily from auto.

Improvements in the quality and relative frequency and relative cost of rail service will also encourage diversion but will generally have a smaller impact than increase in travel time. As a result, rail patronage is expected to increase substantially, resulting in a 9 percent reduction in the 1975 corridor energy requirements by 1980.

This means that the best policies for improving rail passenger service and increasing energy efficiency are likely to be those that seek to significantly attract diverted patronage through improved travel times.

## ACKNOWLEDGMENTS

The authors acknowledge the assistance of Leon Jackson, Michael Trentacoste, Stephen Slavick, Larry Stevens, Axel Rose, and Ram Mittal for providing the data for this study. The assistance of Wilma C. Marhafer and Barbara J. Blowers in preparing the manuscript is gratefully appreciated.

## REFERENCES


1. D. T. Hartgen and G. S. Cohen. Intercity Passenger Travel Demand Models: State-of-the-Art. Planning Research Unit, New York State Department of Transportation Preliminary Research Rept. 112, Dec. 1976.
2. N. S. Erlbaum, M. F. Trentacoste, R. G. Knighton, and S. R. Slavick. NYS Intercity Travel Data, 1975. Planning Research Unit, New York State Department of Transportation Preliminary Research Rept. 113, Feb. 1977.
3. G. S. Cohen, N. S. Erlbaum, and D. T. Hartgen. Intercity Rail Patronage in the NYC-Buffalo Corridor: Models and Forecasts. Planning Research Unit, New York State Department of Transportation Preliminary Research Rept. 115, Apr. 1977.


*Publication of this paper sponsored by Committee on Passenger Travel Demand Forecasting and Committee on Traveler Behavior and Values.*

# Forecasting Travel Demand in Small Areas by Using Disaggregate Behavioral Models

Michael A. Johnson and Aaron Adiv, Institute of Transportation Studies, University of California, Berkeley

A study was done to forecast the patronage of a new transit system proposed for a suburban city in the San Francisco Bay Area, using disaggregate behavioral models of transportation choice. The results suggested that behavioral models of the type used in the study can be applied to travel demand forecasting in small urban areas but that additional development and testing of the models should be done before they are used for policy decisions. The time and expense required for data collection and analysis were within reasonable limits for general application. Although implementation of the forecasting methodologies was quite successful, results of tests of the predictive accuracy of the behavioral models were disappointing.

This report describes a study on forecasting patronage of a new transit system proposed for a suburban city in the San Francisco Bay Area. The forecasting was done with disaggregate transportation choice models, which describe the travel behavior of individual people rather than aggregate populations.

The general forecasting method was (a) to collect survey and transportation supply data for a sample of the people the transit system was intended to serve, (b) to use individual choice models to forecast the probability that each person in the sample would use the system under a variety of policy alternatives, and (c) to aggregate these individual forecasts to obtain general forecasts for the population of potential users, weighting as necessary to correct for disproportionate sampling of different population segments.

## TRAVEL DEMAND FORECASTING IN SMALL AREAS

The study was an example of travel demand forecasting for policies that affect small geographical areas. Typically, such forecasting cannot be done satisfactorily with the data bases and forecasting methods used by metropolitan planning organizations, which are ordinarily intended to investigate policies having impacts on an entire metropolitan region or on large transportation corridors within a region and are on a scale too large to be adequately sensitive to local policy changes.

On the other hand, when policies impact only a small geographical area, it is often feasible to collect data and make forecasts on an ad hoc basis at a level of detail appropriate to the particular policy issues. Disaggregate modes of transportation choice are easily adapted to such applications.

Several previous studies have been reported in which individual choice models were used to forecast travel demand in small geographical areas (1, 2, 3). A detailed review of this research has been presented in a separate paper (4). This study differs from most previous research of this type in several important respects: first in the number of different transportation alternatives considered, second in the detail and accuracy of the transportation supply data used as inputs to the forecasting models, and third in the use of analyses to test and compare the accuracy of the probabilities estimated by different behavioral models, using data available before the bus system was running.

These differences are explained in the remainder of the report.

## STUDY SITE: WALNUT CREEK

The site of the study was the city of Walnut Creek, California, an upper-middle-class suburban community of 50 000 people located approximately 32 km (20 miles) from San Francisco along the route of the Bay Area Rapid Transit (BART) system.

The proposed bus system was designed primarily as a feeder service for peak-hour BART commuters. An additional goal was to serve people traveling to work in downtown Walnut Creek from outlying sections of the city. During the midday off-peak hours, limited service was being considered to serve local nonwork trips, particularly for those for whom automobile travel was not possible. In recent years, similar local transit systems have been started or expanded in a number of suburban communities in the Bay Area.

The methods used in the study were designed to forecast the patronage for the new transit system under different fares, headways, routes, and hours of service. The patronage forecasts made in the study considered only work trips, since this was the primary intended purpose of the system and since choice of travel mode for work trips was the primary research topic for the larger research project of which this study was a component.

## MODELS OF TRANSPORTATION CHOICE

The basis for the forecasts made in this study was made up of two models developed as part of the Urban Travel Demand Forecasting Project (UTDFP), which described how individuals chose among seven alternative travel modes for commuting to work. The travel modes were driving alone, carpooling, bus with walking access, bus with car access, BART with walking access, BART with bus access, and BART with car access. The substance of each model was a set of linear utility expressions that estimated the utility of each of the seven travel modes as a weighted sum of measured variables. The expressions took the form

$$V_{im} = \sum_k B_{km} Z_{ikm} \tag{1}$$

where

$V_{im}$ = the estimated utility of travel mode m for individual i,
$B_{km}$ = coefficient reflecting the estimated influence of variable k on the utility of mode m, and
$Z_{ikm}$ = measured value of variable k for individual i and mode m.

The estimated utilities for each alternative were related to choice probabilities by a logistic function

$$p(m/S)_i = \sum_{n \epsilon s} (e^{V_{mi}}/e^{V_{ni}}) \qquad (2)$$

where $p(m/S)_i$ equals the estimated probability that individual i will choose alternative m from a choice set S of available alternatives. Domencich and McFadden (5) have discussed in detail the underlying assumptions, properties, and estimation techniques for models having this general form.

The particular models used in this study were presented in an earlier report (6) that included a description of their derivation. Two models were used, both of which included the same explanatory variables shown below (the numbers in parentheses indicate the following modes to which the variables were assigned: 1 = driving alone, 2 = bus with walking access, 3 = bus with car access, 4 = BART with walking access, 5 = BART with bus access, 6 = BART with car access, 7 = carpooling):

1. Travel cost divided by post-tax wage, in cents divided by cents per minute (1-7),
2. Auto on-vehicle time (1, 3, 6, 7),
3. Transit on-vehicle time (2-6),
4. Walking time (2-6),
5. Transfer waiting time (2-6),
6. Headway of first transit carrier (2-6),
7. Family income with ceiling of \$7500 (1),
8. Family income under \$7500, with floor of \$0 and ceiling of \$3000 (1),
9. Family income under \$7500, with floor of \$0 and ceiling of \$5000 (1),
10. Number of drivers in household (separate coefficients estimated for 1, 3 and 6, 5, 7),
11. Two-valued variable indicating whether or not the traveler is the head of household (1),
12. Employment density at the traveler's work location (1),
13. Three-valued variable indicating whether the traveler's household was in the central business district (CBD), near the CBD, or otherwise (1),
14. Number of autos owned per household driver, with a ceiling of one (separate coefficients estimated for 1, 3 and 6, 5, 7), and
15. Constant term (separate coefficients estimated for 1, 3-7).

The two models used in this study differed only with respect to the coefficients assigned to some of the variables. In the first model, the calibration of the coefficients did not take into account differences between urban and suburban residents. A single set of coefficients was estimated to apply to a general sample of Bay Area urban and suburban commuters. In the second model, the calibration of coefficients did take into account differences between urban and suburban residents. For selected variables (the first four variables in the list above), separate coefficients were estimated for urban and suburban residents. The coefficients calibrated for the suburban residents were used in this study.

## SURVEY DATA

The study used data from two interview surveys: a telephone interview survey of a random sampling of Walnut Creek households and a questionnaire survey of a random sampling of BART users contacted at the two stations to be served by the proposed local bus system.

### Walnut Creek Telephone Survey

The telephone interview survey was conducted in the fall of 1975 and was designed to include specified numbers of households from each of five geographical areas of the city of Walnut Creek. In each area, households to be called were selected randomly from a reverse telephone directory.

If available when the call was completed, the head of the household was interviewed; otherwise, any available household member was interviewed. This sampling strategy was intended to compensate for the tendencies of heads of household to be home and to answer the phone less frequently than other household adults.

Interviews were completed with 511 respondents representing 75 percent of the households selected for the sample. Of these respondents, 236 worked at least 20 h a week outside the home. Only data for this subsample of workers were used in this study. On the average, the interviews lasted approximately 15 min. A more complete description of this survey, including a reproduction of the interview questionnaire, has been reported elsewhere (7, 8).

### BART Passenger Survey

The BART passenger survey was also conducted in the fall of 1975. A sample of BART users was contacted at each station of the BART system.

Data were collected with a check-answer questionnaire. Interviewers administered the first few questions orally to obtain information about trip origin, destination, and purpose and then handed the partially completed questionnaires to patrons for completion during the trip on BART.

The 529 completed questionnaires represented a 76 percent rate of response when returned by the BART users contacted at the two stations to be served by the proposed Walnut Creek bus system.

Of the people who returned their questionnaires, 156 provided traceable home and work addresses, allowing calculation of transportation supply data, lived within the area to be served by the new bus system, and were using BART for going to work. Data from this subsample were used in the present study.

A more detailed description of the BART passenger survey, including a reproduction of the survey questionnaire, is contained in a 1976 report by the Metropolitan Transportation Commission (9).

## TRANSPORTATION SUPPLY DATA

To forecast the patronage of the new bus system, it was necessary to supplement the interview data with transportation supply data describing the time and cost characteristics of work travel for each person in the supply sample. A separate set of supply variables was calculated for each of the travel modes represented in the forecasting models.

For the purposes of calculating the supply data, each work trip was separated into two segments: the portion inside Walnut Creek (the area served by the proposed bus system) and the portion outside Walnut Creek. Calculations for the two trip segments are described separately.

### Trip Segments Outside Walnut Creek

For the trip portion outside Walnut Creek, calculations were based on existing transportation network data, describing travel between all pairs of 440 transportation zones into which the region was partitioned.

These data were obtained from the Metropolitan Transportation Commission, which is the regional transportation planning agency in the San Francisco Bay Area. The supply data for portions of trips outside Walnut Creek described, for each person, typical trips by auto and BART between the transportation zone in which the person's home was located and the transportation zone in which his or her work place was located.

To increase the accuracy of the auto trip data, zonal averages of parking costs at work were replaced by the parking costs reported by each person in the telephone interview. The latter change was done only for the telephone survey sample; no question on parking cost at work was included in the BART-user sample.

### Trip Segments Within Walnut Creek

The supply calculations for trip portions within Walnut Creek were based on a set of bus routes contained in a transit plan developed by the Walnut Creek Transportation Commission.

### Supply Calculations

The supply calculations were done with a specially devised procedure, based on hand-calculated data, which included map measurements of the walking distance to and from the new bus system for each person in the study sample. The procedure was intended to provide highly accurate supply data with a minimum of map measurement. The calculations for transit travel were based on the specified system of routes for the new bus system. The calculations for auto travel were based on the existing system of roads.

The procedure was essentially a miniature version of a regional transportation network. A system of approximately 30 transportation nodes covering the city was connected to five major destinations within the city, two of which were BART stations. The characteristics of travel from each node to each major destination were calculated by hand for both auto and transit.

In addition to node-to-destination travel characteristics, transit trip calculations included the walking distance from home to the appropriate bus line, based on individual map measurements for each person in the sample. For each person who worked within Walnut Creek, the calculations also included the walking distance measured from the person's work place to the nearest major destination.

When an individual's work trip could be made by more than one route within Walnut Creek (i.e., more than one node to major destination pair) a simple minimum path procedure was used to select a single best route on which the supply calculations were based. A computer was used to combine the individually measured walk distances with the node-to-major-destination travel characteristics and to calculate the minimum path routes.

For the total sample of 369 persons, the map measurements for the supply data calculations required 60 h or approximately 10 min/person.

### Aggregate Forecasts

The survey and supply data provided values for the variables in the seven-alternative models of travel mode choice for each person in the study sample. The models were then used to estimate a probability of bus use for each person in the study sample. To convert these estimated probabilities to aggregate forecasts of bus patronage for the city population, it was necessary to assign a weight to each person in the sample and to calculate a weighted sum of the estimated probabilities.

For the people in the Walnut Creek telephone survey, the weights assigned were used (a) to adjust for differences between the size of the sample and the size of the city population, (b) to adjust for the undersampling of workers living in households with more than one adult—since only one adult per household was interviewed, persons from multiadult households were underrepresented—and (c) to correct for differences in socioeconomic characteristics between the study sample and the city population.

The weighting procedure involved the use of an iterative proportional fitting algorithm and was described in an earlier report (8).

For the people interviewed in the BART station survey, the aggregation weights assigned were used to correct for differences between the size of the sample and the size of the population of BART users at the two stations of interest and to adjust for different sampling ratios of interview respondents to total patrons at these stations during different periods of the day.

The weighting procedure was straightforward; the details are discussed in the report by the Metropolitan Transportation Commission (9).

## TESTS OF THE SEVEN-ALTERNATIVE MODELS

Before the patronage of the new transit system was forecast, tests were carried out to evaluate and compare the predictive validity of the two seven-alternative models. The tests were used to select one of them for use in making the forecasts and to make a rough evaluation of the expected accuracy of the forecasts.

Selecting the most accurate forecasting model from a set of available models is a problem commonly faced in travel demand forecasting. Although information obtained during the calibration of the models provides some indication of relative accuracy, it is often difficult to determine the extent to which the calibration results generalize to a particular forecasting application.

One set of tests used in this study evaluated the ability of each model to predict the travel modes currently used by the study sample for work commuting. A second set of tests compared people's intentions to use the bus system, as reported in the Walnut Creek telephone survey, to the estimated probabilities of using the bus, as calculated with each of the seven-alternative models. Both types of tests were carried out on a person-by-person basis. The two types of tests are described separately below.

### Prediction of Current Mode Choice

For each person interviewed in the Walnut Creek telephone survey, both seven-alternative models were used to estimate the probability of going to work by each of the modes available at the time of the survey: driving, carpooling, BART with walking access, and BART with car access. Then an analysis was done to examine the relationship of these estimated probabilities to actual behavior.

The results of these analyses are summarized in Table 1 for the general model and in Table 2 for the suburban model. Each table contains, for each travel mode, the average estimated probabilities for the people who usually commuted by that mode. Because only four people in the sample usually commuted by BART with walking access, the results for this mode were combined with the results for BART with car access into a single set of results for BART users.

If the models had predicted current behavior perfectly, the tables would contain ones in the diagonal cells

**Table 1. General demand model averages of estimated probabilities of commuting by available modes.**

| | Actual Mode | | Predicted Mode Probability | | | |
|---|---|---|---|---|---|---|
| Usual Mode | No. | % | Driving Alone | Carpooling | BART | Total |
| Driving alone | 143 | 67 | 0.77 | 0.19 | 0.04 | 1.00 |
| Carpooling | 43 | 20 | 0.67 | 0.19 | 0.14 | 1.00 |
| BART | 27 | 13 | 0.61 | 0.17 | 0.22 | 1.00 |
| Total | 213 | 100 | 0.73 | 0.19 | 0.08 | 1.00 |

**Table 2. Aggregate suburban demand model averages of estimated probabilities of commuting by available modes.**

| | Sample Size | | Predicted Mode Probability | | | |
|---|---|---|---|---|---|---|
| Usual Mode | No. | % | Driving Alone | Carpooling | BART | Total |
| Driving alone | 143 | 67 | 0.74 | 0.19 | 0.07 | 1.00 |
| Carpooling | 43 | 20 | 0.61 | 0.18 | 0.21 | 1.00 |
| BART | 27 | 13 | 0.54 | 0.14 | 0.32 | 1.00 |
| Total | 213 | 100 | 0.69 | 0.18 | 0.13 | 1.00 |

and zeros elsewhere, indicating that for the modes usually used the average estimated probabilities were one and that for the modes not usually used the average estimated probabilities were zero. If, on the other hand, the probabilities estimated for each person had been unrelated to the mode usually used, the tables would contain the same entries in each column. Based on this reasoning, a suitable statistic to measure model accuracy is the proportion of the estimated probabilities in a table that are in the diagonal cells. If a model predicted perfectly, this proportion would be one. On the other hand, if the estimated probabilities had had no relationship to behavior, this proportion would be equal to 1/K, where K is the number of available modes (three in this case).

As Tables 1 and 2 indicate, both models performed poorly. In Table 1, for the general model, only 39 percent of the estimated probabilities are in the diagonal cells, only slightly more than the figure of 33 percent that would have occurred if the estimated probabilities had had no relationship to the modes usually used. On an aggregate basis the general model underestimated BART use by a substantial amount; 8 percent of the sample were predicted to commute regularly by BART, while 13 percent actually did.

In Table 2, for the suburban model, the results are only slightly better: 41 percent of the estimated probabilities are in the diagonal cells. On an aggregate basis, however, the suburban model was more accurate than the general model. In particular, the proportion of the sample predicted to commute regularly by BART agreed with the observed proportion of 13 percent.

## Comparisons With Reported Intentions

The second set of tests compared the estimated probabilities that individuals in the study sample would use the proposed bus system as calculated by the two models with the individuals' intentions of using the system as reported in the Walnut Creek telephone survey. Intentions were measured by asking each respondent in the survey how likely he or she would be to use the proposed bus system as part of his trip to work under each of a variety of service combinations. The response categories were "definitely," "probably," "might," "probably not," and "definitely not." The service combinations consisted of different levels of fares (15, 25, and 35 cents), walking distances from home to the bus stop (two, three, and four blocks), and headways between buses (10, 15, 20, and 30 min). Slightly different question formats were used, as appropriate, for respondents who worked within Walnut Creek, worked outside Walnut Creek, and currently commuted by BART.

For each service combination for which reported intentions were obtained, probabilities of bus use were estimated by using each seven-alternative model. Then, for each service combination, the relationships between the estimated probabilities and reported intentions were examined to see if people reporting greater subjective likelihoods of using the system tended to have higher estimated probabilities. In essence, the reported intentions were used as proxies for subsequent behavior to evaluate the accuracy of the probabilities estimated by the model.

This analysis was based on the assumption that reported intentions have substantial predictive validity, that is, there is a substantial relationship between what people say they will do in response to a hypothetical question and what they actually do in a real situation. This assumption requires some justification.

Over the years many researchers have concluded that the predictive validity of reported intentions is low. In particular, when reported intentions have been used to forecast changes in travel behavior, follow-up studies have sometimes found a substantial "non-commitment bias" (10), an apparent tendency of people to overestimate the likelihood of changing their behavior when they are under no commitment to actually do so.

In contrast to these findings, Fishbein and Azjen (11, Chapter 8), after a careful review of numerous studies in which data on reported intentions were related to subsequent behavior on a person-by-person basis, concluded that reported intentions are ordinarily a highly accurate method of predicting specific behavior. They summarized their review by stating that "prediction of single-act criteria is not only possible . . . it is relatively easy. . . . the best predictor of a person's behavior is his intention to perform the behavior."

Fishbein and Azjen's conclusions applied to studies in which the expressed intentions and the follow-up examination of behavior were both concerned with the same specific and clearly defined set of circumstances and corresponding actions. These conditions have frequently not been met in research examining the predictive validity of reported intentions; this includes research concerning changes in travel behavior.

It should be stressed that the appropriateness of the comparisons done in this study did not require that the reported intentions be totally free of error in general or of noncommitment bias in particular. It was merely necessary that the reported intentions have a substantially positive relationship to behavior change, for example, that people who say that they will definitely use the proposed bus under a specific set of circumstances are more likely to actually do so than people who say that they probably, might, or will definitely not do so. Existing evidence seems to indicate that reported intentions are sufficiently accurate for this application.

To measure the relationship between the reported intentions and estimated probabilities, for each of eight service combinations, product-moment correlation coefficients were calculated. For the purposes of these calculations, the response categories for the intentions were assigned consecutive values of from five to one. A zero value was assigned to people for whom BART was impossible (that is, a value of zero was assigned to people who said they could not use the bus as a means of getting to BART because going to work by BART was not possible for them). The results are shown in Table 3. The correlations were low and nearly identical for the two models, ranging from 0.21 to 0.29 for both models.

Similar sets of coefficients, also shown in Table 3, were calculated with the people who said BART was impossible omitted from the data set. Again, the pattern of correlations was nearly identical for the models; in this case, however, none of the correlations for either model was statistically significantly greater than zero; in fact, the correlations were slightly negative.

Thus, the correlations shown in Table 3 indicate that there was very little relationship between the reported intentions and the probabilities estimated by either of the two models. Except for the people for whom BART was impossible, the average estimated probabilities were approximately the same regardless of the intention reported. Actually, the correlation coefficients reflect only linear relationships between the reported intentions and estimated probabilities; however, related analyses, not shown in the table, indicated the absence of any substantial nonlinear relationships as well.

Another perspective from which the reported intentions and estimated probabilities were compared was the sensitivity of the aggregate predictions to changes in the policy variables. To examine this, statistics were calculated to estimate the overall proportion of the sample who would use the bus system under each of the eight service combinations.

For the reported intentions data, two such statistics were calculated. One was the proportion of the sample who said they would "definitely" use the bus. The other was an average probability obtained by assigning a probability value to each person based on reported intention and then averaging these values over the population. The values assigned were 1.00 for definitely, 0.75 for probably, 0.50 for might, 0.25 for probably not, 0.00 for definitely not, and 0.00 for people working outside Walnut Creek who said commuting by BART was impossible for them.

For each of the seven-alternative models, the statistic calculated to indicate aggregate prediction was the average estimated probability of bus use, over the sample.

The values of the aggregate prediction statistics for each of the eight service combinations are presented in Table 4. As the table indicates, the aggregate predictions of bus use based on the reported intentions data

**Table 3. Product-moment correlations between estimated probabilities and reported intentions of bus use.**

| Service Characteristics | | | Sample | | | |
|---|---|---|---|---|---|---|
| | | | All Cases | | BART Possible Only | |
| Fare (¢) | Wait (min) | Walk (blocks) | General Model | Suburban Model | General Model | Suburban Model |
| 25 | 20 | 3 | 0.27 | 0.28 | -0.02 | -0.01 |
| 15 | 20 | 3 | 0.24 | 0.25 | -0.08 | -0.08 |
| 35 | 20 | 3 | 0.21 | 0.21 | -0.06 | -0.05 |
| 25 | 10 | 3 | 0.29 | 0.29 | -0.07 | -0.06 |
| 25 | 15 | 3 | 0.26 | 0.26 | -0.09 | -0.08 |
| 25 | 30 | 3 | 0.24 | 0.23 | -0.00 | -0.01 |
| 25 | 20 | 4 | 0.24 | 0.24 | -0.10 | -0.10 |
| 25 | 20 | 4 | 0.22 | 0.23 | -0.03 | -0.02 |

**Table 4. Estimated aggregate bus use based on reported intentions and mathematical models.**

| Service Characteristics | | | Estimation Method | | |
|---|---|---|---|---|---|
| Fare (¢) | Wait (min) | Walk (blocks) | Reported Intention Composite | Reported Intention "Definitely" | Logit Mathematical Model |
| 25 | 20 | 3 | 0.31 | 0.14 | 0.15 |
| 15 | 20 | 3 | 0.35 | 0.18 | 0.15 |
| 35 | 20 | 3 | 0.20 | 0.08 | 0.15 |
| 25 | 10 | 3 | 0.36 | 0.19 | 0.20 |
| 25 | 15 | 3 | 0.35 | 0.18 | 0.17 |
| 25 | 30 | 3 | 0.15 | 0.04 | 0.11 |
| 25 | 20 | 2 | 0.35 | 0.18 | 0.16 |
| 25 | 20 | 4 | 0.17 | 0.05 | 0.14 |

**Table 5. Estimated probabilities of current BART commuters using local bus.**

| Service Characteristics | | Current Mode | | | | | |
|---|---|---|---|---|---|---|---|
| | | BART/Walk (N = 32) | | BART/CAR (N = 124) | | BART/Total (N = 156) | |
| Fare (¢) | Wait (min) | Probability | Total | Probability | Total | Probability | Total |
| 15 | 10 | 0.26 | 297 | 0.28 | 1154 | 0.27 | 1451 |
| | 15 | 0.25 | 276 | 0.26 | 1071 | 0.26 | 1347 |
| | 20 | 0.23 | 254 | 0.24 | 983 | 0.23 | 1237 |
| | 30 | 0.19 | 208 | 0.19 | 801 | 0.19 | 1009 |
| 25 | 10 | 0.26 | 294 | 0.27 | 1143 | 0.27 | 1436 |
| | 15 | 0.24 | 273 | 0.26 | 1058 | 0.25 | 1331 |
| | 20 | 0.22 | 250 | 0.21 | 970 | 0.23 | 1221 |
| | 30 | 0.18 | 204 | 0.19 | 789 | 0.19 | 992 |
| 35 | 10 | 0.26 | 290 | 0.27 | 1131 | 0.27 | 1421 |
| | 15 | 0.24 | 269 | 0.25 | 1046 | 0.25 | 1315 |
| | 20 | 0.22 | 246 | 0.23 | 958 | 0.23 | 1204 |
| | 30 | 0.18 | 199 | 0.19 | 776 | 0.18 | 975 |

**Table 6. Estimated probabilities of current car commuters using local bus.**

| Service Characteristics | | Work Location | | | | | |
|---|---|---|---|---|---|---|---|
| | | Within Walnut Creek (N = 36) | | Outside Walnut Creek (N = 150) | | Total (N = 186) | |
| Fare (¢) | Wait (min) | Probability | Total | Probability | Total | Probability | Total |
| 15 | 10 | 0.09 | 216 | 0.14 | 1403 | 0.13 | 1619 |
| | 15 | 0.07 | 174 | 0.12 | 1215 | 0.11 | 1390 |
| | 20 | 0.06 | 139 | 0.10 | 1046 | 0.09 | 1185 |
| | 30 | 0.03 | 87 | 0.08 | 756 | 0.07 | 842 |
| 25 | 10 | 0.08 | 205 | 0.14 | 1373 | 0.13 | 1578 |
| | 15 | 0.07 | 165 | 0.12 | 1189 | 0.11 | 1354 |
| | 20 | 0.05 | 131 | 0.10 | 1022 | 0.09 | 1153 |
| | 30 | 0.03 | 82 | 0.07 | 736 | 0.07 | 818 |
| 35 | 10 | 0.08 | 194 | 0.13 | 1345 | 0.12 | 1539 |
| | 15 | 0.06 | 156 | 0.12 | 1163 | 0.11 | 1319 |
| | 20 | 0.05 | 124 | 0.10 | 998 | 0.09 | 1122 |
| | 30 | 0.03 | 77 | 0.07 | 717 | 0.06 | 794 |

were noticeably more sensitive to variations in policy variables than were the aggregate predictions based on either of the seven-alternative models. For example, for the three transit fares of 15, 25, and 35 cents the average probabilities estimated by the models had the same values, 0.15 for the general model and 0.19 for the suburban model, while the proportion of the sample who said they would definitely use the bus varied from 0.08 to 0.18.

Overall, like the attempts to predict current behavior, the tests of the seven-alternative models using reported intentions data were disappointing. On an individual basis, the probabilities estimated by the models showed little relationship to people's reported intentions. On an aggregate basis, the estimated probabilities did not reflect the sensitivity to changes in policy variables revealed by the reported intentions data. To the extent that the reported intentions data can be interpreted as proxies for measures of subsequent behavior, these analyses indicate that the probabilities estimated by the models in this study had questionable validity.

## FORECASTING

The analyses just described created doubts regarding the accuracy of the probabilities of bus use calculated by using either of the seven-alternative models. Nevertheless, forecasts of aggregate patronage for the new bus system, for various combinations of service levels, were made using the methods of data analysis described previously in the report. Because the suburban model was slightly more accurate in the first of the two types of accuracy tests—predicting which of the currently available travel modes were usually used by the people in the study sample—the suburban model was selected for use in forecasting the patronage of the new bus system. Separate patronage forecasts were made for different combinations of fares (either 15, 25, or 35 cents) and headways (either 10, 15, 20, or 30 min).

The results are shown in Tables 5 and 6. The forecasts indicated a substantial latent demand for local bus service, especially as a feeder service for people currently commuting by BART. The proportions of current BART commuters projected to use the bus feeder ranged from 18 to 27 percent, depending on the fares and headways selected. For current drivers, the proportions were lower, ranging from 6 to 13 percent. The variation in the forecasts was chiefly a function of the headway between buses. The forecasts showed little sensitivity to the specified differences in fares.

## CONCLUSIONS

One purpose of the case study was to test the forecasting methodologies to determine their feasibility for making travel demand forecasts in small geographical areas. In general, the implementation of the methodologies was quite successful. The time and expense required for data collection and analysis seemed within reasonable limits for widespread application. Furthermore, the Walnut Creek telephone interviews and the supply variable calculations provided a variety of useful data in addition to that needed for the work-trip forecasts done in this study.

A second purpose of the case study was to evaluate the predictive accuracy of the behavioral models that were used. The tests of predictive accuracy were disappointing. The estimated probabilities of using currently available modes had little relationship to current behavior, and the estimated probabilities of using the proposed bus system bore little similarity to people's reported intentions.

Overall, the results suggest that behavioral models of the type used in this study can be feasibly applied to travel demand forecasting in small geographical areas but that additional development and testing of the models should be done before they are used as a basis for policy decisions in such situations.

## ACKNOWLEDGMENT

This research was supported by the National Science Foundation through grants from the Research Applied to National Needs Program to the University of California, Berkeley.

## REFERENCES

1. J. H. Shortreed and J. A. Ireland. Feeder Bus Service for Commuter Rail. Canadian Journal of Civil Engineering, Vol. 2, No. 4, pp. 530-539.
2. Berkeley Coordinated Transit Development Project. W. Smith and Associates and Curtis Associates, Berkeley, 1976.
3. N. Tahir and M. A. Hovind. A Feasibility Study of Potential Feeder Bus Service for Homewood, Illinois. Proceedings of the 14th Annual Meeting of the Transportation Research Forum, 1973, pp. 553-570.
4. A. Adiv. Existing Methods of Estimating Travel Demand in Small Urban Areas: A Bibliographical Review. Urban Travel Demand Forecasting Project, Institute of Transportation Studies, Univ. of California, Berkeley, Working Paper No. 7710, 1977.
5. T. A. Domencich and D. McFadden. Urban Travel Demand: A Behavioral Analysis. American Elsevier, New York, 1975.
6. K. Train. A Post-BART Model of Mode Choice: Some Specification Tests. Travel Demand Forecasting Project, Institute of Transportation Studies, Univ. of California, Berkeley, Working Paper No. 7620, 1976.
7. M. A. Johnson and A. Adiv. Field Materials and Data Tape Codebook for the 1975 Walnut Creek Transportation Survey. Urban Travel Demand Forecasting Project, Institute of Transportation Studies, Univ. of California, Berkeley, Working Paper No. 7708, 1977.
8. M. A. Johnson and A. Adiv. Forecasting Travel Demand in Small Areas Using Disaggregate Behavioral Models: A Case Study. Urban Travel Demand Forecasting Project, Institute of Transportation Studies, Univ. of California, Berkeley, Final Rept., Vol. 2, 1977.
9. Metropolitan Transportation Commission. 1976 BART Passenger Profile Survey: Field Methods and Processing Procedures. Metropolitan Transportation Commission, San Francisco, 1976.
10. D. T. Hartgen and C. A. Keck. Forecasting Dial-a-Bus Ridership in Small Urban Areas. Planning and Research Bureau, New York State Department of Transportation, Preliminary Research Rept. 60, 1974.
11. M. Fishbein and I. Ajzen. Belief, Attitude, Intention, and Behavior: An Introduction to Theory and Research. Addison-Wesley, Menlo Park, CA, 1975.

*Publication of this paper sponsored by Committee on Passenger Travel Demand Forecasting and Committee on Traveler Behavior and Values.*

# Manual Techniques and Transferable Parameters for Urban Transportation Planning

Arthur B. Sosslau and Maurice M. Carter, COMSIS Corporation, Wheaton, Maryland
Amin B. Hassam,* Peat, Marwick, Mitchell and Company, Washington, D.C.

**This paper summarizes research conducted under the National Cooperative Highway Research Program to identify contemporary transportation policy issues and to evaluate current travel estimation models and procedures in terms of their abilities to respond to such issues. A set of manual techniques and transferable parameters corresponding to the commonly used four-step transportation planning process is described. Brief descriptions are provided for trip generation, trip distribution, mode choice, traffic assignment, time-of-day characteristics, car occupancy factors, capacity analysis, and land development and highway spacing relationships. The travel estimation material developed has been organized in the form of a user's guide, which also includes applications to three scenarios of realistic situations. The manual methods are more advantageous than the computer methods in that transferable parameters allow for quick response in terms of the time required to collect and process local information.**

Much of the emphasis of past urban transportation planning has been put on the development of long-range transportation plans such as the Chicago Area Transportation Study (CATS) of the late 1950s and early 1960s (1). The complex travel-estimating procedures in use then (and now) were designed primarily to evaluate regional transportation systems, in particular highway systems, and to ultimately provide design volumes.

Most initial studies or major updates proceed on a 2- to 3-year time schedule; much of this time is taken up by very costly data collection, data processing, and model calibration. These long-range data-intensive planning processes have often been criticized; their relevance has been questioned; and recommendations have been made for alternative approaches to planning (2, 3).

Recently, however, issues such as energy considerations and the promotion of public transportation modes have taken a much larger role in the planning process. Also, ever-increasing input to the planning process from citizens and elected officials, preparation of environmental impact statements, corridor hearings, and consideration of low-capital and no-build options all demand that the planning process be able to provide analytical support to decision makers in a very short time frame.

Concurrently, emphasis is beginning to shift away from long-range planning to relatively shorter time horizons. Recent papers by Heanue (4) and Manheim (5) have highlighted such a shift in planning strategies.

In light of these trends, it is quite evident that existing planning procedures often fail to permit an analytical response to the various policy issues within the desired time and cost constraints. What is needed are simplified planning methodologies that are easy to understand, relatively inexpensive to apply, and, above all, responsive to the policy issues of the day. It is quite possible to simplify the conventional planning procedures so that more resources can be devoted to other areas of concern (6). Typical improvements on these very involved planning techniques might

1. Avoid dependence on computerized models and instead use manual estimation techniques;
2. Reduce data-collection efforts by utilizing readily available data, transferable parameters, or synthetic models;
3. Analyze regional plans at the district rather than the zone level and focus planning efforts on corridor and subareas; and
4. Estimate travel for only one purpose, and then expand these trips to obtain total travel.

Much work has already been done along these lines. A fair degree of success has been attained, for both highway and transit analysis (7, 8, 9, 10, 11, 12). If such modifications can be applied to the existing planning structures, it would be possible to achieve quicker response capabilities for the travel estimation techniques. This in turn would enable simplified but rapid evaluation of transportation policy alternatives.

On this basis, the National Cooperative Highway Research Program (NCHRP) contracted with the COMSIS Corporation and the Metropolitan Washington Council of Governments to undertake a research and development study to develop manual techniques and transferable parameters for quick response to urban policy issues. The research approach is summarized below.

## RESEARCH APPROACH

The study was conducted in two separately funded phases. Specific objectives, tasks, and results for each phase were as follows.

### Phase 1

The phase 1 research effort involved completion of the following objectives:

1. Objective 1-A: Identification and categorization of contemporary urban policy issues for which travel estimates are required.
2. Objective 1-B: Evaluation of current and emerging travel estimation models and procedures with respect to their ability to satisfy the requirements of policy issues.
3. Objective 1-C: Preparation, on the basis of objectives A and B, of a fully supported set of recommendations for the subsequent phase of the project.

The research approach and results of the phase 1 study have been fully documented (13).

### Phase 2

The results of phase 2 are the focus of this paper. The phase 2 research effort required the following objectives:

1. Objective 2-A: Development of a user's guide to describe transferable parameters and their use with manual and computer techniques for providing quick response travel estimation.

2. Objective 2-B: Identification of areas of potential high payoff for development efforts beyond the scope of the current study.

Details of the phase 2 study are contained in the user's guide (14).

## MANUAL TECHNIQUES AND TRANSFERABLE PARAMETERS DEVELOPED

A set of manual, noncomputerized techniques was developed as a main feature of the phase 2 research project. This set of techniques parallels the classical four-step transportation planning elements of trip generation, trip distribution, mode choice, and trip assignment. The corresponding elements are very similar to procedures used by most transportation planners. However, several shortcuts were made to these elements. To cite a few, model calibration has been eliminated through the use of selected parameters produced from past research studies (e.g., trip rates and friction factors). In some instances, various input data have been minimized by providing estimates from simple nomographs such as zone-to-zone travel times. Overall, the level of applicational effort has been minimized through the provision of step-by-step instructions and simplified work sheets for calculations.

In addition to the four-step components, transferable parameters were provided for the analysis of automobile occupancy, the determination of directional distribution of traffic by time of day, the analysis of highway volume and capacity, and the estimation of facility spacing requirements for alternative land development densities.

### General Capabilities of the Manual Techniques

It is intended that a transportation planner or analyst with a 2- to 3-year experience level can apply the techniques contained in the user's guide. The user can follow these procedures without referring to other sources and with nothing more sophisticated than a hand-held electronic calculator. It is also possible to train a technician to use portions or all of the methods. Ideally, the procedures are most suitable for small-scale transportation projects or localized land-use impacts. Specific projects might include the evaluation of system needs within a single corridor, the assessment of impacts of a transit route extension, or the analysis of increased frequency of transit service. Also, the manual techniques have been designed to adequately address the traffic impacts of a proposed major development on the surrounding street system.

The techniques are also capable of allowing a transportation analysis at the regional level. If a regional analysis is contemplated, it is recommended that the number of analysis areas (zones) be limited to allow application within a reasonable time frame.

The manual methods have proved manageable in application, and it has been found possible to produce reasonable results quite rapidly for many applications (14). For example, the transit demand potential on a single route was estimated in a few hours; spacing requirements based on alternate land development policies were determined within two to three person-days. Further, the transportation impacts of a major residential site were calculated within a week, and a proposed improvement in a corridor was evaluated in about the same time.

In order to fully realize the potential of the manual techniques, it is necessary that the user-planner modify conventional ideas about the planning processes. First, it must be understood that since the manual methods provided are quick and simple, clerical and technical help can be substituted for the computer and computer specialist. Therefore, manual analysis can be very cost-effective. Also, it is anticipated that in most practical cases, through application of the methods, the user would acquire a deeper understanding of the circumstances surrounding the problems than if all comparable work were done by computer. Consequently, his or her ability to clearly present the process and results to clients, elected officials, and the public would be enhanced. Finally, the manual approach stresses simplicity rather than precision in its application and output, thus enabling a larger degree of flexibility and versatility than the computerized planning process does. It must be pointed out, however, that the manual methods are not offered as a replacement for the computer models but rather as an extension of existing analysis techniques.

### Use of Transferable Parameters

Recent transportation research has revealed that certain parameters, factors, and relationships from one study area can quite satisfactorily suffice when transferred to another area having similar characteristics (8, 10, 15, 16). In the NCHRP project study, therefore, every effort was made to capitalize on these conclusions. A large array of transportation data was compiled for use as "default" values. Where more pertinent local information is not available, or where collection of new data is not warranted, these transferable values can be useful in manual and computer applications. This material, which has been supplied in the user's guide, is in the form of tables, charts, nomographs, and formulas. In the development of such manually applicable information, data consistency was maintained throughout. That is, wherever possible and appropriate, the parameters have been grouped together and reported for the four urban population groups—50 000 to 100 000 people, 100 000 to 250 000 people, 250 000 to 750 000 people, and 750 000 to 2 000 000 people—and the three trip purposes—home-based work (HBW) trip, home-based nonwork (HBNW) trip, and non-home-based (NHB) trip.

## MANUAL TECHNIQUES AND TRANSFERABLE PARAMETERS

As mentioned above, the manual techniques and transferable parameters developed have been documented in the user's guide (14). The following sections highlight the capabilities of the major travel estimation components of the transportation planning process contained in the user's guide.

### Trip Generation

Numerous reference sources (15, 16, 17, 18, 19, 20) were used to develop the trip-generation characteristics provided in the user's guide. The information retrieved from this review was compiled into tables and graphs representing (a) average vehicle trip rates and other trip characteristics of generators, (b) detailed trip-generation characteristics by household income (Table 1), and (c) generalized trip-generation parameters for trip productions and attractions.

By knowing the percentage of households by income group or auto ownership per household (for an analysis zone), it is possible to arrive at the estimate of

**Table 1. Trip-generation characteristics for an urban population of 100 000-250 000.**

| Income, 1970 ($000s) | Average Autos Per Household | Average Daily Person Trips Per Household | Households Owning No. of Autos (%) | | | | Average Daily Person Trips Per Household by No. of Autos | | | | Average Daily Person Trips by Purpose[a] (%) | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | 0 | 1 | 2 | 3+ | 0 | 1 | 2 | 3+ | HBW | HBNW | NHB |
| 0-3 | 0.49 | 4.0 | 57 | 37 | 6 | 0 | 1.0 | 7.5 | 10.5 | 13.8 | 20 | 63 | 17 |
| 3-4 | 0.72 | 6.8 | 36 | 56 | 8 | 0 | 1.7 | 9.2 | 13.3 | 16.4 | 22 | 60 | 18 |
| 4-5 | 0.81 | 8.4 | 29 | 61 | 10 | 0 | 2.5 | 10.2 | 14.5 | 17.6 | 22 | 58 | 20 |
| 5-6 | 0.94 | 10.2 | 21 | 65 | 13 | 1 | 3.5 | 11.4 | 14.5 | 19.0 | 22 | 58 | 20 |
| 6-7 | 1.01 | 11.7 | 17 | 66 | 16 | 1 | 4.5 | 12.5 | 15.6 | 20.5 | 20 | 58 | 22 |
| 7-8 | 1.14 | 13.6 | 12 | 65 | 21 | 2 | 5.4 | 13.8 | 17.0 | 22.2 | 20 | 57 | 23 |
| 8-9 | 1.25 | 15.3 | 9 | 61 | 28 | 2 | 5.8 | 15.0 | 17.5 | 23.0 | 20 | 57 | 23 |
| 9-10 | 1.34 | 16.2 | 6 | 58 | 33 | 3 | 6.3 | 15.8 | 18.0 | 23.5 | 19 | 57 | 24 |
| 10-12.5 | 1.50 | 17.3 | 4 | 50 | 40 | 6 | 6.8 | 16.0 | 19.0 | 24.5 | 19 | 57 | 24 |
| 12.5-15 | 1.65 | 18.7 | 2 | 40 | 51 | 7 | 7.0 | 16.0 | 20.4 | 25.0 | 19 | 56 | 25 |
| 15-20 | 1.85 | 19.6 | 2 | 28 | 57 | 13 | 7.2 | 15.0 | 21.0 | 25.5 | 18 | 56 | 26 |
| 20-25 | 2.01 | 20.4 | 1 | 20 | 61 | 18 | 7.5 | 15.0 | 21.0 | 25.5 | 18 | 55 | 27 |
| 25+ | 2.07 | 20.6 | 1 | 19 | 59 | 21 | 7.5 | 15.0 | 21.0 | 25.2 | 18 | 55 | 27 |
| Weighted average | 1.55 | 14.5 | 14 | 48 | 33 | 6 | 5.4 | 13.7 | 18.4 | 22.4 | 20 | 57 | 23 |

[a]Source is Baerwald (41).

the average daily person-trips by purpose for that zone by using Table 1, for the 100 000-250 000 urban population group.

## Trip Distribution

Various trip-distribution methods were investigated for transformation to manual application (21, 22, 23). Since the gravity model (GM) has been the most widely used technique, the model was structured to operate in a manual environment. The conversion required a streamlining of its mode of operation—for instance, calibration of the model friction factors for the four urban population groups and the three trip purposes was totally eliminated by using other information (24). Also, the socioeconomic (K) factors in the computer GM formulation were discarded altogether, since these cannot be handled efficiently manually. One major assumption was that the interzonal travel-time matrix, which has to be developed for input to the GM, is triangular; that is, the travel time from zone i to zone j is the same as that from j to i.

Input data to the GM consist of the balanced productions and attractions by zone, the interzonal travel times obtained from the travel-time matrix, and the corresponding friction factors. In order to perform the GM calculations efficiently, a simplified work sheet was designed. To assess the time requirements for conducting trip distribution at the regional level, the manual GM was tested at a 34 × 34-district "real" example for Atlanta, Georgia. The entire trip distribution process (i.e., developing the interdistrict and intradistrict travel times and the corresponding friction factors, and undergoing two iterations) required approximately 26 person-hours to complete using an electronic desk calculator with memory.

The manual GM was also applied in a 19 × 19-district, three-purpose "site development impact scenario" for Boise, Idaho, and an 18 × 18-zone, two-purpose "corridor analysis scenario" for Columbus, Ohio. The Boise scenario required 14 person-hours for the HBW trip distribution, and a total of 19 person-hours for the HBNW and NHB distribution. The Columbus HBW trip distribution was completed in approximately 20 person-hours.

An empirical relationship was formulated between the time required to carry out manual trip distribution versus the number of analysis areas. This was done to allow the user-planner to estimate the applicational time requirements.

Manual trip distribution probably constitutes the most time-consuming element of the manual procedures provided in the user's guide. But, overall, manual trip distribution was found to be quite manageable and accurate and compares reasonably well with computerized applications. Manual trip distribution is recommended for up to 50 analysis areas.

Other important and useful material developed and provided in the user's guide for the trip distribution phase included nomographs for the development of zone-to-zone travel-time and friction-factor matrixes for the four population groups (Figure 1), gravity model travel-time exponents for three urban population groups and five trip purposes, a method for distributing trips around a site by reversing productions and attractions, the use of accessibility indexes (once computed from a manual GM application) for quick determination of interzonal trip interchanges, and trip-distribution patterns for selected generator sites.

The use of the travel-time and friction-factor nomographs warrants some discussion. Essentially, in-vehicle travel times can be derived by first measuring the zone centroid-to-centroid airline distance on a map, then estimating the proportion of travel on arterials or freeways, next determining the distances traveled in each subregion (CBD or central city or suburbs), and last entering nomographs such as the one illustrated in Figure 1. Appropriate nomographs must be selected according to whether the travel is totally within a subregion or across two or three subregions. The nomographs also provide the origin-destination (O-D) terminal times, which, when added to the in-vehicle travel time, result in the total O-D travel time. The user can then read the corresponding friction factors for each of the three trip purposes (HBW, HBNW, NHB).

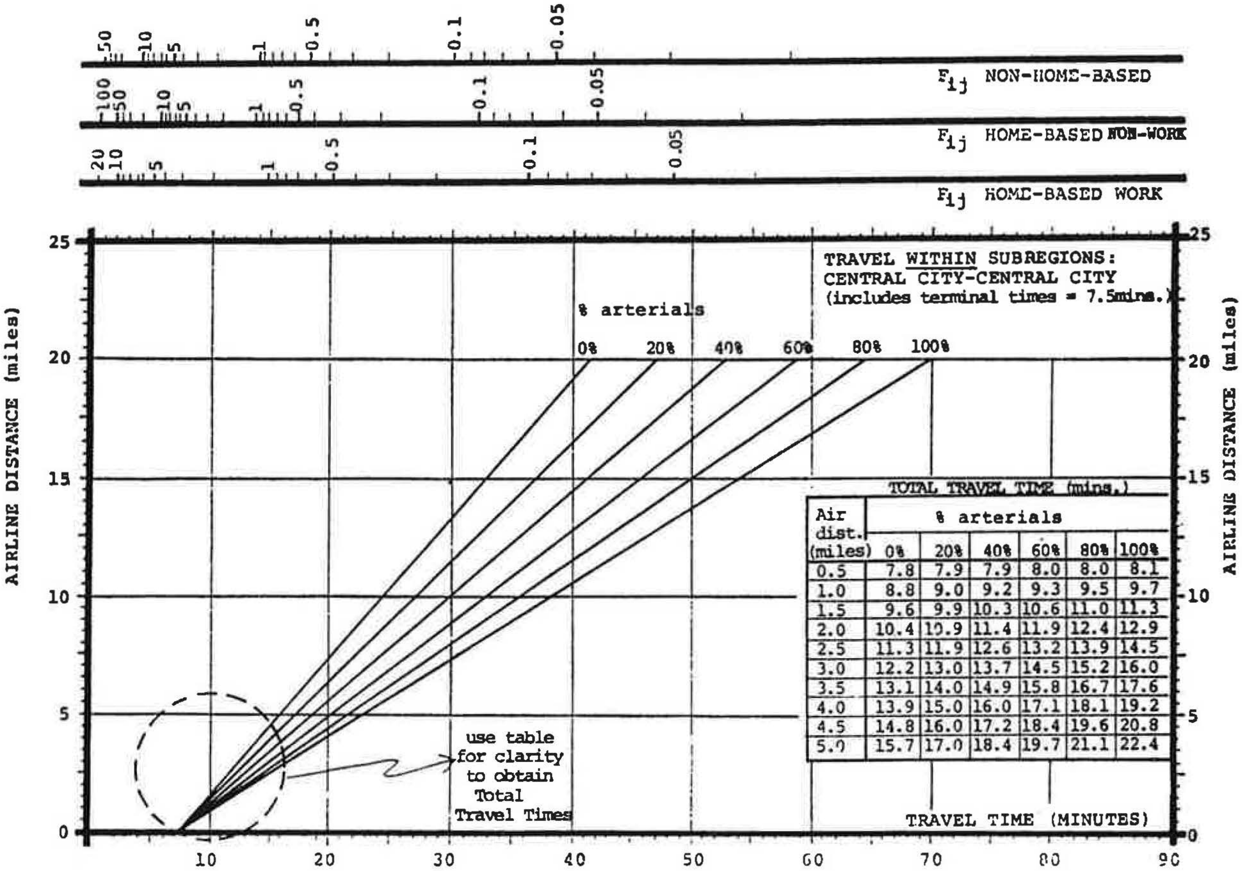
In summary, these nomographs constitute a set of practical tools for determining the travel-time and friction-factor matrix. The user's guide provides instructions for the planner to allow construction of these nomographs to suit particular local conditions, if so desired.

## Mode Choice

A thorough literature review (25, 26, 27) was undertaken to identify mode-choice models having potential for manual conversion. Ultimately, the Urban Mass Transportation Administration default model contained in the program UMODEL (14) was selected for transformation to the manual format. The default model is a simultaneous logit model for trip distribution and modal split.

The transformation described in the user's guide and by Carter (28) converts the logit model into a very sim-

**Figure 1. Airline distance versus travel time versus distribution factors by trip purpose for an urban population of 100 000-250 000.**




TOTAL TRAVEL TIME (mins.)

| Air dist. (miles) | % arterials 0% | 20% | 40% | 60% | 80% | 100% |
|---|---|---|---|---|---|---|
| 0.5 | 7.8 | 7.9 | 7.9 | 8.0 | 8.0 | 8.1 |
| 1.0 | 8.8 | 9.0 | 9.2 | 9.3 | 9.5 | 9.7 |
| 1.5 | 9.6 | 9.9 | 10.3 | 10.6 | 11.0 | 11.3 |
| 2.0 | 10.4 | 10.9 | 11.4 | 11.9 | 12.4 | 12.9 |
| 2.5 | 11.3 | 11.9 | 12.6 | 13.2 | 13.9 | 14.5 |
| 3.0 | 12.2 | 13.0 | 13.7 | 14.5 | 15.2 | 16.0 |
| 3.5 | 13.1 | 14.0 | 14.9 | 15.8 | 16.7 | 17.6 |
| 4.0 | 13.9 | 15.0 | 16.0 | 17.1 | 18.1 | 19.2 |
| 4.5 | 14.8 | 16.0 | 17.2 | 18.4 | 19.6 | 20.8 |
| 5.0 | 15.7 | 17.0 | 18.4 | 19.7 | 21.1 | 22.4 |

ple modal choice formula given by

$$MS_{t(ij)} = [I_a^b/(I_a^b + I_t^b)]\,100 \tag{1}$$

where

$MS_{t(ij)}$ = market share percentage on transit for any ij zone pair,
$I_a$ = auto impedance for the ij zone pair,
$I_t$ = transit impedance for the ij zone pair, and
b = exponent of time (similar to gravity model travel-time exponents).

The user's guide provides a nomograph based on Equation 1 for each of the three trip purposes (see Figure 2 for HBW trips). Once the auto and transit impedances have been calculated for any ij pair, the nomograph can be entered to arrive at the market share percentage of transit. The user has an option of using localized values of b for the specific urban area under study. The auto and transit impedances are computed by using special nomographs drawn from procedures used elsewhere (12). Basic input information such as highway and transit airline distances and auto-operating and parking speeds is necessary for the application of these graphs.

Some other practical tools supplied in the user's guide for mode-choice analysis include a nomograph for converting highway airline distances to average operating speed, simplified worksheets for calculating auto and transit trips, and simple rules of thumb for quick estimates of transit demand.

The mode-choice technique was tested using travel data from Washington, D.C., and Atlanta, Georgia. Good results were obtained and have been documented (28). The overall success of the manual mode-choice procedure prompted San Diego to incorporate the technique for nonwork travel analysis in a current transportation study. Also, Atlanta has replaced its previous mode split models with a computerized procedure using this technique.

## Auto Occupancy

Two major data sources utilized for auto occupancy factors and relationships are found elsewhere (28, 29). In addition, numerous urban transportation studies were reviewed. A series of tables delineating the variations in average daily auto occupancy rates with respect to other exogenous factors was developed. Typical tables included in the user's guide deal with auto occupancy rates by each of the four urbanized area population groups and by trip purpose (Table 2), auto occupancy rates by income level of trip maker and parking cost at trip destination, auto occupancy rates by urban population and land use at destination, auto occupancy adjustment factors by time of day, and auto occupancy adjustment factors by trip length.

The user's guide also presents several illustrative examples to accustom the user to the application of the tables.

## Time-of-Day Distribution

The majority of the manual techniques and parameters contained in the user's guide are based on average daily travel conditions. For an analysis of particular highway facilities, transit routes, and other related work, peak-period or specified hour demand estimation is often necessary. The time-of-day analysis information provided permits various types of conversion.

The material, in the form of tables, explicitly recognizes the characteristics of travel by time of day according to location within the study area (CBD, central city, or suburb) and to orientation of the facility in relation to the core area (radial or cross-town). Facilities considered are freeways and expressways, arterials, and collectors. Much of the material developed here has been obtained from another study (30).

For example, the following relationships and procedures have been incorporated in the user's guide: hourly distribution (a) of internal driver travel by each of the four urban population groups and by trip purpose, (b) of internal driver and total vehicle travel by urban population, and (c) of total travel on various highway facilities by urbanized area population; and conversion factors (a) for critical time periods of internal person travel by urban population (see Table 3) and (b) for critical time periods of transit patronage.

These factors might prove particularly handy for traffic impact analyses, trip-purpose mix studies, and, in view of the critical role of transportation system management (TSM) requirements, for such a management study.

**Figure 2. Mode-choice nomograph.**

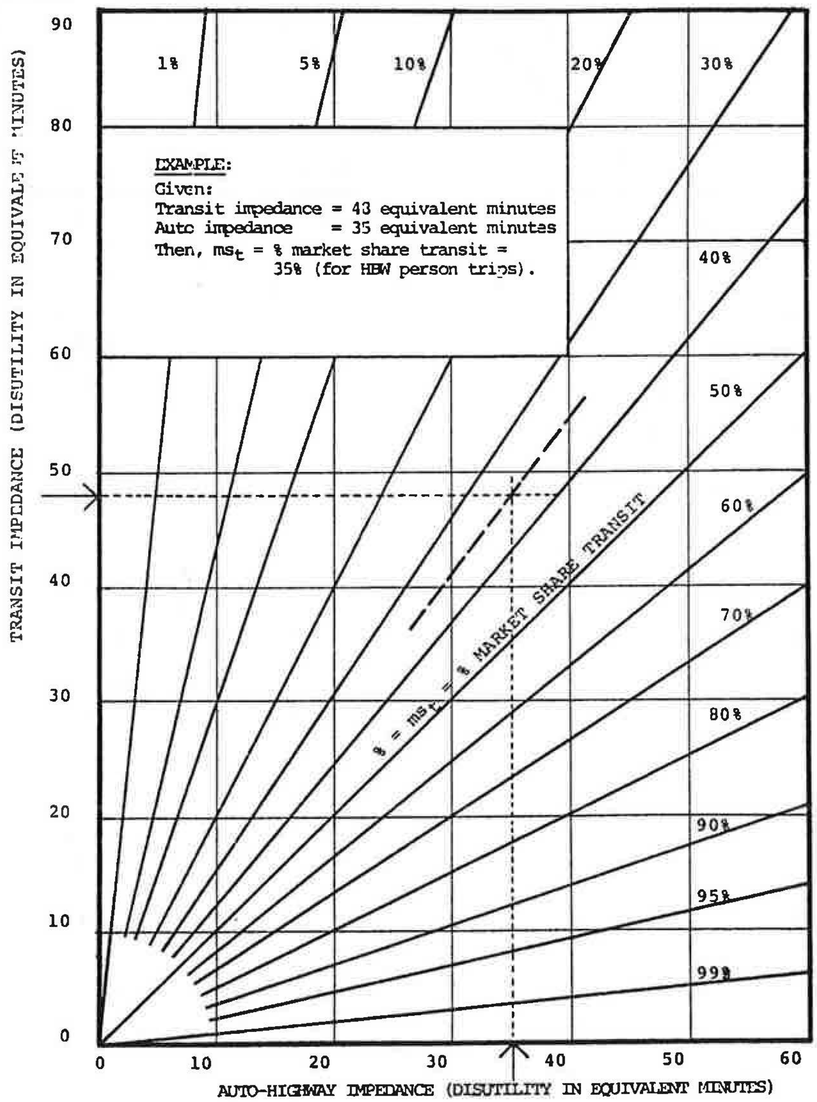



**Table 2. Average daily auto occupancy rates by urban area and trip purpose (42).**

| Urban Population | Trip Purpose: HBW | HB Shopping | HB Social-Recreational | HB Other | HBNW[a] | NHB | All Purposes[b] |
|---|---|---|---|---|---|---|---|
| 50 000-100 000 | 1.38 | 1.57 | 2.31 | 1.52 | 1.82 | 1.43 | 1.50 |
| 100 000-250 000 | 1.37 | 1.57 | 2.31 | 1.52 | 1.81 | 1.43 | 1.50 |
| 250 000-750 000 | 1.35 | 1.57 | 2.30 | 1.52 | 1.77 | 1.43 | 1.50 |
| 750 000-2 000 000 | 1.33 | 1.58 | 2.29 | 1.51 | 1.74 | 1.43 | 1.51 |

[a]Weighted average of auto occupancy rates for HB Shop, HB Social-Recreational, and HB Other trip purposes.
[b]Weighted average of auto occupancy rates for all trip purposes.

**Table 3. Conversion factors for critical periods of internal auto travel for urban population of 100 000-250 000.**

| | Travel Hours | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Travel Type | Total | Total Work[a] | Combined Peak Period | Peak-Hour Total | Combined Peak-Period Work | Morning Peak-Period Work | Evening Peak-Period Work | Peak-Hour Work |
| Total | | 0.200 | 0.322 | 0.099 | 0.116 | 0.062 | 0.053 | 0.039 |
| Total work | 5.000 | | 1.610 | 0.495 | 0.579 | 0.312 | 0.267 | 0.195 |
| Combined peak-period total | 3.106 | 0.621 | | 0.307 | 0.360 | 0.194 | 0.166 | 0.121 |
| Peak-hour total | 10.101 | 2.020 | 3.253 | | 1.170 | 0.630 | 0.539 | 0.394 |
| Combined peak-period work | 8.621 | 1.727 | 2.781 | 0.855 | | 0.539 | 0.461 | 0.337 |
| Morning peak-period work | 16.026 | 3.205 | 5.160 | 1.587 | 1.856 | | 0.856 | 0.625 |
| Evening peak-period work | 18.727 | 3.745 | 6.029 | 1.854 | 2.169 | 1.169 | | 0.730 |
| Peak-hour work | 25.641 | 5.128 | 8.256 | 2.538 | 2.969 | 1.600 | 1.369 | |

[a]Work refers to HBW trips. Total is (HBW + HBNW + NHB) trips. See text for definitions of travel for the various time periods.

## Trip Assignment

After a comprehensive literature review on existing trip assignment methodologies, three manual assignment techniques were selected for inclusion in the user's guide. The first is the traditional all-or-nothing assignment process (7, 31, 32). Major modifications of this commonly used method included the assumption that minimum time paths can be selected by judgment; then, a procedure for smoothing assigned volumes between a set of parallel facilities (33) was provided. Finally, simplified work sheets were designed to systematically keep track of the resulting trip volumes.

The second method was generally guided by a report by Gruen Associates (34). This method enables the estimation of traffic generation and attenuation and the corresponding highway facility requirements such as number of lanes and spacing. Improvements on this method permit the use of more specific estimates of generated trips, for example, by employing Table 1, by using a more responsive decay function, and by providing directional sensitivity.

The third procedure is based upon the multiroute probabilistic process developed by Dial (35). The manual formulation presented in the user's guide provides a means of determining the probable shifts (divisions) in volumes between competing facilities in a corridor.

Examples of some of the products resulting from these three techniques include simplified assignment work sheets, a series of charts for estimating street requirements based on land use, and a graph for determining traffic shifts between facilities in a corridor.

## Capacity Analysis

Capacity analysis addresses the question of how much system is required to satisfy the estimated travel demand or how much traffic the existing street system can accommodate before intolerable congestion develops. Two types of techniques are included in the user's guide for analyzing capacity. First, a corridor analysis procedure is described to investigate volume-to-capacity (V/C) conditions within a highway corridor and to profile these relationships along a corridor route and, second, an intersection analysis procedure to evaluate vehicle movements through intersections (36).

The corridor approach draws upon and extends existing procedures for analysis at screenlines and cutlines. The approach is to analyze V/C conditions in an aggregate sense at key points along a corridor.

The intersection analysis method utilizes turning and through lane movements to determine the critical volume of an intersection. It is presumed that such an intersection capacity analysis would be used if a user were investigating the impacts of a site on local street conditions. The technique requires trip assignment, including the tabulation of turning movements at an intersection.

Using capacity information (37), several manually applicable tables and nomographs were constructed for use in the V/C analysis. Examples of some are generalized capacity measures for freeways, expressways and arterials and capacity nomographs for one- or two-way streets, with or without parking.

## Land Development Density and Highway Spacing Analysis

The basic purpose of the land-use and highway spacing relationships described in the user's guide is to permit the rapid development of a "first-cut" estimate of future highway needs based on a desired level of highway service. Given a distribution of land use in a study area, either in terms of activities (people, households, jobs) or subarea by type of use, and given the presence of an existing highway system, future vehicle trip ends are computed and then adjusted for improved transit service. Next, the average trip distance is computed from counts or from curves provided and adjusted for the future.

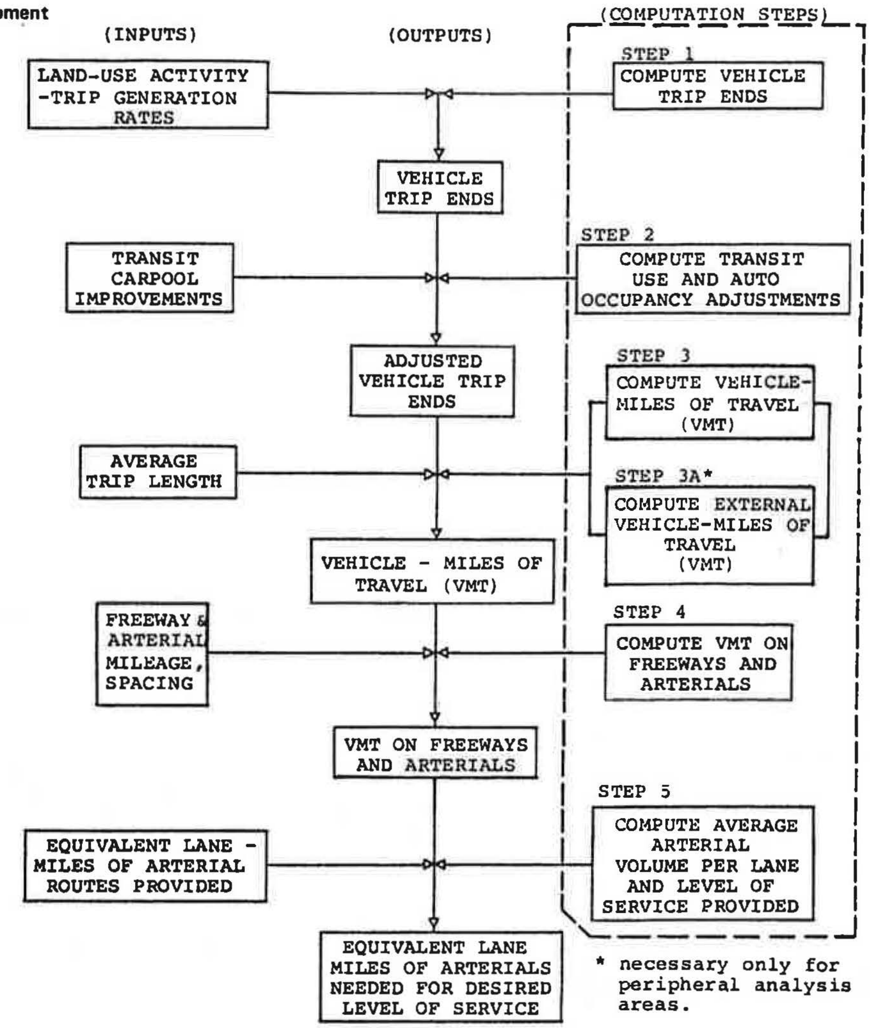
Average arterial volumes, by subarea, for a given spacing of freeways and arterials can then be determined from the computation of vehicle-kilometers of travel and the level of service provided. A comparison of the level of service with a desired level gives a measure of highway needs for the study area. A description of the method in flowchart form is shown in Figure 3.

Some of the analytic techniques in the form of graphs and charts, based on other sources (38, 39, 40), include a graph for least-cost volumes for various facilities, graphs for determining freeway and arterial spacing based on the magnitude of travel, and information on level-of-service volumes for various facilities.

## SCENARIO APPLICATION OF THE MANUAL ESTIMATION TECHNIQUES

To ascertain the capabilities of the manual methodologies and transferable parameters described in the user's guide, extensive applications were made to three differ-

Figure 3. Diagram of development density and highway-spacing methodology.




ent types of transportation scenarios. Each of these scenarios was based on authentic field conditions and data input obtained from the three study areas: Boise, Idaho; Columbus, Ohio; and Fairfax County, Virginia. The choice of these study areas was dictated by and based on their population variations (small, medium, and large, respectively) and their geographical distribution.

The Boise scenario was based on the investigation, for the year 2000, of traffic impacts on the surrounding highway system of a proposed residential development and a large shopping center. Almost all of the manual techniques contained in the user's guide were put to use to quantify these effects. All three trip purposes were analyzed, the analysis itself yielding satisfactory results. The entire investigation—trip generation through trip assignment and capacity analysis—required a total of 60 person-hours. It was estimated that if only HBW trips were developed and then expanded to total trips (using factors such as those contained in Table 3), the work effort would have been reduced to about 40 person-hours.

The object of the Columbus scenario was to determine, for the year 2000, the impacts of a proposed corridor development located on the outskirts of the region and the current growth. Again, most of the manual estimation techniques described above were used for the corridor impact analysis. The scenario was conducted in about 66 person-hours and produced output that was in reasonable agreement with local forecasts.

The Fairfax County scenario determined the base year and future year (1985) levels of service provided by the current and planned transportation systems in the county. The manual techniques described in the user's guide were used to estimate present and future travel, to allocate this travel by subarea to freeway and arterial facilities, and then to compute the resulting levels of service. The scenario required approximately 22 person-hours and produced acceptable results.

## CONCLUSIONS

This paper has presented a brief summary of the research effort undertaken to identify contemporary urban policy issues, to evaluate currently available methods and procedures, and to develop manual travel estimation techniques and transferable parameters. On the basis of the test applications, it is believed that these manual methods are applicable to many transportation planning problems. Further, the manual methods will result in time and cost savings for various applications when compared to computer-oriented solutions.

Since the final report (14) was only recently distributed, the manual techniques have not yet undergone widespread testing and application. We hope this will occur as the techniques are put to use. Plans are under way to develop instructional material for use in training sessions. These sessions, similar in nature to the

Highway Capacity Manual workshops, will be conducted to assist state and local planners in the application of the numerous techniques contained in the user's guide. The Transportation Center of the University of Tennessee will assist COMSIS in the implementation of this phase.

For the manual procedures to achieve full potential and acceptance, additional experimentation is needed. It would also be worthwhile to extend the analytic and estimation features of the techniques summarized in this paper, and to conduct further research to develop other noncomputerized transportation planning techniques for which a need exists. Such techniques could prove useful in responding to the ever-changing issues of the day in shorter time frames.

## ACKNOWLEDGMENTS

The research summarized in this paper was sponsored by the American Association of State Highway and Transportation Officials in cooperation with the Federal Highway Administration. The study was conducted under the National Cooperative Highway Research Program. The research effort for the NCHRP projects on travel estimation procedures for quick response to urban policy issues was undertaken jointly over a period of 3 years by the COMSIS Corporation and the Metropolitan Washington Council of Governments (WASHCOG). Participants in the project for COMSIS included Mark E. Roskin and Martin J. Fertal, and for WASHCOG included George V. Wickstrom, Robert T. Dunphy, and Robert E. Griffiths. We take this opportunity to thank the many individuals and organizations for their cooperation in providing information and support to the project, especially the state and local agencies that provided assistance in the scenario applications in Boise, Idaho; Columbus, Ohio; and Fairfax County, Virginia. A special word of thanks is extended to Robert E. Spicher for his constant encouragement and direction throughout the course of this study.

## REFERENCES

1. Chicago Area Transportation Study. Chicago, IL, 3 Vols., 1959-1960.
2. G. V. Wickstrom. Are Large-Scale Home-Interview Origin-Destination Surveys Still Desirable? Paper presented at the 50th Annual Meeting of the TRB, Jan. 1971.
3. H. S. Levinson. Simplifying the Comprehensive Transportation Planning Process—Prospects and Perspectives. Paper presented at the 51st Annual Meeting of the TRB, Jan. 1972.
4. K. E. Heanue. Changing Emphasis in Urban Transportation Planning. Paper presented at the 56th Annual Meeting of the TRB, Jan. 1977.
5. M. L. Manheim. The Emerging Planning Process: Neither Long-Range Nor Short-Range, but Adaptive and (Hopefully) Decisive. Paper presented at the 56th Annual Meeting of the TRB, Jan. 1977.
6. A. B. Sosslau and G. V. Wickstrom. Quick Response to Urban Transportation Issues. Paper presented at the ASCE Annual Convention, New Orleans, Apr. 1975.
7. DeLeuw Cather Canada Ltd. A New Procedure for Urban Transportation Planning. Paper prepared for the Canada Department of Highways, Toronto, Sept. 1969.
8. H. M. Hajj. Synthesis of Vehicle Trip Patterns in Small Urban Areas. HRB, Highway Research Record 369, 1971, pp. 181-198.
9. T. Hillegass, C. Fleet, and K. E. Heanue. A Proposal for a Simplified Urban Transportation Modeling Process. U.S. Department of Transportation, Federal Highway Administration, 1973.
10. S. Khasnabis and M. R. Poole. Synthesizing Travel Patterns for a Small Urban Area. Traffic Engineering, Vol. 45, No. 8, Aug. 1975.
11. K. W. Heathington. Simplitied Procedures for Preliminary Evaluation of Public Transportation Alternatives. Paper presented at the 56th Annual Meeting of the TRB, Jan. 1977.
12. DeLeuw, Cather and Company. Transit Corridor Analysis—A Manual Sketch Planning Technique. Prepared for U.S. Department of Transportation, Urban Mass Transportation Administration, May 1976.
13. COMSIS Corporation. Travel Estimation Procedures for Quick Response to Urban Policy Issues. Final Rept., prepared for NCHRP, June 1977.
14. COMSIS Corporation. Quick-Response Urban Travel Estimation Manual Techniques and Transferable Parameters: A User's Guide. Prepared for NCHRP, Nov. 1977.
15. L. E. Keefer. Urban Travel Patterns for Airports, Shopping Centers, and Industrial Plants. NCHRP, Rept. 24, 1966.
16. L. E. Keefer and D. K. Witheford. Urban Travel Patterns for Hospitals, Universities, Office Buildings, and Capitols. NCHRP, Rept. 62, 1969.
17. Comparison of Virginia Urban Trip Generation Studies With Similar Investigations Conducted by the States of Maryland and California. Virginia Department of Highways, Metropolitan Transportation Planning Division, Richmond, VA, July 1972.
18. Trip Generation—An Informational Report. ITE, Committee 6A6, Arlington, VA, 1976.
19. Trip Generation by Land Use, Part I: A Summary of Studies Conducted. Maricopa Association of Governments, Transportation and Planning Office, Phoenix, AZ, Apr. 1974.
20. COMSIS Corporation. Trip Generation Analysis. Prepared for U.S. Department of Transportation, Federal Highway Administration, Aug. 1975.
21. T. J. Fratar. Vehicular Trip Distribution by Successive Approximations. Traffic Quarterly, Jan. 1954.
22. M. Schneider. Panel Discussion on Inter-Area Travel Formulae. HRB, Bulletin 253, June 1960.
23. Calibrating and Testing a Gravity Model for Any Size Urban Area. U.S. Department of Commerce, Bureau of Public Roads, Oct. 1965.
24. Urban Trip Distribution Friction Factors. U.S. Department of Transportation, Federal Highway Administration, 1974.
25. M. J. Fertal and others. Modal Split—Documentation of Nine Methods for Estimating Transit Usage. U.S. Department of Commerce, Bureau of Public Roads, Dec. 1966.
26. R. H. Pratt. A Utilitarian Theory of Travel Mode Choice. HRB, Highway Research Record 322, 1970, pp. 40-53.
27. P. R. Rassam, R. H. Ellis, and J. C. Bennett. The n-Dimensional Logit Model: Development and Application. HRB, Highway Research Record 369, 1971, pp. 135-147.
28. M. M. Carter. Development of a Simple Logit Choice Model. Paper presented at the 56th Annual Meeting of the TRB, Jan. 1977.
29. Estimating Auto Occupancy: A Review of Methodology. U.S. Department of Transportation, Federal Highway Administration, no date.
30. Peat, Marwick, Mitchell and Company. An Analysis of Urban Area Travel by Time of Day. Pre-

pared for U.S. Department of Transportation, Federal Highway Administration, Jan. 1972.
31. COMSIS Corporation. Traffic Assignment. U.S. Department of Transportation, Federal Highway Administration, Aug. 1973.
32. R. J. Paquette, N. Ashford, and P. H. Wright. Transportation Engineering: Planning and Design. Ronald Press, New York, 1972.
33. R. H. Pratt and Associates. A Method for Distributing Traffic Volumes Across Cutlines. Kensington, MD, 1976.
34. Gruen Associates. Land Use and Arterial Spacing in Suburban Areas. Prepared for U.S. Department of Transportation, Federal Highway Administration, May 1977.
35. R. B. Dial. A Probabilistic Multipath Traffic Assignment Model Which Obviates Path Enumeration. Prepared for U.S. Department of Transportation, Federal Highway Administration, May 1970.
36. H. McInerney and S. Peterson. Intersection Capacity Measurement Through Critical Movement Summations: A Planning Tool. Traffic Engineering, Vol. 41, No. 1, Jan. 1971.
37. Highway Capacity Manual. HRB, Special Rept. 87, 1965.
38. R. Creighton and others. Estimating Efficient Spacing for Arterials and Expressways. HRB, Bulletin 253, 1960.
39. A. Haikalis and H. Joseph. Economic Evaluation of Traffic Networks. HRB, Bulletin 306, 1961.
40. System Considerations for Urban Freeways. ITE, Washington, D.C., Informational Rept., Oct. 1967.
41. J. E. Baerwald, ed. Transportation and Traffic Engineering Handbook. Prentice-Hall, Englewood Cliffs, NJ, 1976.
42. Nationwide Personal Transportation Study. U.S. Department of Transportation, Federal Highway Administration, Repts. 1 and 8, Apr. 1972.

*Publication of this paper sponsored by Committee on Passenger Travel Demand Forecasting and Committee on Traveler Behavior and Values.*

**Mr. Hassam was with the COMSIS Corporation when this research was being conducted.*

# Tabulating Demand Elasticities for Urban Travel Forecasting

Y. Chan and F. L. Ou, Pennsylvania Transportation Institute, Pennsylvania State University, University Park

**This paper presents a compendium of demand elasticities in a tabulated form in order to facilitate urban travel forecasting. A number of elasticity estimates have been reported for a variety of cities over the past decade, but the scenarios or base conditions differ from one site to another. In order to systematically tabulate these disparate estimates, demand elasticities were pooled into four cells according to urban size (large versus medium) and urban structure (core-concentrated versus multinucleated). Such a classification has been verified to stratify cities into groups sharing common socioeconomic and travel patterns. Demand elasticities can be divided into two categories: empirical elasticities and calibrated elasticities. The former were measured in the field before and after notable incidents such as a fare increase in the transit system, while the latter were derived from demand models. The elasticities can be further identified as either aggregate or disaggregate depending on whether they are calculated from areawide or subarea data. All these result in a collection of elasticities that have rather different values. This paper tries to explain some of these differences to gain insights into the general characteristics of elasticities for urban areas of different sizes and structures. The elasticity tabulation and the general properties of the elasticities provide both practitioners and researchers with factual information for estimating urban travel demand simply and systematically.**

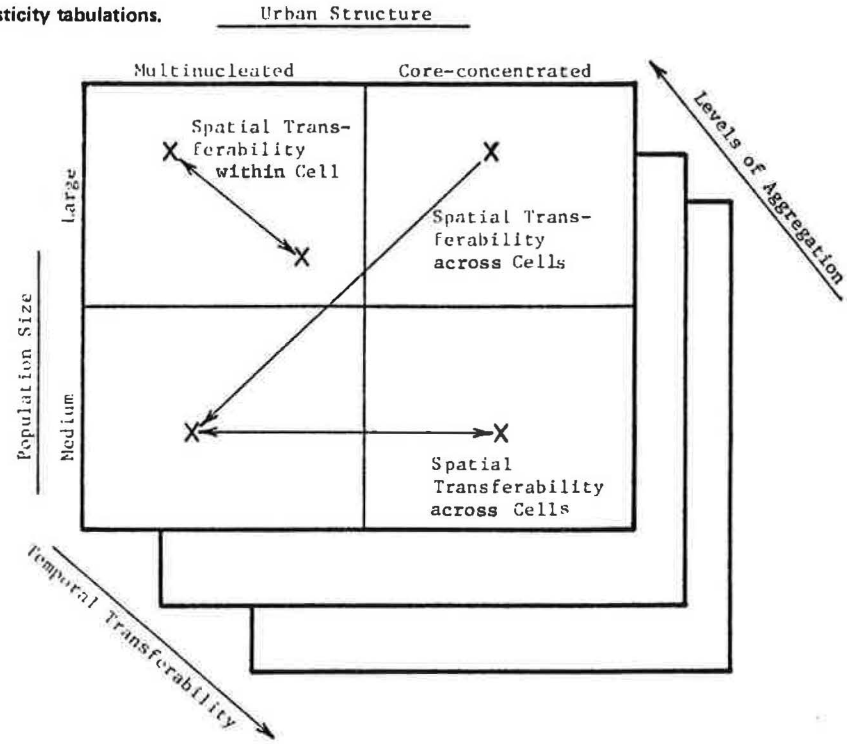
Demand elasticities are often used in conjunction with urban travel forecasting. They have been applied frequently, however, under circumstances that are inconsistent with the assumptions under which they were derived. The purpose of this paper is to resolve some of these inconsistencies and to provide some guidelines—including a systematic tabulation of the available elasticities—for their consistent application in demand estimation.

There are three areas where inconsistencies may be introduced. First, elasticities are often applied in a scenario very different from the base conditions from which they were empirically developed. For example, a fare elasticity of -0.13 measured during the New York subway fare increase of January 1970 refers specifically to the base conditions that existed at that time, including the patronage and fare level. To apply the elasticity indiscriminately for other fare and patronage levels is a futile exercise at best. Unfortunately we found many cases where elasticities are cited out of context and, hence, erroneous inferences are drawn.

Demand elasticities found in a large metropolis such as New York City provide little information on other cities either smaller or of similar size, since they may have drastically different urban structures. Very limited research has been performed in relating elasticities to cities classified according to size and other urban characteristics. Until a better understanding of such a relationship is gained, our knowledge about elasticities in specific sites cannot help us in demand forecasting in other cities.

The measurement of elasticities was performed by using methods ranging from areawide empirical tabulations to disaggregate demand modeling. These various levels of aggregation can often lead to very different estimates of demand elasticities for the same study area. A case study in Chicago, for example, shows that the difference between areawide and household elasticities can be as high as 40 percent, depending on the homogeneity of travel behavior among households in the area (_1_, Appendix 8). Citing an elasticity without specifying the level of aggregation can therefore result in estimates significantly out of kilter with reality. All these conditions point to the fact that guidelines for applying demand elasticities need to be found. The way the elasticities

Figure 1. Conditions in elasticity tabulations.




are compiled and categorized in this paper is, in our judgment, a step in this direction (Figure 1).

TRANSFERABILITY OF ELASTICITIES

Since elasticity measures are defined for a particular base condition, they cannot be directly applied to a scenario with a different base condition unless steps are taken to guarantee their transferability. In order to apply elasticities to travel forecasting, a group of elasticities must be compiled for cities that share common socioeconomic and travel characteristics. Such a stratification may explain some of the variations in elasticities among areas.

Spatial Transferability

A stratification scheme according to city size is intuitively appealing, since the travel patterns are found to be different among large, medium, and small cities. Travel demand is also found to be affected by the urban structure. For example, the number of trips is probably greater when employment and shopping centers are dispersed than when they are concentrated (which results in a major flow of traffic to and from the city center). A stratification of cities into multinucleated versus core-concentrated categories is therefore advisable when explaining the variations in travel demand.

We set up a hypothesis to group the U.S. cities with at least 50 000 population into four cells: (a) large core-concentrated, (b) large multinucleated, (c) medium core-concentrated, and (d) medium multinucleated.

Several experiments conducted (1) to verify or improve such a hypothesis have utilized techniques such as cluster analysis, factor analysis, regressions, and linear goal programming. The results indicate that such a classification scheme, for the data assembled from 55 percent of the U.S. cities, is statistically significant in explaining the variations in the base travel conditions among the cities within the same cell. When 800 000 population is used as a demarkation between large and medium cities, cities in the same cell of the classification are found to share common relationships on a key urban travel parameter: person-trip hours of travel (PHT). This finding is rather gratifying, since PHT, aside from being a measurement of travel intensity, is the product of travel volume and trip impedance that captures the base conditions for which an elasticity is defined. [Demand elasticity, or the percentage of change in travel in response to 1 percent change in such travel impedance as trip time or cost, is defined as ($\Delta$ volume/base volume)/($\Delta$ impedance/base impedance).] Such a result leads us to believe that the travel responses, or elasticities, are similar among the city groups under the taxonomy scheme. In the following sections, then, we shall discuss the outcomes of classifying elasticities according to city size and urban structure as defined in our city stratification scheme.

Temporal Transferability

The four-cell stratification may guarantee spatial transferability of elasticities among cities in the same cell, enabling the elasticities from one city to be applied in another city (belonging to the same group). However, there is still another major consideration before a table of elasticities can be used in a meaningful way, and that is the problem of temporal transferability (Figure 1), which we explain below.

The use of tabulated elasticities as a measurement of demand changes, with respect to changes in travel time and travel cost, is predicated on the hypothesis that these elasticities are stable over time. If we define $\eta_b$ as the demand elasticity of base year b, and $\eta_f$ for the forecast year f, the assumption of temporal transferability amounts to equating $\eta_b$ with $\eta_f$. Obviously such an assumption needs to be verified.

As pointed out earlier, demand elasticities are derived for a specified volume of demand and trip impedance (or level of service), which are collectively referred to as the base travel conditions. Kannel and Heathington (2) examined the form of household travel relations to determine the stability of these relations over time. The results indicate that the trip-making volumes estimated from the 1964 data could successfully predict household travel in 1971.

A study by Voorhees (3) indicates that travel impedances such as trip time can be estimated from city population as a temporally stable relationship. Examinations by Chan of the time series data collected from 1956 through 1976 for 55 percent of the U.S. cities show reasonable temporal stability among both travel volume and travel impedance as represented by trip time (1).

An intuitive explanation of these stability properties can also be offered. Recent investigations into travel decisions and time budgets suggest that an individual spends a relatively constant percentage of his normal day on travel, which implies that the PHT (frequency and duration of trip making) are relatively stable.

Aside from the base conditions, temporal transferability is also reported for demand elasticities. McFadden (4) and Train (5), in their works with the San Francisco Bay Area Rapid Transit (BART), calibrated modal choice models before the implementation of the rapid transit system in order to forecast BART ridership. All the predicted shares turned out to be within one standard error of the corresponding observed shares after the implementation of BART. The forecasting error in total BART patronage amounts to only 2.3 percent. The research by Atherton and Ben-Akiva (6) provides more evidence of temporal transferability in the time span from 1963 to 1968. While the details of the findings may vary, however, we feel that there is enough evidence of temporal and spatial transferability of elasticities using the proposed urban size and structure stratification to warrant more detailed investigation.

## ELASTICITY TABULATIONS

Having established a way to group cities, let us tabulate the elasticities reported in the literature. These can be classified into two major categories: empirical elasticities and calibrated elasticities. Empirical elasticities are obtained from field measurements before and after a notable incident such as a fare increase. Calibrated elasticities, on the other hand, are derived from various types of demand models.

According to the methodology and data used for the computation, elasticities can also be divided into two categories: aggregate versus disaggregate, where the latter approach is founded on more detailed data (such as an individual or household unit of analysis), while the former approach is based on coarser data (such as the areawide geographic unit). Empirical elasticities, measured typically on a corridor or areawide basis, are therefore more aggregate than calibrated elasticities, often on a zone or household level basis.

According to the above taxonomy, elasticities can be classified as (a) those calibrated by aggregate data using aggregate methodology ($\eta_{AA}$) and (b) those calibrated by disaggregate data using a disaggregate methodology ($\eta_{DD}$).

It is hypothesized that $\eta_{DD}$ tends to overestimate the magnitude of demand response, while $\eta_{AA}$ will underestimate demand elasticities, on the average. The relationship can be expressed by the following inequality:

$$|\eta_{DD}| > |\eta_{real}| > |\eta_{AA}| \quad (1)$$

The reason for the above relationship is that observation errors tend to impart downward bias to aggregate parameters, while the elasticity obtained by aggregating over $\eta_{DD}$ is likely to yield higher values (7). Without further empirical studies one cannot know precisely the quantitative nature of the inequality. It is one of the objectives of this paper to address this issue by reporting on some preliminary findings on a limited set of empirical (areawide) and calibrated (subareal) elasticities.

In the following tabulations, both aggregate and disaggregate demand elasticities will be further stratified by trip purpose (whether it be working, shopping, social and recreational, or others) and a variety of impedance or level-of-service attributes such as time and cost. For example, an excess time elasticity may be defined for work trips, while a fare elasticity may be defined for nonwork travel.

### Empirical Elasticities

An empirically derived set of elasticities is compiled for the transit ridership experiences of 19 cities in the United States. For each city, the elasticities over the years corresponding to the various fare increases (such as in New York City) are recorded. Some of the measurements are performed during the peak hours and some during off-peak hours (such as in St. Louis), but most of the numbers are reported for overall trips without stratification into trip purposes.

The variation in these empirical measurements ranges from -0.07 to -3.80, although all of them are transit demand elasticities with respect to the levels of service. An arrangement of these elasticities according to the four-cell stratification scheme is shown in Tables 1 and 2 (8, 9, 10, 11, 12, 13, 14, 15). After the site and level-of-service attributes are specified, the largest variation in each of the entries is from -0.63 to -3.8, which is the range recorded for the excess time elasticities in large core-concentrated cities. This represents a substantial improvement over the -0.07 to -3.80 range found in the original data before classification. We view this as a partial illustration of the effectiveness of the stratification scheme for our elasticity tabulation.

### Calibrated Elasticities

Two groups of calibrated elasticities will be compiled: those calibrated using zone data models and those calibrated using household data models. Shown in Table 3 (16) and Table 4 (6, 17, 18, 19, 20, 21, 22, 23, 24) are the value ranges of the calibrated elasticities obtained by using zone direct-demand models and household modal-split models. It should be noted that among all trip purposes, only the most complete data set, work elasticities, is reported here for conciseness. Elasticities for other trip purposes are reported elsewhere (1) and will not be reproduced here.

The usefulness of Tables 3 and 4 (and for that matter Tables 1 and 2) can be illustrated via a simple example. The linehaul time elasticity for bus patronage, for instance, ranges from -0.78 to -0.30 for a large core-concentrated city, which may be interpreted to say that, in general, the ridership goes down by about -0.54 percent (the average of the two extremes) in response to a one percent rise of the linehaul travel time.

The reader is reminded that some of these rather wide ranges of the elasticity values as shown in Tables 1-4 can be attributed to several factors. First, they are obtained from different cities with disparate base levels of service and travel volumes. Second, they are obtained from both areawide and subarea (zone or household) estimations that may yield rather different elasticities even for the same study area. Thus far, we have had some success in explaining the variations among cities by using the city classification scheme. The variations due to the level of aggregation, however, remain to be explained.

## Comparison Between Calibrated and Empirical Elasticities

The ad hoc manner in which empirical elasticities were compiled and the severe limitations of the data base make a rigorous comparison with the calibrated values difficult. The fact that these two sets of elasticities are generally compiled for different trip purposes (overall versus work, respectively) further complicates the matter. Finally, the different data base (before and after versus cross-sectional, areawide versus subarea) constitutes the last straw. However, a discussion is still useful as a check as long as we keep in mind that work demand elasticities are more inelastic by nature than elasticities for overall trip purposes.

Amid these complications an experimental relationship between the empirical (areawide) elasticity $\eta_{AA}$ and the calibrated (subarea) elasticity $\eta_{OD}$ was discerned. It is encouraging to find that the magnitude of the corresponding numbers in Tables 1-4 confirms our initial conjecture (Equation 1) that elasticities estimated from a set of more disaggregate data are higher in value than those estimates from a more aggregate set, as can be explained below.

In Figure 2, the same elasticities estimated from empirical versus model calibrations are put on the same plot. It appears that the calibrated elasticities show up consistently higher in value than the empirical ones. While there may be several factors that contribute to this, the level of aggregation could be the dominant factor. The reason is that preliminary analysis shows that the ratios of work to overall elasticity range from 1.0 to 1.3, with the average of the two extremes being 1.15. The slope of regression line is 1.97 in Figure 2. The difference between 1.97 and 1.15 has to be attributed to the level of data aggregation.

**Table 1. Range of empirical transit demand elasticities for overall trips for medium cities.**

| Item | Medium Multinucleated Cities Transit[a] | | | Medium Core-Concentrated Cities Transit[b] | | |
|---|---|---|---|---|---|---|
| | Total | Bus | Rail | Total | Bus | Rail |
| Total | | | | | | |
| Linehaul time | NA | NA | NA | NA | NA | NA |
| Excess time | NA | NA | NA | NA | NA | NA |
| Cost | NA | NA | NA | NA | NA | NA |
| Bus | | | | | | |
| Linehaul time | NA | NA | NA | NA | NA | NA |
| Excess time | NA | NA | NA | NA | -0.83 | NA |
| Cost | NA | -0.12 ~ -0.34 | NA | NA | -0.25 ~ -0.65 | NA |
| Rail | | | | | | |
| Linehaul time | NA | NA | NA | NA | NA | NA |
| Excess time | NA | NA | NA | NA | NA | NA |
| Cost | NA | NA | NA | NA | NA | NA |

[a]Salt Lake City and Springfield, Massachusetts.
[b]Chesapeake, Virginia; Portland, Maine; Tulsa; and York, Pennsylvania.

## GENERAL CHARACTERISTICS OF ELASTICITIES

When one regards urban transportation as a service to the consumer, one can assign a generalized price to purchasing such a service in terms of cost and time. Since demand elasticity is a measure of the percentage of responsiveness of travel demand to 1 percent change in the level of service, it evaluates the marginal demand contributions of a change in cost or time. It is found in our tabulations that, while individual travel demand elasticities in a class of urban areas often share some common characteristics, the elasticities among groups of cities are disparate. The difference between demand elasticities among different groups of urban areas can be explained below.

### Saturation of Demand

According to the theory of demand, the marginal utility of additional trips tends to diminish when demanded trips approach served trips. This explains why, in the larger urban areas where served trips often fall short of demanded trips, demand is highly sensitive to change in level of service. The results of three case studies in Boston (17), Chicago (18), and Louisville (16) strongly support this theory. Bus demand elasticities with respect to linehaul time, excess time, and cost are -1.10, -1.84, and -0.51, respectively, for a large city such as Chicago and -0.19, -0.38, and -0.40, respectively, for a smaller city such as Louisville. The comparison of auto demand elasticity with respect to its own linehaul time and cost between a large city and a medium city also shows this tendency. For instance, auto linehaul time and cost elasticities for Boston are -0.82 and -0.494, while for Louisville they are -0.39 and -0.12.

### Level-of-Service Attributes

According to consumer behavior, when the price of a commodity or a service is high, the response of demand to a change in price is more elastic. This leads to a corollary that states that elasticities are lower in value where level of service is high. In smaller cities where

**Table 2. Range of empirical transit demand elasticities for overall trips for large cities.**

| Item | Large Multinucleated Cities Transit[a] | | | Large Core-Concentrated Cities Transit[b] | | |
|---|---|---|---|---|---|---|
| | Total | Bus | Rail | Total | Bus | Rail |
| Total | | | | | | |
| Linehaul time | NA | NA | NA | | NA | NA |
| Excess time | NA | NA | -0.90(10) | | NA | NA |
| Cost | -0.30(5) | NA | NA | | NA | NA |
| Bus | | | | | | |
| Linehaul time | NA | }-0.55 | NA | | NA | NA |
| Excess time | NA | -0.20 ~ -0.60 | NA | | -0.63 ~ -3.8 | NA |
| Cost | NA | -0.11 ~ -0.64 | NA | | -0.08 ~ -0.60 | NA |
| Rail | | | | | | |
| Linehaul time | NA | NA | NA | | NA | NA |
| Excess time | NA | NA | NA | | NA | -0.24 |
| Cost | NA | NA | NA | | NA | -0.07 ~ -0.19 |

[a]Atlanta, Boston, Detroit, Philadelphia, San Diego, and San Francisco.
[b]Baltimore, Cincinnati, Milwaukee, New York, and St. Louis.

**Table 3. Range of calibrated work-trip elasticities for medium cities.**

| Transportation System Variable | Medium Multinucleated City[a] Para-transit | Transit Total | Transit Bus | Transit Rail | Auto | Medium Core-Concentrated City[b] Para-transit | Transit Total | Transit Bus | Transit Rail | Auto |
|---|---|---|---|---|---|---|---|---|---|---|
| Paratransit | | | | | | | | | | |
| Linehaul time | | | NA | | NA | -0.315 | | 0.194 | | 0.194 |
| Excess time | | | NA | | NA | -0.189 | | 0.117 | | 0.117 |
| Cost | | | NA | | NA | -0.057 | | 0.035 | | 0.035 |
| Total | | | | | | | | | | |
| Linehaul time | | | NA | | NA | NA | | NA | | NA |
| Excess time | | | NA | | NA | NA | | NA | | NA |
| Cost | | | NA | | NA | NA | | NA | | NA |
| Bus | | | | | | | | | | |
| Linehaul time | | | }-2.80 (Total time) | | }1.11 | 0.022 | | -0.19 ~ -0.373 | | 0.022 |
| Excess time | | | NA | | NA | 0.092 | | -0.38 ~ -1.55 | | 0.092 |
| Cost | | | -0.51 | | 0.06 | 0.05 | | -0.40 ~ -0.852 | | 0.05 ~ 0.15 |
| Rail | | | | | | | | | | |
| Linehaul time | | | NA | | NA | NA | | NA | | NA |
| Excess time | | | NA | | NA | NA | | NA | | NA |
| Cost | | | NA | | NA | NA | | NA | | NA |
| Auto | | | | | | | | | | |
| Linehaul time | | | ]2.81 (Total time) | | ]-1.11 | 0.173 | | 0.173 | | -0.138 ~ -0.39 |
| Excess time | | | NA | | NA | 0.173 | | 0.173 | | -0.138 |
| Cost | | | 1.39 | | -0.55 | 0.115 | | 0.115 | | -0.092 ~ -0.12 |

[a] Richmond, Virginia. [b] Louisville, Kentucky.

**Table 4. Range of calibrated work-trip elasticities for large cities.**

| Transportation System Variable | Large Multinucleated Cities[a] Para-transit | Transit Total | Transit Bus | Transit Rail | Auto | Large Core-Concentrated City[b] Para-transit | Transit Total | Transit Bus | Transit Rail | Auto |
|---|---|---|---|---|---|---|---|---|---|---|
| Paratransit | | | | | | | | | | |
| Linehaul time | -0.02 ~ -0.664 | NA | -0.27 | NA | NA | -0.59 | | 0.22 | NA | NA |
| Excess time | -0.122 | NA | NA | NA | NA | -0.28 | | 0.10 | | NA |
| Cost | -0.01 ~ -0.092 | NA | -0.06 | NA | NA | -0.10 | | 0.04 | | NA |
| Total | | | | | | | | | | |
| Linehaul time | NA | -0.20 ~ -0.39 | NA | NA | NA | NA | | NA | | NA |
| Excess time | NA | -0.69 ~ -0.709 | NA | NA | NA | NA | | NA | | NA |
| Cost | NA | -0.09 ~ -0.58 | NA | NA | NA | NA | | NA | | NA |
| Bus | | | | | | | | | | |
| Linehaul time | 0.04 | NA | -0.46 ~ -1.10 | 0.23 | 0.04 ~ 0.14 | NA | | -0.30 ~ -0.78 | | 0.25 |
| Excess time | 0.10 | NA | -0.17 ~ -2.28 | 0 ~ 0.06 | 0.05 ~ 0.373 | NA | | -0.94 | | 0.30 |
| Cost | 0.03 | NA | -0.1 ~ -0.58 | 0 ~ 0.28 | 0.02 ~ 0.138 | NA | | -0.12 ~ -0.20 | | 0.06 |
| Rail | | | | | | | | | | |
| Linehaul time | NA | NA | 0.13 ~ 1.02 | -0.60 ~ -0.80 | 0.10 | NA | | NA | | NA |
| Excess time | NA | NA | 0.03 ~ 1.15 | -0.12 ~ -2.06 | 0.02 | NA | | NA | | NA |
| Cost | NA | NA | 0 ~ 0.25 | -0.86 ~ -1.80 | 0.13 | NA | | NA | | NA |
| Auto | | | | | | | | | | |
| Linehaul time | NA | 0 ~ 0.37 | 0.36 ~ 0.39 | 0.27 ~ 0.41 | -0.02 ~ -0.82 | NA | | 0.17 ]~0.53 | NA | -0.18 |
| Excess time | NA | 0 | NA | NA | -0.027 ~ -1.437 | NA | | 0.33 ] | NA | -0.35 |
| Cost | NA | 0 ~ 0.80 | 0.06 ~ 0.97 | 0.06 ~ 0.97 | -0.01 ~ -0.494 | NA | | 0.15 | NA | -0.16 |

[a] Boston, Chicago, San Francisco, Los Angeles, San Diego, Minneapolis-St. Paul. [b] Washington, D.C.

the level of service is generally higher (meaning that the travel time is shorter and the travel cost is lower), demands are therefore less elastic. However, for headway elasticities where the headway is longer (hence the level of service lower) for smaller cities the reverse is true.

The examples used in explaining the saturation of demand support this assertion. Additional evidence can be found in the case studies on bus demand elasticities in New Bedford, Massachusetts; San Francisco; Washington, D.C.; and Los Angeles (1, 22), where the bus ridership in large cities is more sensitive to changes in travel time.

Another example of transit schedule frequency can be

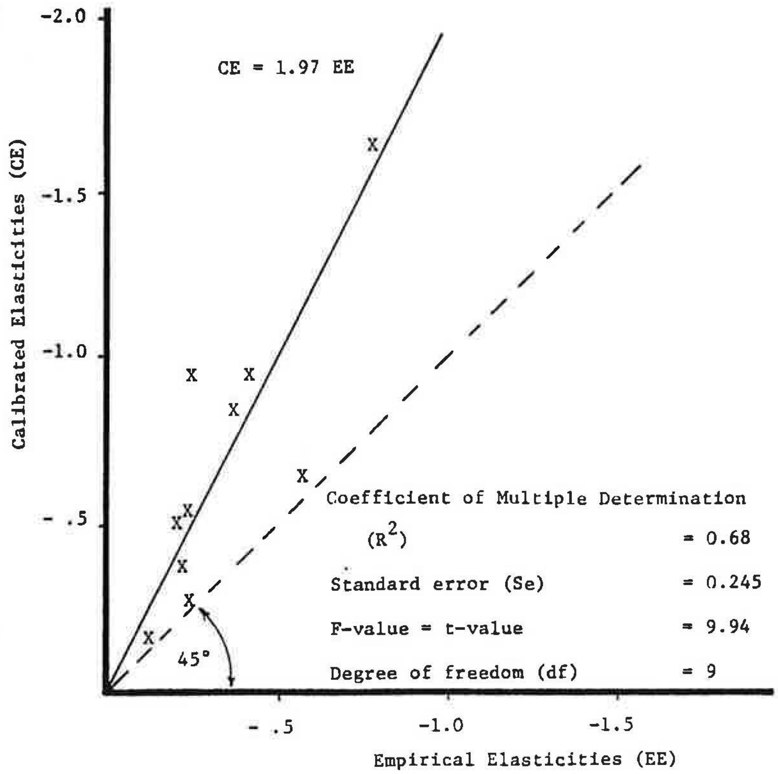



**Figure 2. Comparison of mean values and empirical and calibrated elasticities.**

cited. Large cities generally have higher levels of service in terms of schedule frequency. The headway change of buses from 5 to 10 min has less impact on ridership compared to the same percentage of change in medium cities from 15 to 30 min. This is shown in the empirical findings of Chesapeake, Virginia, a medium-sized city, and several large cities such as Boston, New York, and Detroit (but not Milwaukee). The headway elasticity for the former is -0.83, and the headway elasticities for the latter are -0.60, -0.20, and -0.63. This reasoning is further supported by empirical findings in New York, where ridership impacts corresponding to changes in the schedule frequencies of buses and subways were measured (Tables 1 and 2). The bus (with lower level of service) has an elasticity of -0.63, while the subway (with higher level of service) has an elasticity of -0.24.

### New Mode

Based on modal choice theory, the demand elasticities are more elastic if the trip maker has more than one choice of mode. This is shown in two sets of observations. In the BART study (22), the bus demand elasticities for linehaul time, excess time, and cost are -0.46, -0.17, and -0.45 respectively for pre-BART, and -0.60, -0.19, and -0.58 for post-BART, while the auto demand elasticities with respect to changes in linehaul time and cost are -0.13, -0.32, and -0.22, -0.47 corresponding to two different points in time.

### Summary

Due to the difference of social and economic backgrounds, different areas have different levels of saturation of demand, levels of service, mode choices, and other variables. These lead to the variation of demand elasticities among areas. To some extent, the city classification scheme employed in our analysis accounts for a number of these factors and helps to explain some of the elasticity variations. The size of the city is used to group cities into different levels of demand saturation and level of service. The urban structure, on the other hand, separates core-concentrated cities (where there are generally more choices of mode) from multinucleated cities (where there are fewer mode choices).

If one examines the elasticities in each of the four groups of cities, he or she will notice that, generally speaking, the numerical values are larger in large cities and sites with a core-concentrated urban structure. This constitutes a verification of the two-way classification scheme.

## CONCLUSIONS

The main purpose of this paper is to translate all of the reported demand elasticities into a consistent form that will be useful in demand forecasting. Of special concern are the spatial and temporal transferabilities of these elasticities, in which the former means that parameters calibrated in one area can be applied to other areas, while the latter implies that parameters are stable over time.

We emphasize that demand elasticity alone, being a point estimate, cannot be used to forecast effectively. This is because elasticity is defined only for a particular base condition, which is often characterized by a particular level of service and traffic volume. In order to apply the elasticities obtained in one city to another, one should group cities according to similar socioeconomic and travel patterns. The patterns are collectively referred to as the base conditions.

An analysis was performed to uncover some of the base conditions under which the elasticities were derived, and a city classification scheme was found for grouping cities into generic cells within which elasticities may be transferable. Elasticities were then tabulated for the respective cells according to the stratification of modes and level-of-service attributes. It was found that a good deal of the variations among the time and cost elasticities can be explained by the city classification scheme, which groups cities according to their level of demand saturation, level of service, and choice of modes.

At the same time, the variations among the values obtained for the same elasticity may be attributed to the different level of data aggregation. Empirical elasticities, which are measured from areawide or corridor data, provide the lower bound estimate, while calibrated elasticities, which are obtained from zone or household data, give the upper bound estimate. The true value of elasticities lies somewhere in between the two extremes.

Like many other studies of this nature, this research can be refined by expanding the data base and continuing the methodological investigations. In the interim, there are some very positive contributions reported in this paper. For the practitioners, it provides a handbook of elasticities, thus reducing the replication of demand forecasting model calibration efforts, and serves as a tool to perform travel estimation in a fast and consistent manner using the available elasticity tabulations. For the researchers, it offers some insights into the general properties of elasticity measurement in American cities of different urban size and urban structure.

## ACKNOWLEDGMENTS

Financial support for the work presented in this paper came from the Office of the Secretary of Transportation under its University Research Program. Moral and intellectual support was provided by C. Chamberlain and E. Weiner. Other professionals to whom we wish to express our appreciation include D. Brand, R. Dial, R. Dobson, D. Dunlap, D. Gendell, J. Hamburg, K. Heanue, T. Hillegass, J. Horowitz, I. Kingham, D. Ward, and P. Watson. While we are the chief contributors to the research results reported here, the other two members of the research team, J. Perl and E. Regan, collaborated closely throughout the one-year study.

## REFERENCES

1. Y. Chan, F. L. Ou, J. Perl, and E. Regan. Review and Compilation of Demand Forecasting Experiences: An Aggregation of Estimation Procedures. Pennsylvania Transportation Institute, Pennsylvania State Univ., University Park, Final Rept., PTI 7708, 1977.
2. E. J. Kannel and K. W. Heathington. Temporal Stability of Trip Generation Relations. HRB, Highway Research Record 472, 1973, pp. 17-27.
3. Alan M. Voorhees and Associates. Factors and Trends in Trip Length. NCHRP, Rept. 48, 1968.
4. D. McFadden. The Theory and Practice of Disaggregate Demand Forecasting for Various Modes of Urban Transportation. Paper presented at the Seminar on Emerging Transportation Planning Methods, Daytona Beach, FL, 1976.
5. K. Train. A Post-BART Model of Mode Choice: Some Specification Tests. Urban Travel Demand Forecasting Project, Institute of Transportation Studies, Univ. of California, Berkeley, Working Paper 7620, 1976.
6. T. J. Atherton and M. Ben-Akiva. Transferability and Updating of Disaggregate Travel Demand Model. Paper presented at the 55th Annual Meeting, TRB, 1976.
7. D. McFadden and F. Reid. Aggregate Travel Demand Forecasting From Disaggregated Behavioral Models. TRB, Transportation Research Record 534, 1975, pp. 24-37.
8. J. F. Curtain. Effect of Fares on Transit Riding. HRB, Highway Research Record 213, 1968, pp. 8-20.
9. D. K. Holland. A Review of Reports Relating to the Effects of Fare and Service Changes in Metropolitan Public Transportation Systems. U.S. Department of Transportation, Federal Highway Administration, 1974.
10. M. A. Kemp. Some Evidence of Transit Demand Elasticities. Transportation, Vol. 2, 1973, pp. 25-52.
11. W. Lassow. Effect of the Fare Increase of July 1966 on the Number of Passengers Carried on the New York City Transit System. HRB, Highway Research Record 213, 1968, pp. 1-7.
12. J. I. Scheiner and G. Starling. The Political Economy of Free-Fare Transit. Urban Affairs Quarterly, Vol. 2, No. 10, 1974, pp. 170-184.
13. A. B. Sosslau, K. E. Heanue, and A. J. Balek. Evaluation of a New Modal Split Procedure. HRB, Highway Research Record 88, 1965, pp. 44-63.
14. Tri-State Regional Planning Commission. Urban Densities for Public Transportation. Prepared for the Urban Mass Transportation Administration, 1976.
15. R. H. Pratt and Associates. Traveler Responses to Transportation System Changes: A Handbook for Transportation Planners. U.S. Department of Transportation, 1977.
16. F. C. Kavak and M. J. Demetsky. Behavioral Modeling of Express Bus-Fringe Parking Decision. TRB, Transportation Research Record 534, 1975, pp. 10-23.
17. T. A. Domencich, G. Kraft, and J. P. Vallette. Estimation of Urban Passenger Travel Behavior: An Economic Demand Model. HRB, Highway Research Record 238, 1968, pp. 64-78.
18. A. Talvitie. A Direct Demand Model for Downtown Work Trips. Transportation, Vol. 2, 1975, pp. 121-152.
19. L. A. Lave. Modal Choice in Urban Transportation: A Behavioral Approach. Department of Economics, Stanford Univ., Ph.D. thesis, 1968.
20. S. L. Warner. Stochastic Choice of Mode in Urban Travel: A Study in Binary Choice. Northwestern Univ. Press, Chicago, 1962.
21. T. E. Lisco. The Value of Commuters' Travel Time: A Study in Urban Transportation. Department of Economics, Univ. of Chicago, Ph.D. thesis, 1967.
22. D. McFadden. The Measurement of Urban Travel Demand. Journal of Public Economics, Vol. 3, 1974, pp. 303-328.
23. R. G. McGillivray. Binary Choice of Transport Mode in the San Francisco Bay Area. Univ. of California, Berkeley, Ph.D. thesis, 1969.
24. Peat, Marwick, Mitchell and Co. Implementation of the n-Dimensional Logit Model. Comprehensive Planning Organization, San Diego County, Final Rept., 1972.
25. Development and Calibration of Mode Choice Models for the Twin Cities Area. R. H. Pratt and Associates and DTM, Minneapolis, MN, 1976.

*Publication of this paper sponsored by Committee on Passenger Travel Demand Forecasting and Committee on Traveler Behavior and Values.*

# Technique for Determining Travel Choices for a Model of Nonwork Travel

Thomas E. Parody, Charles River Associates, Cambridge, Massachusetts

**In transportation corridor studies, it is not always clear whether the effects of non-mode-choice decisions for discretionary travel demand should be considered in detailed analyses. This paper presents a manual approach that can be used by planners to determine quickly whether time-of-day, trip frequency, or destination choice effects can be neglected early in the planning process. The approach relies on demand elasticities obtained from disaggregate travel demand models. Demand models that capture the causal structure of shopping trip decisions were first introduced in 1972 in a study performed by Charles River Associates for the Federal Highway Administration. To simplify the modeling approach, the study developed the concept of inclusive price. This paper presents a revised specification of the inclusive price variables and identifies the resulting new elasticity equations for separable discretionary travel demand models. The differences between the previous and revised definitions of elasticity with respect to travelers' responses to changes in transportation level of service are highlighted.**

Until recently, most studies of disaggregate travel demand examined only mode-choice behavior for the work trip. At the present time, however, disaggregate demand models have been calibrated and evaluated for discretionary or shopping trip travel behavior. These models indicate that changes in level of transportation service (e.g., travel times or travel costs) will generally affect choice of time of day, destination, and trip frequency as well as choice of mode. Thus, these models are able to capture the causal structure due to changes in one or more of the transportation level-of-service (LOS) variables.

In certain instances, however, a proposed modification to the transportation system may only affect mode choice. Consequently, time of day, trip frequency (or generation), and destination can be neglected. This assumption is routinely made for certain trip purposes such as for trips to work. But for other nonwork travel decisions it is not always obvious when a planner should concentrate just on shopping mode choice without also having to examine destination, frequency, and time-of-day decisions. Clearly, when these other choices can be ignored (without introducing a significant error), the analysis will be simplified, resulting in quicker and less costly studies.

For some policies the type of analysis under consideration will indicate a priori when time-of-day, destination, or frequency choices can be omitted. As an example, a change in tolls that increases travel costs to an airport may not cause shifting among destinations or to other times of the day. We would expect, however, to observe shifts of modes. Given a large enough increase in the toll, some reductions may also occur in the frequency of travel. On the other hand, an improvement in travel service along a single corridor may attract trips from other destinations, especially if the change is significantly large and involves a sensitive variable that changes travel behavior. As a consequence of this policy, additional choice decisions over and above mode should be explicitly considered in the analysis. The question is how large a change must occur in an independent LOS variable before alternative times of day, destinations, or trip frequencies must be evaluated, in addition to modes. When a priori theory and intuitive judgment are uncertain as to the degree of traveler response for a particular policy, what decision criterion should be employed?

This paper presents a manual approach that can be used by planners to determine quickly whether time-of-day, trip frequency, or destination choice effects can be neglected at an early stage in the planning process. The approach relies on demand elasticities obtained from disaggregate travel demand models. If the time-of-day, destination, and frequency elasticities are sufficiently inelastic, then it is possible that some or all of these elements of choice can be dropped from further analysis. After the responsiveness of each element of choice is evaluated using elasticity measures, the appropriate choice functions remaining can be subjected to a more thorough analysis using disaggregate demand models (5, 6).

## ELASTICITY DEFINITIONS

Elasticities have been widely reported in many of the previous aggregate and disaggregate demand studies in order to indicate the sensitivity of a model with respect to a fixed percentage change in one of the independent variables. Often elasticities become a means of comparing the responsiveness of a recently developed model to other models already cited in the literature.

Elasticities are also used to make quick estimates or forecasts for policy analysis when budget or time considerations prohibit using a more sophisticated demand model. For example, the Simpson-Curtin rule is a widely known fare elasticity that suggests a 0.33 percent decrease in transit patronage for every 1 percent increase in transit fares.

Until recently, elasticities (both direct and cross) were mainly reported in relation to changes in mode choice because most models were concerned specifically with this aspect of travel behavior. This was true for disaggregate demand models until 1972 when Charles River Associates (CRA) (2) reported on the development of disaggregate nonwork models for travel choices such as shopping time of day, destination, and trip frequency.

To reduce the complexity of estimating the sequence of models after mode choice, the concept of an inclusive price was introduced. This allowed the various mode attributes of a trip to be combined into a single, generalized variable. As a result, a substantial savings was realized in the number of parameters that had to be estimated for any given model formulation.

Inclusive price for a particular mode was defined as the sum of the attributes for that mode weighted by the calibration coefficients from the mode-choice function for the trip purpose under consideration. That is,

$$IP'_m = -\sum_{\text{all } k} \alpha_k X_{mk} \qquad (1)$$

where

$IP^t_m$ = inclusive price for mode m for traveler t,
$X_{mk}$ = the value of the kth attribute for mode m, and
$\alpha_k$ = the estimated coefficient for the kth attribute.

The assumption of an inclusive price implies that the relationships or marginal rates of substitution between certain variables (notably time and cost) remain constant over different choice functions (i.e., destination, time of day, and trip frequency). However, the relative effect of a group of variables that appears in more than one model can differ among choices. In fact, this is accounted for by the coefficient of the inclusive price variable.

As employed by CRA (2) and Domencich and McFadden (7) the inclusive price in the time-of-day model, for example, was summed over all modes m, at the time period h, to destination d, given that a trip is made (f = 1), using as weights the expected conditional probability, P(m/hdf), of selecting each mode. That is,

$$IP^t_h = \sum_{\text{all } m} IP^t_m \times P(m/hdf) \qquad (2)$$

Based on an assessment of model assumptions (3, p. C-143) associated with the above linear composition of inclusive price, a revised formulation has been developed and is presented in a later section of this paper. With the new inclusive price formulation it can be shown that the conditional or separable probability models will, under certain assumptions, yield results similar to the joint probability model of shopping demand. Specifically, if the coefficients of the inclusive price variables are restricted to one, then the separable models will be mathematically equivalent to the joint model.

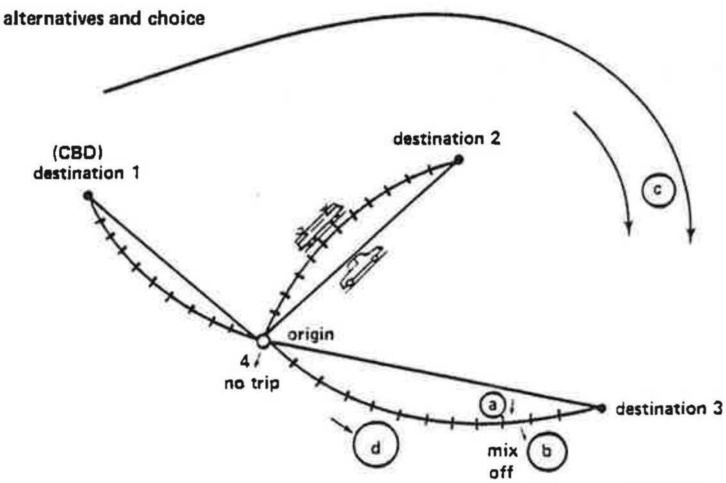
Because of the change in the definition of inclusive price, the elasticity formulas originally presented in the 1972 report will also change. We shall develop the proper elasticity equations for the separable disaggregate demand models and compare the resulting change in elasticities to those based on the old definition of inclusive price.

## SAMPLE APPLICATION

The following illustrative but typical example (Figure 1) is presented to show how demand elasticities can be used to identify, a priori, the approximate magnitude of change in discretionary travel behavior, given a proposed policy that affects transportation level of service. The objective will be to determine when time of day, destination, and frequency effects should be modeled or when they can be ignored when a change will occur in one of the independent variables influencing demand.

In this example it is assumed that for a given traveler three alternative shopping destinations are available: one may be the CBD and the other two may be regional shopping centers. The example could conceivably be a group of individuals stratified by a market segmentation category. Each destination is served by both auto and transit, but, because of service differentials, the mode-

Figure 1. Sample alternatives and choice probabilities.




Legend

Modes, m: —— = auto (a); –+–+– = transit (t)

Time of Day, h: m = mixed period trips; o = off peak trips

LOS Data

off peak transit travel time 0—1 = 20 min.
off peak transit travel time 0—2 = 25 min.
off peak transit travel time 0—3 = 30 min.

Choice Probabilities

Modes $[P_{m/h,d}]$

| | | | |
|---|---|---|---|
| $P_{a/m,1}$ = .30 | $P_{a/o,1}$ = .40 | $P_{t/o,1}$ = .60 | $P_{t/m,1}$ = .70 |
| $P_{a/m,2}$ = .40 | $P_{a/o,2}$ = .50 | $P_{t/o,2}$ = .50 | $P_{t/m,2}$ = .60 |
| $P_{a/m,3}$ = .50 | $P_{a/o,3}$ = .60 | $P_{t/o,3}$ = .40 | $P_{t/m,3}$ = .50 |

Destinations $[P_{d/f}]$

$P_{1/f=1}$ = 1/3
$P_{2/f=1}$ = 1/3
$P_{3/f=1}$ = 1/3

Frequencies $[P_f]$

$P_{f=1}$ = .67
$P_{f=0}$ = .33

Time of Day $[P_{h/d}]$

$P_{o/d}$ = 2/3
$P_{m/d}$ = 1/3

choice probabilities vary with destination and time of day. Figure 1 presents the proposed alternatives under consideration and the choice probabilities for an individual traveler.

Basically, the example assumes that the traveler is twice as likely to make a home-shop-home round trip in a given 24-h time period than not (i.e., $P_{f=1}$ is twice $P_{f=0}$). If he or she does make a shopping trip, the odds of having both legs of the shopping tour occur in the off-peak period are twice as likely as having one of the legs take place in a peak period and the other in an off-peak—what is referred to as a mixed-peak trip (i.e., $P_{o/d}$ is twice $P_{m/d}$). It is assumed that the number of shopping trips that have both segments of the sojourn in the peak period is negligible. For a given shopping trip, each possible destination has an equal likelihood of being selected. If, in the off-peak period, destination 1 is selected, the traveler is more likely to travel by transit, while for destination 2 either mode is equally likely. Last, for destination 3, the odds indicate that the traveler is more likely to choose auto. For mixed-peak shopping trips the odds suggest more transit use, possibly because of arterial roadway congestion. With a high probability that transit will be chosen for destination 1, this may be thought of as a radial corridor to the central business district (CBD) that has very good transit service.

In the example, it is assumed that a proposed policy will alter off-peak transit service in the corridor to destination 3 such that linehaul transit times will be affected. The transit manager would like to obtain the increase that may be expected in off-peak ridership levels. For this example, the increased transit patronage could (a) be diverted from the auto users already traveling to destination 3, (b) result in travelers changing their time-of-day decisions for shopping, (c) result from more trips in corrider 3 that were attracted from destinations 1 and 2, and (d) be the consequence of newly generated trips (see corresponding letters and arrows in Figure 1). The question thus becomes what conditions will require all four of the above effects to be modeled rather than a subset of the choice functions to be examined. By being able to confidently reduce the number of decisions to be analyzed, savings can be realized in time and cost.

The first step in the analysis is to compute the transit direct elasticities for mode, time of day, destination, and trip frequency. The required elasticities for a range of base travel time and travel cost characteristics are shown in Table 1. The actual calculations of the elasticities (1) only have to be performed once.

By examining the elasticities shown in Table 1, a few observations are evident. First, under most conditions the demand for transit travel is inelastic with respect to transit travel time and cost. This applies to time-of-day, destination, and trip frequency elasticities, in addition to those for mode choice. It is immediately obvious from the very inelastic frequency elasticity of demand that infinitesimally small changes in generated shopping travel will occur given a change in transit travel time or costs. Thus, this aspect of choice can be confidently omitted from further analysis in this particular application. Also, because the example is concerned with changes in transit travel times, the remaining discussion will focus on transit time elasticities calculated at the existing condition of 30 min.

Although the remaining three elasticities for mode, time of day, and destination suggest inelastic behavior, the expected percentage change in the independent variable is computed next, since a change of significant magnitude in transit travel time could cause a measurable response in modal, temporal, or destination choices. For this example, we consider the effects of a 5-min reduction in the roundtrip off-peak transit travel time only to destination 3 caused, for example, by a bus signalization scheme's being introduced.

As previously presented, the traveler's current roundtrip in-vehicle transit travel time during the off-peak to destination 3 is 30 min. Consequently, a reduction in travel time from 30 to 25 min represents a -16.67 percent change (5/30 = 16.67 percent) in the independent variable. The transit mode elasticity is -0.72 from Table 1 based on a 1 percent change in transit travel time. Thus, for a -16.67 percent reduction in travel time, the percentage increase in transit modal split to destination 3 may be approximated by taking the product of elasticity and the appropriate percentage change, -0.72 × -16.67, which equals a 12 percent increase.

The time-of-day direct elasticity with respect to off-peak transit travel time is -0.107. The percentage increase in off-peak shopping trips can, therefore, be computed as the product of -0.107 and -16.67 or +1.78 percent.

In a similar fashion, the destination choice direct elasticity for destination 3 with respect to transit travel time is -0.15. Thus, the percentage increase in trips diverted to destination 3 is -0.15 × -16.67 or 2.5 percent.

The new choice probability for destination 3 is now (1 + 0.025) × 0.33 or 0.342. Since it is an implicit property of the logit model that the cross elasticities are equal for all other alternatives, the percentage reduction in travel to the other destinations is the same for each and equal to one-half the increase to destination 3. Thus, the probability of selecting either destination 1 or destination 2 is now equal to approximately 0.329. Consequently, whereas each destination had an equal chance of being selected (i.e., p = ⅓) before the proposed change in LOS, these preliminary forecast destination shares, because of the change in the transportation system, are 0.329, 0.329, and 0.342 (the sum = 1.0) for destinations 1, 2, and 3 respectively.

It would appear at first glance that the time-of-day and destination effects are minimal. However, the final step of the analysis computes the increase in modal shares attributable to travelers changing times of day and destination.

The initial modal shares to destination 3 during the off-peak were 0.6 for auto and 0.4 for transit. As we have determined from travelers who shift modes given that a trip is made to destination 3 in the off-peak

**Table 1. Off-peak transit demand elasticities to destination 3.**

| Travel Component | Travel Time (min) 15 | 30 | 45 | 60 | Travel Cost ($)[a] .25 | .50 | 1.00 | 2.00 |
|---|---|---|---|---|---|---|---|---|
| Mode | -0.36 | -0.72 | -1.08 | -1.44 | -0.25 | -0.51 | -1.01 | -2.02 |
| Time of day | -0.05 | -0.107 | -0.16 | -0.21 | -0.04 | -0.08 | -0.15 | -0.30 |
| Destination | -0.08 | -0.15 | -0.23 | -0.30 | -0.05 | -0.11 | -0.21 | -0.42 |
| Frequency | -0.002 | -0.004 | -0.005 | -0.007 | -0.001 | -0.002 | -0.005 | -0.01 |

[a]Calculated based on an average household income of $8000 (1967).

**Table 2. Numbers of travel changes for a 5-min reduction in transit LOS in corridor 3.**

| Travel Component | Off-Peak Hour Transit | Off-Peak Hour Auto | Off-Peak Hour Subtotal | Peak Hour Transit | Peak Hour Auto | Peak Hour Subtotal | Total |
|---|---|---|---|---|---|---|---|
| Base | 2400 | 3600 | 6000 | 1500 | 1500 | 3000 | 9000 |
| Changes in destination | 2460 | 3691 | 6151 | 1538 | 1538 | 3076 | 9227 |
| Changes in time of day | 2505 | 3756 | 6261 | 1483 | 1483 | 2966 | 9227 |
| Changes in mode | 2805 | 3456 | 6261 | 1483 | 1483 | 2966 | 9227 |

**Table 3. Increase in off-peak transit ridership to destination 3.**

| Travel Component | Incremental Ridership Increase | Percent Increase (2400 base) | Percent Total Increase |
|---|---|---|---|
| Base | 0 | 0 | 0 |
| Destinations | 60 | 2.5 | 15 |
| Time of day | 45 | 1.9 | 11 |
| Modes | 300 | 12.5 | 74 |
| Total | 405 | 16.9 | 100 |

period, transit demand will increase by 12 percent. This results in a new transit share of $0.4 \times 1.12 = 44.8$ percent. Based on the time-of-day elasticity, 1.78 percent more shopping trips will be taken in the off-peak rather than in mixed-peak periods. From the destination analysis we now have 2.5 percent more trips diverted to destination 3 from other destinations.

From the mode-choice analysis it is anticipated that the transit mode would attract about 45 percent of the new trips made in the off-peak period. Likewise, about this same fraction of the new travelers attracted to destination 3 will select the transit mode. The summary results of the entire analysis can be presented more clearly in tabular form (see Table 2).

For this example, it is assumed that the zone where trips originate has a population of 40 500 (see Figure 1). Of these individuals, two-thirds make a home-shop-home trip on a given day, resulting in 27 000 shopping trips that leave the origin zone. The probabilities presented in Figure 1 indicate that one-third (9000) of these trips are destined for destination 3. Of these trips, two-thirds (6000) are taken during the off-peak time periods. The remaining 3000 trips are mixed-period trips having one leg of the trip fall in a peak period and the other in an off-peak period. As previously stated, the assumption is made that a negligible number of shopping trips with both legs of the trip in peak periods occur.

Referring to Table 2, the number of trips in the base (or before) case is allocated to the auto and transit modes based on the modal split probabilities presented in Figure 1. Thus, for example, of the 6000 trips taken to destination 3 in the off-peak, 40 percent or 2400 are made by transit and 60 percent or 3600 by auto. For the mixed period, 1500 trips are made by each mode, since it has been postulated that a traveler has an equal probability of selecting the auto or bus mode in the mixed-peak time period.

From the elasticity analysis, a 16.6 percent decrease in off-peak transit travel time resulted in the number of trips to destination 3 increasing by 2.5 percent to 9227. Two-thirds or 6151 of these trips occur in the off-peak, while the remaining one-third or 3076 are taken in the mixed-peak period. Because of changes in the time of day when trips are made, 1.78 percent of the travelers ($1.0178 \times 6151 = 6261$) switch to the off-peak period. The total number of trips in the corridor remains constant at 9227.

Last, the decrease in off-peak transit travel time to destination 3 results in a 12 percent increase in transit trips to 2805 ($1.12 \times 2505$). Again, the total trips in the corridor remain constant as well as the number of trips in each time-of-day period.

Table 3 presents the absolute and relative increases in transit ridership that can be attributed to each separate component of travel behavior. As can be seen from the table, change in modes has the greatest effect on ridership, while time-of-day choice has the least effect. From this example it could be concluded that since three-quarters (300/405) of the increase in transit ridership is the result of changes in mode, only mode-choice effects should be examined with the more sensitive disaggregate modeling techniques. Given larger changes in transit travel time, however, destination and time-of-day effects may have to be considered, as they account for 15 and 11 percent of the increase in the transit ridership, respectively, depending on the final accuracy desired.

## SEPARABLE SHOPPING DEMAND MODELS

In this section we present the mode-choice disaggregate demand model and the revised specification of inclusive price for the shopping trip purpose. A complete, detailed review of all the separable choice models is not repeated here; rather, reference is made to the Phase II report (1). The mode-choice model is

$$P^t_{m/hdf} = e^{\alpha X_{mhdfk}} / \sum_{\text{all } m} e^{\alpha X_{mhdfk}} \tag{3}$$

where

$P^t_{m/hdf}$ = probability that traveler t will select mode m given that a trip is made (f = 1) to destination d at time period h,

$X_{mhdfk}$ = a vector of k variables for mode m given that a trip is made to destination d at time period h, and

$\alpha$ = a vector of parameter coefficients.

The new inclusive price is defined as

$$I_h = \ln \sum_{\text{all } m} e^{\alpha X_{mhdfk}} \tag{4}$$

where

$I_h$ = inclusive price of travel in time period h summed over all modes m in general, the procedure for computing the inclusive price is sometimes called the "log of the denominator"; in this instance the denominator would be in reference to Equation 3 and

ln = natural logarithm.

## SHOPPING ELASTICITIES

Direct and cross elasticities are presented in this section for each of the separable shopping travel demand models. The relationships are derived based on the theoretical definition of point elasticity of demand.

### Mode Choice

The mode-choice model is not affected by the revised definition of inclusive price. Therefore, the elasticities are identical to those previously reported. In particular,

$$E^{P_{m/hdf}}_{X_{mhdfk}} = [(\partial P_{m/hdf})/(\partial X_{mhdfk})] \times [(X_{mhdfk})/(P_{m/hdf})]$$

$$= \alpha_k \times X_{mhdfk} \times (1 - P_{m/hdf}) \tag{5}$$

where

$E^{P_{m/hdf}}_{X_{mhdfk}}$ = mode m direct elasticity for traveler t with respect to variable k of the vector $X_{mhdfk}$.

The direct elasticity in the above expression varies from zero, when $P_{m/hdf} = 1$, to $\alpha_k X_{mhdfk}$, when $P_{m/hdf} = 0$. In other words, the larger the choice probability becomes, the smaller will be an individual's direct elasticity of demand, all else being equal.

In a similar manner, the cross elasticity ($m' \neq m$) can be defined as

$$E^{P_{m/hdf}}_{X_{m'hdfk}} = [(\partial P_{m/hdf})/(\partial X_{m'hdfk})] \times [(X_{m'hdfk})/(P_{m/hdf})]$$

$$= -\alpha_k \times X_{m'hdfk} \times P_{m'/hdf} \tag{6}$$

where

$E^{P_{m/hdf}}_{X_{m'hdfk}}$ = mode m cross elasticity for traveler t with respect to variable k of the vector $X_{m'hdfk}$.

Notice that the elasticities can only be calculated with respect to changes in mode characteristics of the independent variables (e.g., changes in auto time or transit cost) for the destination d and the frequency level f under consideration. Therefore, if transit travel cost changes to another destination, say d′, this is assumed to have no impact on the modal splits for destination d. Consequently,

$$E^{P_{m/hdf}}_{X_{mhd'fk}} = 0 \tag{7}$$

The direct and cross elasticities presented so far can be expressed in a combination form as

$$E^{P_{m/hdf}}_{X_{m'hdfk}} = \alpha_k X_{m'hdfk} (\delta_{mm'} - P_{m'/hdf}) \tag{8}$$

where

$$\delta_{mm'} = \begin{cases} 1 \text{ if } m = m' \text{ (a direct elasticity) or} \\ 0 \text{ if } m \neq m' \text{ (a cross elasticity).} \end{cases}$$

### Time-of-Day Choice

In the following sections the elasticity identities will be presented in their reduced form using notation consistent with preceding equations. Only elasticities with respect to variables that appear in the mode-choice model are presented here; other elasticities are given in the Phase II report (1).

$$E^{P_{h/df}}_{X_{mh'dfk}} = \phi\alpha_k X_{mh'dfk} [P_{m/h'df}] [\delta_{hh'} - P_{h'/df}] \tag{9}$$

where

$E^{P_{h/df}}_{X_{mh'dfk}}$ = time period h elasticity for traveler t with respect to variable k of the vector $X_{X_{mh'dfk}}$.

### Destination Choice

$$E^{P_{d/f}}_{X_{mhd'fk}} = \theta\phi\alpha_k X_{mhd'fk} [P_{m/hd'f}] [P_{h/d'f}] (\delta_{dd'} - P_{d'/f}) \tag{10}$$

where

$E^{P_{d/f}}_{mh'dfk}$ = destination d elasticity for traveler t with respect to variable k of the vector $X_{mhd'fk}$.

### Frequency Choice

$$E^{P_f}_{X_{mhdf'k}} = \lambda\theta\phi\alpha_k X_{mhdf'k} [P_{m/hdf'}] [P_{h/df'}]\ P_{d/f'} (\delta_{ff'} - P_{f'}) \tag{11}$$

where

$E^{P_f}_{X_{mhdf'k}}$ = the f = 1 frequency elasticity for traveler t with respect to variable k of the vector $X_{mhdf'k}$.

## COMPARISON OF DEMAND ELASTICITIES

The demand elasticity equations above, which were derived on the basis of the revised definition of the inclusive price variable, are compared in this section with the elasticities presented in the original CRA study of 1972 (2). The objective is to provide an indication of the change in magnitude of the new elasticities and to outline the appropriate relationships between the elasticities in each of the separable shopping demand models.

The 1972 CRA study, by employing a linear additive form for generalized prices, appeared to indicate that direct elasticities of demand may increase for each succeeding higher order of choice decisions from mode to time of day to destination and, last, to trip frequency. The resulting inference was that frequency and destination shopping travel choices were as sensitive as or even more sensitive than choice of mode to changes in transportation level-of-service characteristics. This finding, which appears inconsistent with a priori expectation, is corrected with the revised definition of inclusive price.

The effect that the new inclusive price definition has on elasticity can be illustrated very succinctly by examining the differences in the old and new formulations for frequency elasticity of demand. With the old linear, additive specification of inclusive price, the frequency demand elasticity can (using the same notation as previously defined) be represented as

$$E^{P_f}_{X_{mhdfk}} = \lambda\theta\phi\alpha_k X_{mhdfk} (1 - P_f) \tag{12}$$

which is an outdated formulation, or, with the revised definition of inclusive price based on a nonlinear summation of the explanatory variables, as

$$E^{P_f}_{X_{mhdfk}} = \lambda\theta\phi\alpha_k X_{mhdfk} \underbrace{P_{m/hdf} \times P_{h/df} \times P_{d/f}}_{\text{Additional Parameters}} (1 - P_f) \qquad (13)$$

which is the revised formulation.

As highlighted above, the new parameters in the revised frequency elasticity equation are three conditional probabilities. Since, by definition, these probabilities assume values between zero and one, the product of these three probabilities will be a joint probability whose value will generally be much less than one. Consequently, the new frequency elasticities will almost always be much smaller than those calculated using the old elasticity formulation (12).

To illustrate the change in elasticities based on the revised definition of inclusive price, consider the example used above, where we estimated transit travel time elasticities for a traveler whose transit time is 30 min. The coefficient values are taken from the CRA (2) and Phase I (3) reports. The same data from Pittsburgh were used in estimating the models for both studies. The conditional probabilities used are $P_{m/hdf}$ = 0.40, $P_{h/df}$ = 0.67, and $P_{d/f}$ = 0.033, which are approximately equal to the aggregate choice shares observed in the calibration data set. The resulting transit time elasticities for both definitions of inclusive price are

| Travel Choice | Linear Inclusive Price | Log Inclusive Price |
|---|---|---|
| Mode | -1.17 | -0.72 |
| Time of day | -1.35 | -0.107 |
| Destination | -2.88 | -0.15 |
| Frequency | -2.47 | -0.0035 |

From this example it is very apparent that the responsiveness or sensitivity of time-of-day, destination, and frequency choices to changes in transport level of service is much more inelastic than was previously indicated with models based on the linear inclusive price definition. We should note further, however, that the inelastic frequency elasticity is due in part to the small and statistically insignificant coefficient observed for the frequency inclusive price variable.

One additional comparison between the old and new elasticities concerns the magnitude of each direct elasticity as the hierarchy of shopping travel choice decisions proceeds from mode to time of day to destination and to frequency choice. For the analysis to be tractable it is assumed that inclusive price coefficients are near or at unity. (A unitary value is an inherent assumption in a joint-choice model, for example.)

With the original elasticity definitions, a subsequent choice elasticity could increase or decrease depending simply on whether the choice probability in the following choice model increased or decreased. Although highly unlikely, if the conditional probabilities for mode, time-of-day, destination, and frequency choice were identical, the previous elasticity formulations would indicate that all direct elasticities for a given LOS variable for m, h, d, and f would be the same.

Alternatively, with the revised specification of inclusive price, demand elasticities generally (but not always) decrease for each succeeding choice function, since an additional probability is introduced into each following computation. Using the supposition of identical choice probabilities given above, with $P_t$ = 0.5, each subsequent elasticity would be one-half the value of the preceding elasticity. For instance, if the auto mode-choice elasticity with respect to one of the auto independent variables is -1.0, then the time-of-day direct elasticity would be -0.5 (i.e., -1.0 × 0.5); destination elasticity would be -0.25 (i.e., -0.5 × 0.5); and frequency elasticity would be -0.125 (i.e., -0.25 × 0.5).

The one instance in which an elasticity of a following choice decision can be larger than a preceding elasticity occurs when the first conditional choice probability is somewhat higher than the subsequent choice probability. A situation where this occurs can be found in the application reported in an earlier section of this paper.

Although demand elasticities for each of the subsequent shopping choice models do not necessarily decrease monotonically, it is possible to conclude that, in general, subsequent demand elasticities tend to decrease in magnitude and that, in particular, they are much more inelastic than those demand elasticities previously calculated based on a linear inclusive price formulation.

## SUMMARY

To ascertain whether all dimensions of shopping travel behavior, in addition to mode choice, are important, a planner must clearly understand the full implications of the policy being analyzed. One very simple and rapid procedure is to compute demand elasticities for each component of travel choice. For a more complete appreciation of the approximate magnitudes of each change in travel behavior, it is useful to continue the preliminary analysis further than just an examination of demand elasticities. In particular, an explanatory variable that indicates inelastic behavior could have a significant impact on travel choice, given that a large enough change will be made in its value. Therefore, the decision criterion must consider both the elasticity and the magnitude of the proposed change in an independent variable. As in most analyses, a final decision must consider weighting time and cost considerations against the desired accuracy required of the final predictions.

Throughout the preceding example, only direct elasticities were computed. The impression should not remain, however, that these are the only type of elasticities that need to be computed to analyze every policy. For instance, consider the same example above except assume that auto (rather than transit) travel time will be reduced by a given amount by a signalization scheme. The analyst is still interested in the effect on bus ridership. Under these conditions the bus cross elasticity of demand with respect to auto travel time would be calculated.

In instances where the auto is the dominant mode, this cross elasticity can assume values much larger than the direct auto time elasticity (4, 8, 9). For example, if the auto and transit shares are 90 and 100 and an auto disincentive policy decreases the amount of auto travel to 85 (a relatively modest 5.5 percent reduction), bus modal shares will increase to 15, which, consequently, represents a rather substantial 50 percent increase in transit usage.

## ACKNOWLEDGMENTS

This research was supported by the National Cooperative Highway Research Program on disaggregate travel demand models. I would like to acknowledge the assistance provided by William Tye and Daniel Brand of Charles River Associates and Daniel McFadden of the

University of California, Berkeley. The opinions and conclusions expressed or implied in the paper are mine. They are not necessarily those of the Transportation Research Board, the National Academy of Sciences, or the National Cooperative Highway Research Program.

REFERENCES

1. Charles River Associates. Disaggregate Travel Demand Models. NCHRP, Phase II Draft Rept., Project 8-13, Dec. 1977.
2. Charles River Associates. A Disaggregated Behavioral Model of Urban Travel Demand. Prepared for the Federal Highway Administration, Mar. 1972.
3. Charles River Associates. Disaggregate Travel Demand Models. NCHRP, Phase I Rept., Project 8-13, Vols. 1 and 2, Feb. 1976.
4. Charles River Associates. Estimating the Effects of Urban Travel Policies. Transportation Systems Center, Cambridge, MA, April 1976.
5. Charles River Associates. Evaluation of Transit Service Improvements and Transportation Pricing Policies. Prepared for U.S. Department of Transportation, Aug. 1976.
6. Charles River Associates. Regional Management of Automotive Emissions: The Effectiveness of Alternative Policies for Los Angeles. Prepared for the U.S. Environmental Protection Agency (in press).
7. T. A. Domencich and D. McFadden. Urban Travel Demand: A Behavioral Analysis. North-Holland, New York, 1975.
8. T. E. Parody. A Disaggregate Prediction Analysis of Mode Choice. Univ. of Massachusetts, PhD dissertation, Apr. 1976.
9. T. E. Parody. Analysis of Predictive Qualities of Disaggregate Modal-Choice Models. TRB, Transportation Research Record 637, 1978, pp. 51-57.

*Publication of this paper sponsored by Committee on Passenger Travel Demand Forecasting and Committee on Traveler Behavior and Values.*

# Characteristics of Urban Transportation Demand: A New Data Bank

Herbert S. Levinson, Wilbur Smith and Associates, New Haven

**The Characteristics of Urban Transportation Demand is the third manual in an ongoing Urban Mass Transportation Administration-Federal Highway Administration series designed to provide real-world information for practicing transportation professionals. The manual is a single reference source on the characteristics of urban transportation demand. This paper reviews the manual's objectives, sources, content, and structure and illustrates how it can be used for inputs to and cross-checks of the long-range urban transportation planning process. The manual sets forth important controls for the number, type, and length of trips, as well as for freeway and rail transit use. Rail transit ridership, for example, relies closely on the number of people crossing the downtown cordon and on the number of workers in the central business district.**

The need for practical transportation planning tools underlies ongoing research activities of the Urban Mass Transportation Administration and the Federal Highway Administration. Two publications—The Characteristics of Urban Transportation Systems and Traveler Response to Transportation System Changes—have already enjoyed widespread distribution. The third manual in this series, Characteristics of Urban Transportation Demand, was designed as a single reference source on demand characteristics.

This paper describes the objectives, content, and use of this demand manual. It also contains representative exhibits that illustrate how key demand-related steps in the transportation planning process can be checked.

## OBJECTIVES AND SOURCES

The manual provides a compendium of information on urban travel behavior and transportation system use. It is intended to guide practicing professionals in the following tasks:

1. Assessing demands for urban highway and transit systems,
2. Applying and validating transportation planning techniques, and
3. Achieving realistic and reasonable decisions on urban transportation improvements and investments.

Toward these objectives, the manual contains inputs to and cross-checks of the conventional transportation planning process. It suggests simplified procedures for estimating or verifying urban transportation demand—especially for sketch planning. The manual also presents urban area aggregate and disaggregate transportation demand data and sets forth key parameters for urban transportation demand analyses. Finally, it contains summary characteristics of urban transportation systems.

The information contained in the manual was assembled from a wide variety of existing sources, including transit ridership statistics, urban traffic volume counts, U.S. Census journey-to-work and employment data, and comprehensive metropolitan area transportation planning studies. An attempt was made to develop comparable information on a facility-by-facility and city-by-city basis. However, many variations in reporting systems were found among locations, area units, definitions, and survey years. Despite these variations, this information provides a comprehensive picture of the dimensions and nature of urban travel behavior in both city-specific and comparative contexts.

## ORGANIZATION AND USE

The manual is organized for easy use by practicing transportation planners, engineers, and administrators who are selecting parameters, analyzing system performance, and establishing sound transportation planning decisions. Five chapters, each with a corresponding appendix, are keyed to the various steps of the comprehensive urban transportation planning process. The chapters contain general information and corresponding appendixes that contain more detailed, city-specific exhibits.

Chapter 1 presents the manual's objectives, organization, and use and discusses the nature of urban transportation demand. Demand is shown to reflect the magnitude, location and intensity of urban land use; social and economic characteristics of urban residents; supply, attractiveness, and price of urban transportation systems; and public policy decisions relating to land use, transport investment, and travel costs.

Chapter 2 contains guidelines for transportation demand analysis. It presents an overview of the long-range transportation planning process and sets forth inputs, parameters, and cross-checks for principal steps in this process: population, employment, and land use; car ownership; trip generation; trip distribution; and transportation systems assignments.

Chapter 3 contains detailed urban travel characteristics that were derived from more than 60 comprehensive metropolitan area transportation studies conducted over the past several decades. Ten sections on trip rates, trip lengths, and travel variations are keyed to the transportation planning process. These sections include urban travel summaries and trends, trip production parameters, trip generation parameters, trip purpose and car occupancy, truck travel, taxi travel, urban person travel, vehicle trip lengths, trip times, and hourly variations.

Chapter 4 sets forth floor space, employment, and travel characteristics for U.S. and Canadian city centers. The various measures of central business district (CBD) intensity are especially significant in assessing or validating demands for major transit improvements. Four basic sections contain information on employment; floor space, area, travel, and cordon crossings; trip purpose; and parking and pedestrian characteristics. Principal tabulations include all urban areas of over 1 000 000.

The exhibits show a highly skewed distribution of CBD intensity; the top U.S. city centers measured by employment, person destinations, peak-hour cordon crossings, and relative transit use are New York, Chicago, Washington, San Francisco, Boston, and Philadelphia. These six far outstrip other U.S. city centers.

Chapter 5 presents use characteristics for existing rail, bus, and highway transportation facilities. These measures provide important checks of highway and transit system assignments. Five sections contain exhibits pertaining to overall urban transit use, rail transit, bus transit, highway transportation, and traveler response to system supply changes.

These materials provide a sense of scale for existing and future use. Transit fleets of over 1500 vehicles are found only in New York, Chicago, Philadelphia, Washington, Boston, Los Angeles, Montreal, and Toronto; rail transit line volumes of over 200 000 occur only in New York and Toronto. Freeway volumes of over 200 000 vehicles per day are found only in Chicago, Los Angeles, and New York.

The manual may be used to compare travel parameters for a given community with those for other cities, thereby providing a basis for cross-checking and refinement. Similarly, figures for transportation use can be compared with those for comparable systems and cross-checked for reasonableness. The manual provides inputs where information is lacking and contains broad macro-measures for use by decision makers and administrators. It can be used to obtain or verify parameters pertaining to trip generation, purpose, length, and distribution and to assignment steps of the demand-forecasting process.

The manual is intended as a guide and must be used as such. The various relationships represent specific communities at given points in time. Care, therefore, should be taken in applying it to specific situations and sound judgment should be exercised. Judiciously used, the manual should provide a guide to achieving sound urban transportation decisions.

## CHECKING RESULTS

The manual's approach to checking the urban transportation planning process illustrates its scope and potential application. The need to check transportation demand forecasts and systems assignments is extremely important, not only in assessment of the cost effectiveness of new systems but also in the context of resource allocation.

The following questions should be addressed in any check for reasonableness. Are the methods used consistent with established procedures? Are the assumptions and parameters used reasonable in view of past trends, base-year conditions, and projected growth rates? Are the results of the forecasts realistic when compared with actual experiences in the same urban area and in other urban areas of similar size, structure, and economy?

The following parameters should be checked in developing and analyzing demand forecasts:

1. Number of trips (including trips per dwelling unit and trips per capita),
2. Types of trips,
3. Length of trips,
4. Person-hours per capita,
5. Person-kilometers per capita,
6. Vehicle-kilometers (VKMT) or vehicle-miles (VMT of travel per capita and per registered car,
7. CBD modal split versus CBD cordon counts,
8. Rapid transit ridership, and
9. Freeway volumes.

### Areawide Guidelines

General guidelines for cross-checking urban transportation demands are given in Table 1. This table lists some 20 factors, identifies their basic functions, gives ranges in key parameters, and sets forth an illustrative application. Supporting tables are cross-referenced in the manual itself.

Figure 1 shows how these guidelines can be used to develop macro-estimates or cross-checks for a typical urban area of 1 000 000 population. For example, such an urban area is likely to contain 400 000 dwelling units and 360 000 cars and to provide employment for 400 000. Residents of the area can be expected to generate 2 500 000 person-trips each weekday: 600 000 for work and 1 900 000 for other reasons. Approximately 10 percent, or 200 000, would probably be made by public transport. The 2 300 000 trips by car would yield about 1 530 000 auto-driver trips, and there would be another 270 000 truck trips, resulting in some 1 800 000 vehicle trips.

The 1 000 000 residents would generate about 15.1 VKMT (9.4 VMT) per capita, resulting in 151 280 000 daily over-the-road VKMT (9 400 000 daily over-the-road VMT) based on 41.9 VKMT (26.0 VMT) per registered car; this results in a trip length of about 8.1 km (5.0 miles).

### Cross-Check Factors

A more detailed explanation of the various factors follows.

Table 1. Guidelines for checking transportation planning process.

| Item | Function-Input | Parameters | Guidelines | Illustrative Application to Figure 1 |
|---|---|---|---|---|
| Population | Input for employment, land use, and trip generation | - | U.S. Census | Assume 1 000 000 population |
| Dwelling units | Input for employment, land use, and trip generation | - | 3.0-3.5 persons/dwelling unit in U.S. cities | Assume 3.3 persons/dwelling unit |
| Employment | Input for trip attraction | - | U.S. Census, participation ratio of 0.35-0.40 | Assume ratio of 0.40 |
| Developed land area | Input for density and trip attraction | - | Developed urban land will increase more rapidly than population | - |
| Population density | Input for systems planning and trip generation | - | Density gradient will increase more with time; compute density from first four items | - |
| Family income | Input for car ownership and trip production | - | U.S. Census or special surveys | - |
| Car ownership | Trip production and modal split | Income, density, and dwelling-unit size, where available | Will increase more rapidly than population<br>About 1.0-1.2/dwelling unit 1975, overall | Assume 1.2/dwelling unit overall |
| Total person-trips per day (within study area) | Modal analysis and system assignment | Income and car ownership | Trips increase faster than total population in urban area, 2.5-3.0 trips/person is reasonable range for future conditions | 2.5 trips/person/d for 2.75/persons/car |
| Work trips per person per day (production) | Trip distribution and modal analysis | Population | Work trips remain consistent at about 0.6-0.7 trips/person/d | Assume 0.6 work |
| Nonwork trips (production) | Trip distribution and modal analysis | Income and car ownership | Nonwork trips increase faster than work trips | Assume 1.9 trips/person/d |
| Trip attraction | Trip distribution and modal analysis | Employment, floor space, and students | - | - |
| Truck trips | Identify sources of truck travel and trip distribution and assignment | Population, land use, and floor space | Approximately 15-20 percent of all vehicle trips | Trucks represent 15 percent of all vehicle trips; cars ÷ 0.85 = total vehicle trips |
| Truck travel | Trip distribution and assignment | Population | Assume 0.8-1.2 VMT/capita/d | Assume 1.61 vehicle-km (1 VMT) per capita/d |
| Trip distribution | Connect trip productions and attractions | Zonal-distances, times | Use weighted time-distance (or else constrain future system speeds)<br>Friction factors check by comparing with existing frequency distribution | - |
| Person-hours of travel | Check and control of distribution | Population | Person-hours per person should remain relatively constant at 0.75-0.85/d | - |
| Person-kilometers of travel (over-the-road) | Check for control of distribution | Population and system speed | About 10-15 1975, 12-19 future year | Assume 20 person-km (13.0 person-miles) per person/8 km |
| Vehicle trip length-kilometers (over-the-road) | Check for control of distribution | Population and urban structure | Generally 4-6 miles in 250 000-3 000 000 population range | Assume 8 km (5.0 miles) per trip |
| Vehicle kilometers per registered car | Check for control of distribution | Car registration and population | Generally 24-28 miles in 250 000-3 000 000 population range, 10-15 VMT/capita reasonable range | Assume 42 vehicle-km (26 VMT) per day |
| Car occupancy | Obtain vehicle trips | Varies with trip purpose | 1.4-1.6 persons/car reasonable range for 1975 conditions | Assume 1.5 persons/car |
| Modal split | Allocate travel among car, bus, and rail | Car ownership, travel disutilities; develop models on a disaggregate basis | As an areawide guide, 10-15 percent transit (including school bus) represents a reasonable upper limit for most areas | Assume 10 percent of all person-trips |

Note: 1 km = 0.62 mile.

### Population

Population forecasters should recognize that considerable portions of future urban areas often are already in place. Because major growth is not likely in these built-up areas, radical changes in past growth generally should be avoided. This is certainly relevant in view of the decreasing birth rates.

### Land Consumption

Land used for urban purposes should generally increase faster than population, assuming current land-use policies and energy availability. This implies that the density gradient—i.e., the rate at which net residential density decreases with increasing time-distance from the city center—will generally flatten over time.

### Dwelling Units

A range of 3.0-3.5 persons/dwelling is common on a regional or zonal basis based on 1970-1975 conditions. In the future, the number of persons per dwelling unit is likely to decrease slightly because of the increased elderly in the population and the reduced family size.

### Employment

The participation ratio for U.S. urban areas generally represents 35-40 percent of the population. Values outside this range should be reviewed for reasonableness. CBD employment generally will represent less than 20 percent of the region's total, except in small cities. Employment in Manhattan south of 59th Street accounts for about 25 percent of the New York Tri-State region's total.

### Car Ownership

Car ownership relates to family size, income, and population density. Typical relationships, based on the 1970 Nationwide Personal Transportation Study, are

Figure 1. Typical regional travel analysis.

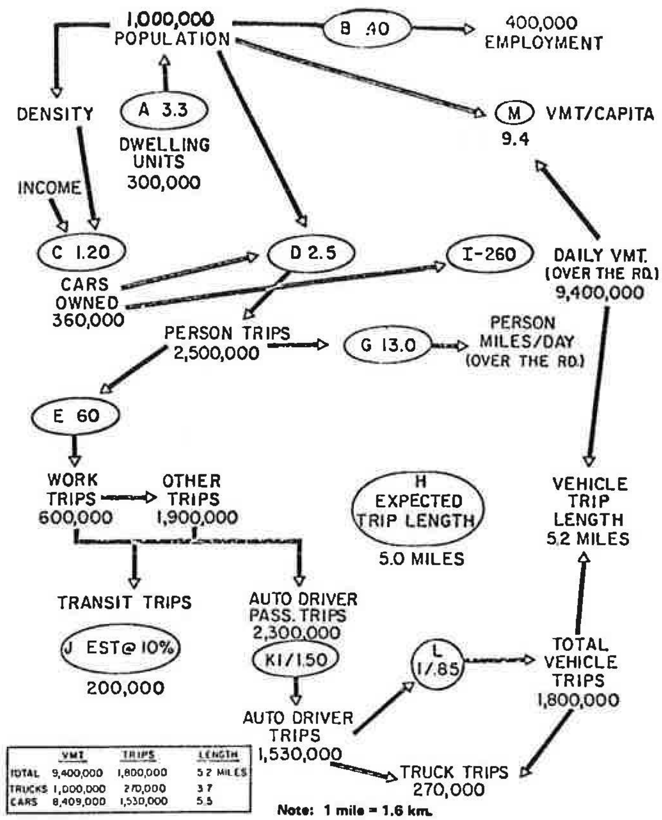



Table 2. Estimated relationships among car ownership, income, and population density.

| Density | Income ($000 000) | Percentage of Dwelling Units per Number of Cars: 0 Car | 1 Car | 2 Cars | 3+ Cars | Total |
|---|---|---|---|---|---|---|
| Medium | 4 | 30 | 60 | 10 | 0 | 100 |
| | 8 | 7 | 60 | 30 | 3 | 100 |
| | 12 | 2 | 45 | 45 | 8 | 100 |
| | 16 | 1 | 33 | 53 | 13 | 100 |
| | 20+ | 0 | 25 | 57 | 18 | 100 |
| High (more than 30 persons/acre) | 4 | 35 | 65 | 5 | 0 | 100 |
| | 8 | 9 | 62 | 27 | 2 | 100 |
| | 12 | 4 | 49 | 41 | 6 | 100 |
| | 16 | 1 | 42 | 47 | 10 | 100 |
| | 20+ | 0 | 34 | 52 | 14 | 100 |
| Low (less than 10 persons/acre) | 4 | 30 | 60 | 10 | 0 | 100 |
| | 8 | 7 | 60 | 30 | 3 | 100 |
| | 12 | 2 | 41 | 48 | 9 | 100 |
| | 16 | 1 | 30 | 55 | 14 | 100 |
| | 20+ | 0 | 22 | 59 | 19 | 100 |

shown in Table 2 and can serve as a planning guide (1). A value of 1.0 to 1.2 cars/dwelling unit reflects 1975-1977 conditions.

## Trip Rates and Purposes

Person-trip production rates are primarily a function of car ownership, which in turn depends largely upon family income. The average urban resident makes 3 trips/d where car ownership averages 2.0 persons/car. This rate drops to 2.0 person-trips/d where car ownership averages 4.0 persons/car. Members of zero-car dwelling units make about 1.0 person-trips/d, as compared with about 2.0-2.5 and 2.5-3.5 person-trips/d for members of single and multicar dwelling units respectively.

Table 3. Car ownership and trip purpose.

| Trip Purpose | Percentage Distribution per Number of Cars per Dwelling Unit: 0 Car | 1 Car | 2+ Cars |
|---|---|---|---|
| Home-based work | 35 | 26 | 21 |
| Home-based shopping | 19 | 16 | 14 |
| Home-based social recreational | 15 | 17 | 18 |
| Home-based school | 9 | 7 | 9 |
| Home-based other | 12 | 13 | 14 |
| Subtotal | 90 | 79 | 76 |
| Nonhome based | 10 | 21 | 24 |
| Total[a] | 100 | 100 | 100 |
| Trips per dwelling unit per day | 1.8 | 7.0 | 12.0 |
| Home-based work trips per dwelling unit | - | 1.8 | 2.5 |

[a]Totals reflect averages for 14 cities. Purpose distribution was based on an urban transportation study of 1965.

The number of work trips has remained constant at about 0.60-0.70/person over the past several decades. For example, according to the Washington Council of Governments in Washington, D.C., the work trips per capita decreased from 0.71 in 1942 to 0.69 in 1955, and 0.63 in 1968. This implies that the greatest increase in future trips will be for nonwork purposes.

General guidelines for estimating trip purposes for zero-car, one-car, and multicar dwelling units are shown in Table 3. For example, in a one-car unit, there are about 1.8 trips daily, of which about 0.6 is for work; in a multicar unit there are about 12.0 trips/d, of which 2.5 are for work purposes.

## Truck Trips

The number of truck trips will vary among urban areas, depending upon the extent to which light trucks (i.e., panels and pick-ups) are used as second passenger cars. Resident goods-vehicle guidelines for a typical medium-sized urban area are 28 registered goods vehicles per 1000 residents, about 10 daily trips per truck, about 43 daily VKMT (27 daily VMT) per vehicle [about 1.6 VKMT (1 VMT) per resident], and about 4.3 km (2.7 miles) per truck trip.

## Trip Distribution

Trip distribution procedures should avoid the overstatement of future person and vehicle travel common to the gravity model. This condition can be alleviated by using realistic travel time assumptions for peak and off-peak periods with system speeds reflecting capacity restraints or by using a utility function that weights time and distance in model calibration. Airline person-kilometers of travel based on 1970-1975 conditions should range from 16 to 24 km (from 10 to 15 miles) per capita per day. Over-the-road kilometers would be about 25-35 percent greater. Approximately 0.75 to 0.85 person-hour per capita is spent in travel each day.

### Vehicle-Miles of Travel

Over-the-road kilometers and vehicle trip lengths provide a valuable cross-check of the reasonableness of traffic forecasts. Regression analysis based on data for some 30 urban areas produced the following relationships (in miles):

$$\log(\text{VMT/cars}) = 0.99 + 0.07 \times (\log \text{pop.}) \quad r = 0.66 \quad Sy = 0.04 \tag{1}$$

$$\log[\text{average vehicle trip length (miles)}] = 0.39 = 0.18 \times (\log \text{pop.}) \quad r = 0.85 \quad Sy = 0.06 \tag{2}$$

Vehicle trip length and vehicle-kilometers of travel guidelines (for 15 percent of trucks included but calculated in miles) based on these formulas are summarized below.

| Urban Population | Total VMT/ Total Cars | Vehicle Trip Length (over-the-road kilometers) |
|---|---|---|
| 100 000 | 22.0 | 5.5 |
| 250 000 | 24.0 | 6.3 |
| 500 000 | 25.0 | 7.1 |
| 1 000 000 | 26.0 | 8.1 |
| 2 000 000 | 29.5 | 9.0 |
| 5 000 000 | 29.0 | 10.5 |
| 10 000 000 | 31.0 | 11.8 |

Total daily kilometer travel per car should be less than 48 (30 miles) except in the largest metropolitan areas. Similarly, average vehicle trip lengths should not exceed 9.7 km (6 miles), except where the urban population exceeds 3 000 000.

## CBD Travel and Transportation Systems Assignments

Results of transportation assignments, both systemwide and on an individual facility basis, should be reviewed for reasonableness. Comparisons with base-year conditions in the same community are a point of departure. Design-year projections also should be compared with use factors for existing facilities in urban areas of comparable size.

Guidelines for assessing CBD travel characteristics and systems assignments are set forth in Table 4. These exhibits are based on a review of existing CBD travel patterns, rail transit ridership, and freeway volumes in American cities. They provide analogy cross-checks for design-year system assignments.

### CBD and Non-CBD Travel Comparisons

Table 5 shows typical travel demand indexes for urban areas of 100 000 and 1 000 000 population, respectively. It also presents detailed employment, cordon crossings, and destinations for centralized downtown areas. These values represent a reasonable upper limit of what might occur in the typical U.S. city center. It may be seen that, as an urban population rises, the central business district will usually grow relatively more slowly than the region as a whole in terms of floor space, employment, and trip generation. An urban population area of 1 000 000 would generate about 9 times the person trips, 12 times the person-kilometers, and about 3 times the CBD cordon crossings that would be generated by a community of 100 000.

### CBD Cordon Counts and Rapid Transit Ridership

Peak-hour travel demands in radial corridors approaching a centralized CBD are shown in Table 6 (2, 3). This tabulation is based on analyses of cordon count data that include modal and submodal splits in existing U.S. and Canadian cities. In cities of over 2 000 000 with extensive rail transit systems (excluding New York City) about 50-80 percent of the total transit ridership is off-street (3). A 65 percent distribution of rail-to-total transit, therefore, appears reasonable as an upper limit on rail transit potentials in large cities.

For an urban area of 1 000 000 persons, a maximum corridor one-way express transit volume of almost 3000 persons/h can be anticipated. For an urban area of 2 000 000, the heaviest corridor would generate a potential rail transit one-way volume from 4400 to 13 800 persons. This would correspond to a daily two-way ridership of 30 000-90 000, assuming that the one-way peak is 5 percent of the daily two-way total travel.

### Rapid Transit Ridership

The relations between rapid transit-commuter rail ridership and CBD employment are shown in Tables 7 and 8. Overall, daily system ridership ranges from 0.5 to 2.6/CBD employee, New York City and Toronto excluded, depending upon the extent of the system. For Toronto, the factor is 3.4; for New York it is 2.5 based on Manhattan south of 59th Street and 4.4 based on midtown and downtown alone. These relationships are generally con-

**Table 4. CBD travel and systems assignment guidelines.**

| Item | Function Input | Parameters | Guidelines and Sources | Illustrative Application |
|---|---|---|---|---|
| CBD employment | CBD trip attractions; check modal distribution and transit assignments | - | CBD employment will become a smaller proportion of metropolitan area employment; for a large urban area it should generally be less than 15 percent of the total | For 1 000 000, a strong CBD would have 15 percent of 60 000 |
| CBD peak cordon crossings and modal split | Check highway and transit system assignments | Employment, density, car ownership, disutility | Peak-cordon volumes of over 100 000 will be limited to urban areas of over 2 000 000; transit will not generally exceed 65 percent of peak in heaviest cordon | CBD cordon crossings in peak would approximate 54 000 up to 27 000 by transit |
| Rapid transit ridership | Cost-effectiveness evaluation | Result of systems assignment; check by CBD employment and cordon crossings | Divide by 15 percent to obtain daily two-way volumes | Maximum corridor one-way ridership for urban areas 1 000 000 = 3 000/h |
| | | | Reasonable assumptions that peak auto trips to CBD will remain constant, with no capacity increase; allocate growth to CBD cordon to rapid and surface transit | |
| | | | System ridership generally from 1-2.4 times CBD employment, depending upon extent of system; maximum line ridership about 0.4 employment | Maximum line in urban area of 1 000 000 with 60 000 CBD jobs about 24 000/d |
| Rapid transit ridership and CBD orientation | - | - | 65-80 percent of ridership has origin or destination in CBD | - |
| Modes of arrival at stations | - | - | 25-35 percent are likely to walk to and from stations; may reduce to 15-25 percent in suburban-oriented systems | - |
| Freeway volumes | Cost-effective evaluation | Systems extent configuration; VMT, urban population | Maximum load point on system under 200 000 unless urban population exceeds 2 000 000, when it may reach 250 000 | Range in maximum load point for urban population of 1 000 000 is 60 000-150 000 |

sistent, although not as precise, as those found by Pushkarev and Zupan (4, p. 28), who found that office floor space relates closely to commuter rail and rapid transit ridership on a logarithmic scale.

It is clear that the intensity of downtown development has important bearing on rail transit ridership and provides an important cross-check for evaluating system assignments.

Ranges for the maximum rail transit line ridership are shown below.

| Urban Population | Range ADT (rounded) |
|---|---|
| 250 000 | 35 000-60 000 |
| 500 000 | 40 000-100 000 |
| 1 000 000 | 60 000-150 000 |
| 2 000 000 | 80 000-200 000 |
| 3 000 000 | 100 000-210 000 |
| 4 000 000 | 110 000-220 000 |
| 5 000 000 | 120 000-230 000 |
| 10 000 000 | 180 000-250 000 |

This table shows ridership of 0.3-0.5 times the CBD employment for U.S. cities (1973-1975 conditions). Thus, a broad cross-check of maximum line ridership might approximate 40 percent of the downtown employment. This results in a maximum load-point daily volume of 40 000-60 000 where CBD employment is under 150 000. These values would compare with the 30 000-90 000 daily ridership figures implied in Table 7. Maximum line volumes of over 100 000 should be rechecked for reasonableness in evaluating ridership potentials of new rapid transit systems.

The additional factors that should be taken into account in evaluating rapid transit ridership projections are, first, that approximately 65-85 percent of the trips will have either origin or destination in the CBD, depending on CBD size and system configuration. For example, over 80 percent of all rail transit trips in New

**Table 5. Demand indexes for urban areas.**

| Item | Urban Population | | |
|---|---|---|---|
| | 100 000 | 1 000 000 | Ratio |
| Daily person trips (all modes) | 270 000 | 2 500 000 | 9.2 |
| Daily person-kilometers of travel | 1 770 000 | 15 000 000 | 11.8 |
| Vehicle-kilometers of travel | 1 320 000 | 15 100 000 | 11.5 |
| CBD area (square kilometers) | 0.78 | 1.86 | 2.4 |
| CBD employment | 10 000 | 60 000 | 6.0 |
| CBD floor space (thousands of square meters) | 418 | 232 | 5.6 |
| CBD cordon entrants per 12 h (persons) | 100 000 | 270 000 | 2.7 |
| CBD cordon crossings (persons outbound in peak hour) | 20 000 | 54 000 | 2.7 |
| CBD maximum person accumulation | 25 000 | 75 000 | 3.0 |
| CBD person destinations | 35 000 | 135 000 | 3.9 |
| Percent to work | 30-35 | 50-55 | 1.7 |
| CBD parking | | | |
| Total | 4 200 | 18 000 | 4.3 |
| Off-street spaces | 2 500 | 16 000 | 6.4 |
| On-street spaces | 1 700 | 2 000 | 1.2 |
| Daily parkers | 14 000 | 38 000 | 2.7 |

Note: 1 km = 0.62 mile, 1 km$^2$ = 0.39 mile$^2$, and 1 m$^2$ = 10.76 ft$^2$.

**Table 6. Peak-hour demand on main corridors to a centralized CBD.**

| Item | Urban Population | | | |
|---|---|---|---|---|
| | 250 000 | 500 000 | 1 000 000 | 2 000 000 |
| CBD destinations per day (persons) | 40 000 | 80 000 | 135 000 | 225 000 |
| PM peak-hour outbound across cordon (persons) | 16 000 | 32 000 | 54 000 | 100 000 |
| Heaviest corridor 25-35 percent (persons) | 4 000-5 600 | 8 000-11 000 | 13 500-19 000 | 25 000-33 000 |
| Percent transit | 10-15 | 25-35 | 40-50 | 50-65 |
| Number of transit passengers | 400-800 | 2 000-3 300 | 5 400-8 500 | 12 500-21 300 |
| Percent express bus or rail transit of total transit | 10-15 | 15-20 | 25-35 | 35-50-65[a] |
| Express transit | 40-120 | 300-660 | 1 350-2 900 | 4 400-10 600-13 800 |
| Equivalent express buses (50 passengers/bus) | 1-3 | 6-12 | 27-58 | 88-212[b] |
| Equivalent rail cars (100 passengers/car) | -[c] | -[c] | 14-29 | 44-106-138 |

[a] For existing cities with extensive rail transit, this percentage ranges from about 50-80 (New York excluded). A 65 percent factor therefore appears reasonable in estimating rail system potentials in large cities.
[b] Outside domain of buses.
[c] Outside domain of rail.

**Table 7. Macroanalysis of rapid transit ridership.**

| City | Daily Travel | | | No. of CBD Employed | Total Travelers Divided by CBD Employed |
|---|---|---|---|---|---|
| | Rapid Transit | Commuter Rail | Total | | |
| New York | 3 717 000<br>155 000<br>17 000 | 536 000 | 4 425 000 | 1 777 000[a]<br>1 000 000 | 2.5<br>4.4 |
| Chicago | 512 000 | 269 000 | 781 000 | 300 000 | 2.6 |
| Philadelphia | 350 000<br>25 000<br>41 000 | 114 000 | 530 000 | 225 000 | 2.4 |
| Boston | 420 000 | 42 000 | 462 000 | 263 000 | 1.8 |
| San Francisco | 120 000<br>100 000 | 18 000 | 238 000 | 282 000 | 0.8 |
| Cleveland | 42 000<br>17 000 | - | 59 000 | 117 000 | 0.5 |
| Toronto | 682 000 | 25 000 | 707 000 | 210 000[b] | 3.4 |

[a] Manhattan south of 59th Street. [b] Central area south of Bloor.

**Table 8. Rapid transit volumes versus CBD employed (1974-1975).**

| City | Line | No. of Riders | No. of CBD Employed | Rides per Employee |
|---|---|---|---|---|
| New York | Lexington Ave. Express and Local | 524 000 | 1 000 000[b] | 0.39-0.54 |
| | Queens Blvd. Express (two tracks) | 375 000 | 1 777 000 | 0.21-0.38 |
| Chicago | NS-North Side | 120 000 | 300 000 | 0.40 |
| Philadelphia | Market Franford west or north of CBD | 103 000 | 225 000 | 0.46 |
| Boston | Green west of CBD | 110 000 | 263 000 | 0.42 |
| San Francisco | Surface cars (LRT) | 100 000[a] | 282 000 | 0.36 |
| | BART Transbay | 84 000[a] | 282 000 | 0.30 |
| Cleveland | West Side | 34 000 | 117 000 | 0.29 |
| U.S. range | | | | 0.29-0.54 |
| Toronto | Yonge | 265 000 | 210 000 | 1.26 |

[a] Total riders on maximum line; maximum load point ridership would be less. [b] Manhattan south of 59th Street.

York City are to or from Manhattan south of 59th Street. Corresponding figures are 84 percent for Boston; about 70 percent for Chicago, Toronto, Cleveland, and Philadelphia; and 64 percent for BART in San Francisco. Second, approximately 20-35 percent of all riders walk to and from stations on most existing systems (New York excluded). This range may reduce to 15-25 percent for proposed suburban-oriented systems.

Freeway Volumes

Freeway use depends upon the extent of the system and the size of the urban area. The proportion of vehicle-kilometers on freeways increases in general proportion to the relative amount of total capacity provided by freeways. For example (1 km = 0.62 mile):

| Percent Freeway Capacity | Percent Vehicle-Kilometers on Freeways for Population 50 000-200 000 | 1 000 000+ |
|---|---|---|
| 20 | 7 | 12 |
| 40 | 20 | 30 |

From the ranges in observed maximum freeway volumes for urban areas of various sizes, it is clear that forecast daily volumes of over 200 000 should be carefully rechecked, since these are found only in a few very large cities today, on roadways with more than eight lanes.

OVERVIEW

This paper has set forth a summary of a broader research effort undertaken to provide meaningful guidelines and parameters for use in inputs to or verification of the demand forecasting process. The paper and the manual from which it has been extracted provide an important data resource for contemporary urban planning efforts.

Plans call for distribution of the manual to various transportation agencies for their use and review. This will be followed by progressive updating of the manual's 250 tables and charts as new data become available. In this way, the manual will be able to respond to ongoing needs and priorities.

ACKNOWLEDGMENTS

I wish to acknowledge Sam Zimmerman and Thomas Hillegass of the Urban Mass Transportation Administration for their guidance and many constructive suggestions.

REFERENCES

1. F. H. Wynn and H. S. Levinson. Some Considerations in Appraising Bus Transit Potentials. HRB, Highway Research Record 197, 1967, pp. 1-24.
2. Transportation and Parking for Tomorrow's Cities. Wilbur Smith and Associates. New Haven, CT, 1966.
3. Urban Transportation Concepts—Center City Transportation Project. Wilbur Smith and Associates. New Haven, CT, 1970.
4. B. Pushkarev and J. Zupan. Public Transportation and Land Use Policy. Indiana Univ. Press, Bloomington, 1977.

*Publication of this paper sponsored by Committee on Passenger Travel Demand Forecasting and Committee on Traveler Behavior and Values.*

# Minimizing Error in Aggregate Predictions From Disaggregate Models

Fred A. Reid, University of California, Berkeley

**This paper presents empirical tests of aggregated prediction error on a sample of work-trip mode choices for the San Francisco area and systematic criteria for choosing classification variables. It introduces a more efficient utility scale classification criterion for aggregated prediction. Aggregation error is found to be much larger than previous tests have indicated. The choice of classification variables that produces the smallest error is found to vary with the scale of the prediction aggregates. Level-of-service variables are more important for large aggregates, socioeconomic variables for smaller. Classification of the sample based on the scales of the total utility of the explanatory variables in each alternative is found to be much more efficient in error reduction than classification by individual variables.**

A number of papers have developed the theory of making aggregate predictions from disaggregate (individual) demand models (1, 2, 3, 4). Koppelman (1) integrated these theories into general guidelines for aggregate prediction and tested the relative accuracy of different proposed approaches by Monte Carlo simulation and with empirical data for one situation—the demand for different travel modes by Washington, D.C., commuters. He developed the classification method as the most accurate approximate method for aggregate prediction and encouraged classification based on "choice-set availability" variables such as the number of cars per driver in a traveler's household. This contrasted strikingly with the traditional approach to data aggregation—classification on the basis of geographic groups—an approach shown to produce biased results.

Dunbar (5) and Koppelman and Ben-Akiva (6) have further promoted prediction based on cross-classification by the values of the most influential variables in choice.

There are two important problems with these results. The classification developments, although clearly emerging in principle to define relatively homogeneous choice segments, leave practitioners with no systematic method of choosing the type and number of classification variables, especially considering the difference in the relative variance of the variables at different geographic levels of aggregation. Koppelman's tests show that the number of cars per driver is the most important classifier for small geographic aggregates. Dunbar's and my results show that level-of-service variables are more important classifiers for large aggregate forecasting. The empirical variability of these results and the newness of the procedures to planners suggest that systematic methods are needed.

A second problem is that the Koppelman study—the only empirical results isolating the error of the different approximate methods for aggregation—is based on only one choice model and setting and then for an optimistic case. His data were for work-trip mode choice restricted to one destination, the Washington, D.C., central business district (CBD). The data also happened to have nearly equal choice shares over the full sample (the zero aggregation bias condition). Thus aggregation error, at least for regional mode choice, should be higher than shown by Koppelman.

This paper gives aggregation tests by different methods on San Francisco region commuter data, presents a more systematic criterion for identifying the disaggregate data (variables) most important for low error, and shows that a classification method based on the total utility scales of the choice alternatives, rather than on the values of the individual variables, is much more accurate and efficient for aggregation.

## AGGREGATION ERROR TESTS

I have used the models and data from the urban travel demand forecasting project at the University of California, Berkeley, to perform tests isolating error by different methods in different situations (7, 8, 9).

The tests of this study were done on a sample of 771 workers drawn from about half of the San Francisco Bay Area. Their individual mode-choice probabilities for commuting to work were described by the four-alternative logit model

$$p_j = \left(\exp \sum_K \beta_{jk} z_{jk}\right) \Big/ \left(\sum_J \exp \sum_K \beta_{jk} z_{jk}\right) \qquad (1)$$

where $\beta_{jk}$ are model coefficients of the K variables $z_{jk}$ for the jth alternative. These are shown in the table below, where 1 is auto alone (429 people), 2 is bus with walking access (134 people), 3 is bus with auto access (30 people), and 4 is carpooling (178 people). Each independent variable takes the described value in the alternatives listed in parentheses and zero in unlisted alternatives.

| Independent Variable | Estimated Coefficient | t-Statistic |
|---|---|---|
| Cost divided by post-tax wage, in cents divided by cents per minute (1-4) | -0.019 06 | (-2.646) |
| Auto on-vehicle time, in minutes (1, 3, 4) | -0.059 86 | (-5.24) |
| Transit on-vehicle time, in minutes (2, 3) | -0.024 17 | (-2.751) |
| Walking time, in minutes (2, 3) | -0.068 92 | (-5.301) |
| Transfer waiting time, in minutes (2, 3) | -0.055 99 | (-2.361) |
| Number of transfers (2, 3) | -0.060 39 | (-0.442 9) |
| Headway of first bus, in minutes (2, 3) | -0.029 51 | (-2.997) |
| Family income with ceiling of $7500, in dollars per year (1) | 0.000 006 238 | (0.070 16) |
| Family income minus $7500 with floor of $0 and ceiling of $3000, in dollars per year (1) | -0.000 066 25 | (-0.492 0) |
| Family income minus $10 500 with a floor of $0 and a ceiling of $5000, in dollars per year (1) | -0.000 029 11 | (-0.475 0) |
| Number of persons in household who drive (1) | 0.989 4 | (4.667) |
| Number of persons in household who drive (3) | 0.975 6 | (3.272) |
| Number of persons in household who drive (4) | 0.859 3 | (4.223) |
| Dummy if person is head of household (1) | 0.716 4 | (3.736) |
| Employment density at work location (1) | -0.002 888 | (-3.945) |
| Home location in or near CBD (2 is in CBD, 1 is near CBD, 0 is otherwise (1) | -0.454 6 | (-3.694) |
| Autos per driver with a ceiling of one (1) | 4.992 | (9.600) |
| Autos per driver with a ceiling of one (3) | 2.333 | (2.769) |
| Autos per driver with a ceiling of one (4) | 2.385 | (5.315) |
| Auto alone alternative dummy (1) | -5.210 | (-5.868) |
| Bus with auto access dummy (3) | -5.584 | (-5.476) |
| Carpooling alternative dummy (4) | -3.795 | (-6.333) |

The goodness-of-fit statistics are 0.4484 for the likelihood ratio index, -1069.0 for the log likelihood at zero, and -589.6 for the log likelihood at convergence. All cost and time variables were calculated for round trips. The dependent variable is the alternative chosen (value of one for chosen alternative, zero otherwise).

The overall exact mode shares were 55.6 percent auto alone, 17.4 percent bus with walking access, 3.9 percent bus with auto access, and 23.1 percent carpooling. The data were from household surveys and transportation network minimum-path simulations, modified to produce trip attributes temporally and spatially disaggregated to the commuters' schedules and trip ends.

Exact aggregate choice shares are obtained by summing the probabilities Equation 1 over the individual $z_{jk}$ values for the prediction aggregate of interest (the enumeration method). Other (approximate) methods are motivated by the large data and computation requirements of this enumeration in practice. The extreme approximation—the "naive" method—assumes that Equation 1 represents aggregate shares when the $z_{jk}$ are simply the average values for the prediction group.

The measure of aggregation error in the tests here is the percent root mean square (RMS) of choice shares:

$$\epsilon_{RMS} = \left(\sum_J \{[(\hat{P}_j - P_j)/P_j]^2\} P_j\right)^{1/2} \qquad (2)$$

where

J = set of choice alternatives,
$\hat{P}_j$ = aggregate share of alternative j estimated by the tested method, and
$P_j$ = aggregate share by enumeration.

The percentages of error from the naive aggregation method applied at three geographic levels on our sample are shown below. Predictions are all for the total region. The input data were the averages in the geographic aggregates shown. Thus, results represent either of two forecasting situations: geographic classification for full region predictions or the average absolute errors when making separate predictions for all the cells at a geographic (classification scale) level. These errors are approximately

| Classification Scale | No. of Cells | Percent Error |
|---|---|---|
| Region | 1 | 40.0 |
| Cities | 17 | 17.9 |
| Traffic analysis zones | 150 | 13.8 |

The errors are large. It is obvious that geographic classification alone is not an adequate aggregation method for this region and model. Errors decrease with geographic scale. Smaller area samples apparently do have less choice variability. Since classification-aggregation was done on the basis of residential district only, these results do not represent what would be expected from trip interchange forecasts. It is expected that average interdistrict aggregation errors would be about half those shown. Individual interdistrict aggregate errors would be worse.

Koppelman's results were lower in magnitude and did not vary with geographic scale. He showed an 8.5 percent error for a superdistrict scale similar to our cities, which show 17.9 percent. The regional errors were 10.4 and 40 percent, respectively. Several reasons account for this difference: he had a CBD-trip-oriented sample only; his average shares were nearly equal; the choice model he used was simpler; and the level-of-service data were less disaggregate (10).

Both tests aggregated (classified) the input data based only upon the origin of residence. Other sources of incomparability—the number of choice alternatives and the nonlinearity of the measure—are not large for these data. A much simpler model, equivalent to Koppelman's variable set, was found to produce 80 percent of the naive error on otherwise equivalent conditions for this data set (9, Chapter 6).

The error on our sample shown by five available methods of aggregation is

| Method | No. of Cells | Percent Error |
|---|---|---|
| Naive | – | 40.0 |
| Taylor series | – | 121.0 |
| Classification by city | 17 | 17.9 |
| Classification by auto ownership | 4 | 21.7 |
| Classification by utility scale | 4 | 3.1 |

Choice share prediction is again for the total (regional) sample. The naive method and the geographic classification method are the same as for the previous table. The Taylor series method is that of Talvitie (10). The by-variable value classification method is Koppelman's (1). The utility classification method is described later.

All of the methods except the Taylor series reduce error below that of the naive procedure. These results confirm earlier tests that the Taylor series approximation produces counterproductive results due to its poor convergence properties on large variance data. Unfortunately, it is in such data that the correction is important. Geographic classification, as shown above, is very inefficient in reducing aggregation error.

The predictions by classes of auto ownership also only reduce about half the naive error, giving unacceptable accuracy at the regional scale of forecasting. Koppelman showed auto availability classification to reduce a smaller regional naive error by two-thirds, with a result of 3 percent. Granting that the variable cars per driver will reduce two-thirds of the naive error, this classifier would still leave an unacceptable error in general regional aggregate travel predictions (for trips to all destinations and with unequal shares). Choice set availability classifiers used on predictions for subcity or city interchange level aggregates may yield more acceptable errors (below a third of 18 percent).

The method of utility classification gives a much greater reduction relative to naive error than the other methods. Only four class cells were used. This method is discussed below.

## UTILITY SCALE CLASSIFICATION

Although classification by the values of one or two variables gives unacceptable error for large aggregate predictions, the error can be made small by cross-classification of more of the variables in a model. However, if more than a few variables contribute significantly to the variance of the utility of the choices, the number of cross-classifications must be large to achieve small error.

A more efficient method of classification is possible for the category of simply scalable models such as those of the logit form. Cross-classification between individual variables includes much information that does not matter to simple scales. The essential information needed to predict each individual's choice in these models is contained in the J utility scales of the attributes of the J alternatives of choice. Cross-classification between the utilities of the different alternatives picks up the full-scale variances and between-scale covariances, thus describing the full distribution of individual choice factors in an aggregate prediction sample. Regardless of scale complexity, this procedure bypasses those individual variable cross-classification trade-offs, which do not change the scale values.

Thus, the procedure requires fewer classes. Classification on the total utility includes the variances of the minor variables, not just the variance of the limited number of interactions feasible in classification by a subset of model variables. This further increases its efficiency.

To define relatively homogeneous classes of utility combinations across modes is to probe the essence of the classification approach: the grouping of individuals with uniform choice situations. Since the procedure operates on the utility scales, it is termed the utility scale classification method of aggregation.

Utility class sizes and boundaries cannot be defined in the same way as can individual variable classifications. Utilities are not discrete, and intuition does not give the guide on thresholds of utility in choice that it does on a variable such as walking time to a bus in mode choice. However, utility values are clearly related to choice function thresholds. For the binary logit case, the optimum divisions would differentiate utilities near the maximum nonlinearity of the choice function (±1.6 on the utility difference scale). For multiple-choice logit model classification, the criteria should concentrate on pairs of alternatives with these same differences of utilities.

Cluster analysis techniques could be employed to achieve general isolation of classes (11, 12). The tests here and in supporting research show that ad hoc class divisions are adequate and more appropriate, considering the above nonlinearities and the press of obtaining predictions (9).

One such procedure is the successive division of an aggregate sample about the median (or mean) values of the differences of the utilities of pairs of alternatives, starting with the pair with the largest variance in utility difference. The procedure cycles through all pairs of utilities, further subdividing the first pairs if the variance of utility differences in the resulting classes are still large. The variance criteria for desired accuracy can be estimated with the covariance analysis procedure discussed in the next section.

The utility ranges gave a much smaller error than any of the individual variable or geographic classifications considered in the tests in the tables with only two utility classes on a four-alternative model. Clearly this method is more efficient in error reduction than the others. This is to be expected at the regional level, where the total utility variance of all of the variables is present. It would also be true for predictions for small aggregates unless only a minor subset of model variables dominated the variability of these subsamples. Such is not the case for within-city aggregates, as will be seen in the following percentages of aggregation error by cell count.

| Class Cell Count | Percent Error |
|---|---|
| 1 (naive method) | 38.4 |
| 2 | 6.4 |
| 4 | 2.3 |
| 6 | 2.6 |
| 8 | 0.5 |

This shows the variation in aggregation error with the number of class cells used in the utility classification method. These results are shown for a three-alternative subset of the four-mode choice model used in the previous tests (auto alone, bus with walking access, and carpooling). The cell definition criteria were those above. The errors seem acceptable with only four cells. With eight cells error becomes negligible. The instability in error reduction with cell count is due partly to the ad hoc cell definition criteria used but also to the nonlinearity of the RMS measure. An error measure composed of a linear sum of absolute errors in each alternative decreased monotonically with cell count (9).

Since utility scale classification does not differentiate the joint distributions of all variables in a model, it does not apply to general models that are not simply scalable. Probit models and others that allow estimation of different coefficients for the same variable in different alternatives are examples (13, 14). Aggregate predictions with these, in general, require individual variable cross-classification or covariance procedures (see the paper by Bouthelier and Daganzo in this Record). However, even non-simply-scalable models are not likely to estimate cross-alternative coefficients for many variables.

In view of the laborious nature of cross-classification and the efficiency of utility classes on the simple scales, the latter may be the most effective even in this case. Alternatively, it may be efficient to cross-classify utility classes with an important cross-alternative variable.

The focus of this aggregation method on the total utility value need not preclude the retention and association of the values of specific explanatory variables with the classes. This makes it possible to intuitively interpret the cells, analyze the effects of policy changes on choices, and predict subsegment choices. These can all be accommodated by characterizing each class cell by the average value of any desired model variable or socioeconomic descriptor, in addition to the average value of the J total utilities of its members. Together with the values of the underlying model scale coefficients, the mean utilities may be adjusted for policy changes on specific variables to simply analyze their effects on aggregate choices. Population subgroup shares may be predicted by summing the products of the proportions of these groups and the shares in each cell. A complete description and examples of the use of utility classification method in subsegmented regional travel prediction and policy analysis are given in a related report (15).

This, like any accurate classification method, requires that individual choice-maker data observations be available to compute the cell mean values. Hence, the principal advantage of the method is in the efficiency of the predictive calculations, given the data. This is no small advantage, since predictions can require exponentiation over many choices for each individual in large samples or they may be desired for numerous policy (input) changes. Once the disaggregate data information is reduced to cell means and parameters, predictions for different policies or output segments or both are simple enough to make hand calculation feasible (15).

It is tempting to believe that by-variable-value class means are simply available from aggregated planning or census data. However, data cannot be directly collected in autos per driver classes or other cross-classes necessary for accurate aggregation. These require individual observations. This method may also indirectly aid the data-collection process. By identifying the minimum amount of information necessary for prediction, it focuses on these requirements. What is necessary for logit models are only the utility scale variances, not the unique within-alternative distributions of variable values. It may be possible to directly extract the utility scale distribution information in a sample directly from census cross-tabulations (9, 16) or from minor subsets of the model variables. The latter approach is discussed in the next section.

## COVARIANCE MATRIX ANALYSIS OF THE EXPLANATORY VARIABLES

Whatever the method of aggregation, the requirement for disaggregate data against a past experience of aggregate data collection suggests that data collection may have to focus on the major sources of error. Thus, it is necessary to know which variables in a model contribute the greatest part of the covariances of the utilities within any prediction aggregate. Precedent and intuition have shown the importance of some variables such as socioeconomic descriptors and transit access variables in mode choice. The problem is that the important variables for classification vary with the geographic scale of the aggregate predictions and, to the extent that there is no universal model of behavior, with the model being used. The limited precedents available may be for the wrong cases. It is unreasonable to expect expert judgment or research-level optimization of aggregation accuracy in practical planning situations.

For effective use of cross-classification methods, where necessary, one must have knowledge of the variables contributing the most to utility covariances at the level of aggregation.

A systematic picture of priorities of individual variable data can be gained by looking at the covariance matrixes of the within-aggregate variation of the explanatory variables at the desired level of aggregation. Binary choice aggregation theory has established the direct relationship between this matrix and aggregate shares for probit models (7). I have shown that this simplified analytic form of the aggregation method for probit models and normally distributed explanatory models can be extended from probit to an approximate relation for logit models with an error less than 1.2 percent of total shares (17).

$$S_j/N = 1 + \exp\left((\overline{V_i - V_j})/\{[1 + \mathrm{Var}(V_i - V_j)/2.79]^{1/2}\}\right) \qquad (3)$$

where

$S_j/N$ = choice share of the alternative j,
$V_j$ and $V_i$ = utilities of the two choices, and
$\overline{V_i - V_j}$ and $\mathrm{Var}(V_i - V_j)$ = mean and variance of their difference in the aggregate of N travelers.

The assumption of normality need apply only to the scale utility difference, not to each explanatory variable. Tests have shown this assumption to contribute 4.0 percent error for regional aggregate predictions that had 43 percent error by the naive method, suggesting that normality is a reasonable assumption (9). The method is still limited to binary choice.

Although only a more complicated relationship of this type exists in the multiple choice case, the covariance matrixes are still better indicators of which variables are appropriate classifiers than a priori knowledge, especially if these vary with aggregate level.

Table 1 shows a normalized covariance matrix of the intracity aggregate values of the major explanatory variables in two alternatives of a prediction model. The model used was shown at the beginning of the paper. The covariance elements are normalized by dividing the largest of their values (in this case it is the variance of the

**Table 1. Normalized covariance matrix of explanatory variables for intracity portions of individual observations.**

| Variable[a] | Cars/ Driver | Auto On-Vehicle Time | No. of Drivers | Employment Density | Cost-Auto | Bus With Walking Access | Bus On-Vehicle Time | Headway | Transfer Waiting Time |
|---|---|---|---|---|---|---|---|---|---|
| Cars/driver | 1.0 | 0.14 | <[b] | < | < | < | < | < | |
| Auto on-vehicle time | 0.14 | 0.59 | < | 0.10 | 0.10 | -0.09 | 0.27 | < | 0.06 |
| No. of drivers | < | < | 0.25 | < | | < | < | | |
| Employment density | < | 0.10 | < | 0.11 | | -0.05 | | | |
| Cost-auto | < | 0.10 | | | 0.04 | < | | | |
| Bus with walking access | < | -0.09 | < | -0.05 | | 0.98 | 0.38 | 0.05 | |
| Bus on-vehicle time | < | 0.27 | < | < | < | 0.38 | 0.25 | < | 0.07 |
| Headway | < | < | < | | | 0.05 | < | 0.07 | < |
| Transfer waiting time | < | 0.06 | | | | < | 0.07 | < | 0.15 |

[a] Covariance values are normalized by multiplying the variances of the variables by their respective model coefficients and dividing by the maximum such utility component ($\beta_0 S_{00}^2 \beta_0$) in this case the utility variance of cars per driver having the value 1.92. Individual observations are minus the mean variable values for all travelers in sample with work trips between the same cities of origin and destination.

[b] < indicates the element is less than 0.04 (0.08 unnormalized); blank indicates the value was not computed but judged to be less than 0.04.

utility of cars/drivers in the traveler's household) in order to give a picture of the ranking of the contributions of each variable to the overall utility variance.

The underlying model is a logit function, linear in its explanatory variables. Hence, each element in the matrix is the product of the covariances of the individual variables and their corresponding model coefficient is

$$\sum_{kl}^{2} = \beta_k S_{kl}^2 \beta_l / \beta_0 S_{00}^2 \beta_0 \tag{4}$$

where

- K and l = matrix indexes of the K variables in the model ($1 \le k$, $1 \le K$),
- $\beta_k \beta_l$ = linear utility coefficients for the kth and lth variables, and
- $\beta_0 S_{00}^2 \beta_0$ = corresponding utility variance for the largest matrix element.

These elements are called "utility component covariances," since each expresses the part of the total sample variance of the linear utility for one mode contributed by a single variable (diagonal variance elements) or pairs of variables (covariance elements). The sum of the matrix elements for any mode equals the total variance of the utility for that mode.

For binary logit choice models the relationship between these covariances and aggregation error is understood under the assumption of a normal distribution of the utility scale of the explanatory variables (Equation 3). Thus Table 1 provides a convenient way to rank the individual variables, and their combinations, by their importance for reducing aggregation error for binary choice.

The larger the variance of the difference of utilities between the two modes, the larger the aggregation error.

Since most of the variables in this model are unique to a mode, variances of intermodal utility differences nearly equal the individual modal elements. Where the same variables appear in the utility expression for two modes, the differences of utilities can be obtained from the table by this expression for the variance of such a difference: Var (X - Y) = Var X + Var Y - 2 Cov (X, Y).

The variables that have the individual values most important for reducing aggregation are the ones with the largest variance values in Table 1, such as bus with walking access, cars/driver, and auto on-vehicle time. The intermodal covariances show the relative importance of pairs of variables in aggregation. However, since intermodal elements contribute to the variance of binary utility difference with negative sign, it is the negative values such as that between cars/driver and bus with walking access that increase the aggregation correction. Positive values indicate a compensating reduction in aggregation error.

The amount of aggregation error that will result in a forecast using the individual values of only selected variables can be estimated by comparing Equation 3 with and without the unnormalized values of the sum of the matrix elements for the variables considered. [Only the individual values of some variables or their variance may be available in a forecasting situation. In some cases, variances are possible to obtain or estimate where covariances are not (10).]

Inspection of Table 1 shows that a large part of the variance of the utilities, and hence aggregation correction, can be recovered by considering only a few of the variables in the model. For example, three of the eleven variables in the model used here—cars/driver, bus with walking access, and bus on-vehicle time—constitute 64 percent of the utility variance. When only their individual values are used for binary prediction, they correct for 80 percent of the aggregation error produced by the naive method. Auto on-vehicle time, in contrast, contributes nothing to the correction of aggregation error by its inclusion with this set, since its covariance with bus with walking access cancels out its variance contribution alone.

The covariance matrix also contains the information for guiding the classification method in its implied correlations between the classifier variable(s) and the others. The additional power of a classifier from this effect is given by the sum of the product of covariances and correlation coefficients of all variables correlated to the classifier. With this procedure, one can minimize aggregation error, given limits on the number of classification variables, or minimize the number of classifiers, given tolerable limits on error. This is systematically outlined elsewhere (11).

This analysis is only simply applicable to binary choice. For multiple choice it requires a relationship analagous to Equation 3 between covariances and choice. The only example of this is the approximate formulas of Bouthelier and Daganzo for probit models.

In principle, it is possible to use any of the above effective subsets of disaggregate variable data for aggregation accuracy using these formulas and covariances between all pairs of alternatives. In practice, or if the model in question is not multinomial probit, simpler approximate guides to the important aggregation variables can be gained by observing the covariance matrixes of major pairs of the alternatives. These should be alternatives that have aggregates different from 1/J and considerably different from one another.

The procedure also does not account for skewed utility distributions. Skewedness will weaken the analysis. This does not appear to be a problem in the tests on this data set (9).

Covariance analysis on a small population sample for which policy-forecasting data are received can guide the subsequent cost-effective emphasis on which variables to focus collection of individual observations in the larger forecasting sample.

## CONCLUSIONS

On the basis of the above exposition, I have drawn the following seven conclusions.

1. The tests show that aggregation error by all methods is larger than previously revealed. Naive method error can be larger than other errors in prediction, such as model specification or data measurement.
2. The relative ranking of the accuracy of existing approximate aggregation methods by Koppelman is confirmed. However, error is found to grow substantially with aggregate size. Previous studies have underestimated this and general error level because their samples had limited trip types, choice share ranges, and data disaggregation.
3. None of the existing simple approximation methods gives even marginally acceptable (≤13 percent) error at the urban regional scale of mode-choice aggregation using multinomial logit models.
4. The method of classifying on the differences of utility scales of major choice alternatives gives at least a five times lower error to class cell-count ratio than other classification criteria. It can reduce naive error by a factor of 77 (to 0.5 percent) with only eight cells.
5. The choice homogeneity of utility classification cells eliminates the need to collect data and make separate aggregations for each sector of a population for which share predictions are desired. It replaces such repetitive choice function evaluation with a smaller number of class segment proportion calculations.
6. Analysis of the covariance matrix of the utility components of the individual model variables answers the question of which variables should be stressed in disaggregate data collection. Doing this on a small sample for the prediction population of interest can guide the subsequent data collection to the subset of variables that gives the least accuracy return. This analysis can also give systematic guidance to the selection of by-variable classification methods for predictions with non-simply scalable models.
7. Predictions, even for large aggregates, can be made accurately with disaggregate data on only a minor subset of critical variance variables. Aggregation can be accomplished with the minimum of choice function computations using utility classification on the full data sample rather than by-variable classification on every prediction segment.

Classification methods of prediction, especially those based on selecting homogeneous utility or choice groups, realize the full potential for efficiency of disaggregate models that was first seen when it was found that they could be calibrated on a fraction of the observations necessary for aggregate models. Now prediction can also be done on samples large enough to identify only a limited number of utility classes, that is, homogeneous choice markets.

Predictions for subaggregates require larger samples but are simply a process of determining the relative numbers of the population in each of these classes. The prediction problem has been reduced from one of computing choice shares for great numbers of aggregate outputs to locating the proportions of these aggregate cases in a small number of behavioral market segments. Selected disaggregate data are necessary to gain the accuracy of behavioral models, but predictive computations are greatly reduced. Reid (9) and Reid and Harvey (15) discuss these aggregation methods and their application to policy analysis in greater depth.

## ACKNOWLEDGMENT

This research was supported in part by a National Science Foundation grant through the Research Applied to National Needs Program to the University of California, Berkeley. I am indebted to Daniel McFadden and Ibrahim Hasan for discussion and comment on these subjects. Mr. Hasan, Chris Murano, and Rafi Melnick performed valuable supporting computations. I am solely responsible for the results.

## REFERENCES

1. F. S. Koppelman. Travel Predictions With Models of Individual Choice Behavior. Center for Transportation Studies, Massachusetts Institute of Technology, Cambridge, Rept. 75-7, 1975.
2. D. McFadden and F. A. Reid. Aggregate Travel Demand Forecasting From Disaggregated Behavioral Models. TRB, Transportation Research Record 534, 1975, pp. 24-37.
3. A. Talvitie. Aggregate Travel Demand Analysis With Disaggregate or Aggregate Travel Demand Models. Proc., Transportation Research Forum, 1973.
4. R. B. Westin. Predictions From Binary Choice Models. Journal of Econometrics, 1974, pp. 1-16.
5. F. C. Dunbar. Policy-Contingent Travel Forecasting With Market Segmentation. TRB, Transportation Research Record 637, 1977, pp. 27-32.
6. F. S. Koppelman and M. E. Ben-Akiva. Aggregate Forecasting With Disaggregate Travel Demand Models Using Normally Available Data. Paper presented at the World Conference on Transport Research, Rotterdam, April 26-28, 1977.
7. D. McFadden and others. Demand Model Estimation and Validation. Urban Travel Demand Forecasting Project, Institute of Transportation Studies, Univ. of California, Berkeley, Final Rept. Series, Vol. 5, June 1977.
8. F. A. Reid. Disaggregated Supply Data and Computation Procedures. Urban Travel Demand Forecasting Project, Institute of Transportation Studies, Univ. of California, Berkeley, Final Rept. Series, Vol. 3, 1977.
9. F. A. Reid. Aggregation Methods and Tests. Urban Travel Demand Forecasting Project, Institute of Transportation Studies, Univ. of California, Berkeley, Final Rept. Series, Vol. 7, 1978.
10. A. Talvitie and Y. Dehghani. Supply Model for Transit Access and Line Haul. Urban Travel Demand Forecasting Project, Institute of Transportation Studies, Univ. of California, Berkeley, Working Paper No. 7614, July 1976.
11. P. E. Green and Y. Wind. Multi-Attribute Decision Marketing: A Measurement Approach. Holt, Rinehart and Winston, New York, 1973.
12. H. P. Friedman and J. Rubin. On Some Invariant Criteria for Grouping Data. American Statistical Association Journal, Dec. 1967.
13. J. A. Hausman and D. A. Wise. A Conditional Probit Model for Qualitative Choice: Discrete Decisions Recognizing Interdependence and Heterogeneous Preferences. Department of Economics, Massachusetts Institute of Technology, Cambridge, Working Paper No. 173, 1976.
14. D. McFadden. A Closed Form of Multinomial Choice Model Without the Independence from Ir-

relevant Alternatives Restrictions. Urban Travel Demand Forecasting Project, Institute of Transportation Studies, Univ. of California, Berkeley, Working Paper No. 7703, 1977.
15. F. A. Reid and G. Harvey. Regional Policy Analysis Case Study. Urban Travel Demand Forecasting Project, Institute of Transportation Studies, Univ. of California, Berkeley, Final Rept. Series, Vol. 9, 1978.
16. D. McFadden and others. Demographic Data for Policy Analysis. Urban Travel Demand Forecasting Project, Institute of Transportation Studies, Univ. of California, Berkeley, Final Rept. Series, Vol. 8, June 1977.
17. F. A. Reid. A Set of Models for Optimizing the Benefits of a Transportation Plan. Univ. of California, Berkeley, M.S. thesis, 1974.

*Publication of this paper sponsored by Committee on Passenger Travel Demand Forecasting and Committee on Traveler Behavior and Values.*

# Disaggregate Demand Model for Nonwork Travel

Joel Horowitz, U.S. Environmental Protection Agency

**Daily nonwork travel by urban households frequently involves visits to several destinations during a single roundtrip from home or several round trips. This paper describes a disaggregate approach to modeling the demand for nonwork travel, including multidestination travel. The approach is presented in two parts. First, a theoretical framework for modeling the demand for nonwork travel is developed that uses tours (i.e., round trips from home) and sojourns (i.e., visits to nonwork destinations) as the basic units of travel, and relates tour frequency, sojourn frequency, and destination choice to household characteristics, destination characteristics, and transportation level of service. An empirical model of nonwork travel demand that is based upon the theoretical framework is then presented. The empirical model enables tour and sojourn frequencies, destination choice, and nonwork vehicle-kilometers traveled to be computed as functions of household characteristics, destination characteristics, and transportation level of service. Several tests of the validity of the empirical model are described. The model was found to perform well.**

One of the problems that must be addressed in modeling the demand for urban nonwork travel is that of incorporating multidestination travel into the modeling framework. Current operational models of nonwork travel demand use one of two methods to treat multidestination travel. In the first, only two nonwork travel options are explicitly modeled: the option of doing no nonwork travel during a day and the option of taking a single round trip between home and nonwork destination during a day. Multidestination travel is ignored or its effects are approximated through relatively crude factoring procedures. This method is used frequently in connection with disaggregate models of nonwork travel demand (1, 2, 3).

The other method of treating multidestination travel uses individual trip links as the basic units of travel and models the demand for these links. Effects of multidestination travel are represented by changes in the demand for home-based and non-home-based trip links. This method is used in conventional aggregate demand models (4) and in at least one disaggregate demand model (see the paper by Ben-Akiva, Sherman, and Kullman in this Record). It also is used in Markovian activity sequencing models (5, 6, 7). The method implicitly assumes that travelers' decisions are based only on the characteristics of individual trip links without regard to the characteristics of groups of links or travel patterns. Distortions that this assumption causes in travel demand models are described elsewhere (8, 9).

Neither of the foregoing methods provides a satisfactory treatment of multidestination travel: the first virtually ignores multidestination travel, and the second ignores relationships between the trip links that make up multidestination travel patterns.

Two other approaches to treating multidestination travel have been suggested as means of overcoming these difficulties. In one of these approaches complete travel patterns of households during 24-h periods are used as the basic units of travel, and the demand for individual travel patterns is modeled (8). This approach offers a theoretical framework in which travel decisions depend on relationships between trip links as well as on characteristics of individual links. However, it provides no means of distinguishing between the set of travel patterns actually considered by households in the process of making travel decisions and the virtually infinite set of travel patterns that are, in principle, available for consideration.

The second approach uses the tour and the sojourn as the basic units of travel (10). A sojourn is a visit to a place other than home or work. A tour is a movement that begins and ends at home or work and includes one or more sojourns. The approach consists of modeling the demand for tours and sojourns. The detailed sequence of sojourns within tours is not modeled, thus avoiding the need to consider a potentially infinite variety of tour structures. On the other hand, because the tour is a multidestination unit of travel, it is not necessary to assume that travel decisions are made without regard to collective characteristics of trip links.

This paper describes a disaggregate model of the demand for nonwork tours and sojourns. The model is presented in two parts. First, using the household as the basic decision-making unit, a theoretical framework is developed that relates tour frequency, sojourn frequency, and destination choice to household characteristics, destination characteristics, and transportation level of service. An empirical model of nonwork travel demand that is based upon the theoretical framework is then presented. The empirical model, which was estimated using travel data from the Washington, D.C., area, enables tour and sojourn frequencies, destination choice, and nonwork vehicle-kilometers traveled to be computed as functions of household characteristics, destination characteristics, and transportation level of service.

The modeling approach that is presented here is intended to be prototypical rather than operational. In keeping with this orientation, certain aspects of nonwork travel demand are not treated. Tours that originate or terminate at work are not dealt with, and it is assumed that the automobile is the only available mode for nonwork travel. Only automobile driver tours and sojourns are represented explicitly in the framework. These can be converted to person tours and sojourns if exogenous information on automobile occupancy is available, but this is not done here. These limitations are not intrinsic to the modeling approach, but they greatly simplify its development and application. With the exception of the exclusion of transit travel from the framework, the limitations are present in most operational disaggregate models of nonwork travel. Means of removing the limitations are discussed at the end of this paper.

## THEORETICAL FRAMEWORK

A theoretical framework for modeling nonwork travel demand is presented in this section. The principal assumptions of the framework are described in the discussion that begins in the next paragraph and ends with Equation 2. The mathematical derivation of a demand model based on these assumptions follows Equation 2. The presentation and interpretation of the resulting model begin with Equation 43.

Households are assumed to be the basic decision-making units. Let A be the set of nonwork destinations available to a household. A is defined to include home. Let t be the time of day, and let $\Delta t$ be a time interval sufficiently short that the household can begin at most one automobile driver trip in the period t to t + $\Delta t$. Finally, let $A_t \subseteq A$ be the set of nonwork destinations to which the household members can begin travel during t to t + $\Delta t$. $A_t$ will not necessarily contain all of the elements of A. For example, if all members of the household are at nonwork destination i at time t, then $i \in A$ but $i \notin A_t$.

The travel choices available to the household during t to t + $\Delta t$ are

1. An automobile driver, possibly accompanied by others, begins a trip to destination $i \in A_t$ ($i \neq$ home) as part of a tour from home to one or more nonwork destinations to home. Destination i is not necessarily the first destination of the tour.
2. An automobile driver, possibly accompanied by others, begins a trip from a nonwork destination to home. This option is available only if home is in $A_t$.
3. No trips of the above types are begun.

The following utilities are associated with each of these options:

Option 1. $U_i(x, s, z, N_o, t, \Delta t) + \epsilon_i$
Option 2. $U_h(s, N_o, t, \Delta t) + \epsilon_h$
Option 3. $U_o(s, N_o, t, \Delta t) + \epsilon_o$

where

U = deterministic component of utility,
$\epsilon$ = random component of utility,
x = vector of transportation level-of-service variables relevant to the choice of destination i,
s = vector of household characteristics,
z = vector of destination characteristics other than transportation level of service relevant to the choice of destination i, and
$N_o$ = number of automobile driver trips to nonwork destinations other than home that begin at times other than t to t + $\Delta t$.

$N_o$ is included in the utility functions on the hypothesis that a household has limited travel resources (e.g., time, money, automobiles) and, hence, that the decision to begin a trip during t to t + $\Delta t$ may depend on the number of trips taken at other times of day. The model thereby incorporates the concept that a household considers both its past travel decisions and future travel plans when making current travel decisions.

It is assumed that the distribution properties of the $\epsilon$'s are such that the probabilities of the various travel options can be represented by the multinomial logit model (1). Then the probability of beginning travel to $i \in A_t$ ($\bar{i}$ = home) during t to t + $\Delta t$ is

$$p(i|N_o,t,\Delta t,A_t) = \exp(U_i)/[\exp(U_o) + \sum_{j\epsilon A_t} \exp(U_j)] \quad (1)$$

and the probability of beginning travel to home if home is in $A_t$ is

$$p(h|N_o,t,\Delta t,A_t) = \exp(U_h)/[\exp(U_o) + \sum_{j\epsilon A_t} \exp(U_j)] \quad (2)$$

The members of a household can make only a finite number of trips during a day. Thus, it is reasonable to suppose that as $\Delta t$ approaches zero, the probability of travel during t to t + $\Delta t$ also approaches zero and that for small $\Delta t$ the probability of travel is proportional to $\Delta t$

$$p(i|N_o,t,\Delta t,A_t) = p(i|N_o,t,A_t)\,\Delta t \quad (3)$$

$$p(h|N_o,t,\Delta t,A_t) = p(h|N_o,t,A_t)\,\Delta t \quad (4)$$

This implies that for small $\Delta t$

$$U_i(x,s,z,N_o,t,\Delta t) = V_i(x,s,z,N_o,t) + \ln \Delta t \quad (5)$$

$$U_h(s,N_o,t,\Delta t) = V_h(s,N_o,t) + \ln \Delta t \quad (6)$$

$$U_o(s,N_o,t,\Delta t) = V_o(s,N_o,t) \quad (7)$$

Thus for small $\Delta t$ Equations 1 and 2 become

$$p(i|N_o,t,A_t)\,\Delta t = [\exp(V_i - V_o)]\,\Delta t \quad (8)$$

$$p(h|N_o,t,A_t)\,\Delta t = [\exp(V_h - V_o)]\,\Delta t \quad (9)$$

The dependence of the travel probability Equations 8 and 9 on $A_t$ now is removed by forming the following marginal travel probabilities:

$$p(i|N_o,t)\,\Delta t = \sum_{\substack{A_t \\ i\epsilon A_t}} P(A_t)\, p(i|N_o,t,A_t)\,\Delta t \quad (10)$$

$$p(h|N_o,t)\,\Delta t = \sum_{\substack{A_t \\ h\epsilon A_t}} P(A_t)\, p(h|N_o,t,A_t)\,\Delta t \quad (11)$$

where $P(A_t)$ is the probability that $A_t$ is the set of available travel options during t to t + $\Delta t$. Substituting Equations 8 and 9 into Equations 10 and 11 yields

$$p(i|N_o,t)\,\Delta t = [\exp(V_i - V_o)]\,\Delta t \sum_{\substack{A_t \\ i\epsilon A_t}} P(A_t) \quad (12)$$

$$p(h|N_o,t)\,\Delta t = [\exp(V_h - V_o)]\ \Delta t \sum_{\substack{A_t \\ h\epsilon A_t}} P(A_t) \quad (13)$$

The summations in Equations 12 and 13 respectively are the probabilities that destinations i and home are available to the household during t to t + $\Delta t$. These probabilities can be absorbed into $V_i$ and $V_h$ through the transformations

$$V_i \rightarrow V_i + \ln \sum_{\substack{A_t \\ i\epsilon A_t}} P(A_t) \quad (14)$$

$$V_h \rightarrow V_h + \ln \sum_{\substack{A_t \\ h\epsilon A_t}} P(A_t) \quad (15)$$

to yield the following travel probabilities:

$$p(i|N_o,t)\,\Delta t = [\exp(V_i - V_o)]\ \Delta t \quad i\epsilon A \quad (16)$$

$$p(h|N_o,t)\,\Delta t = [\exp(V_h - V_o)]\ \Delta t \quad (17)$$

The conditioning of the travel probabilities on $N_o$ now is removed by forming the marginal probabilities

$$p(i|t)\,\Delta t = \sum_{N_o} P(N_o)\ p(i|N_o,t)\,\Delta t \quad (18)$$

$$p(h|t)\,\Delta t = \sum_{N_o} P(N_o)\ p(h|N_o,t)\,\Delta t \quad (19)$$

where $P(N_o)$ is the probability that $N_o$ trips to nonwork destinations other than home start during times other than t to t + $\Delta t$. Substituting Equations 16 and 17 into Equations 18 and 19 yields

$$p(i|t)\,\Delta t = \sum_{N_o} P(N_o)\ [\exp(V_i - V_o)]\ \Delta t \quad (20)$$

$$p(h|t)\,\Delta t = \sum_{N_o} P(N_o)\ [\exp(V_h - V_o)]\ \Delta t \quad (21)$$

At this point it is useful to be more specific about the structure of the utility functions. Let $V_i - V_o$ and $V_h - V_o$ have the following functional forms:

$$V_i - V_o = F_i(x,s,z) + G(N_o,s) + L(t) \quad (22)$$

$$V_h - V_o = H(N_o,s) + K(t) \quad (23)$$

Then Equations 20 and 21 become

$$p(i|t)\,\Delta t = \exp[F_i(x,s,z) + L(t)]\ \Delta t \sum_{N_o} P(N_o) \exp[G(N_o,s)] \quad (24)$$

$$p(h|t)\,\Delta t = \exp K(t)\ \Delta t \sum_{N_o} P(N_o) \exp H(N_o,s) \quad (25)$$

Note that $\exp H(0, s) = 0$. A trip to home during t to t + $\Delta t$ cannot take place unless there was travel to a nonwork destination away from home during a previous time period. $N_o = 0$ implies that no such travel occurred.

The $N_o$ dependence of $\exp G(N_o, s)$ now is approximated by a first order Maclurin's series in $N_o$:

$$\exp G(N_o,s) \simeq \exp G(0,s) + [\partial G(0,s)/\partial N_o][\exp G(0,s)]\ N_o \quad (26)$$

To develop an analogous approximation for $\exp H(N_o, s)$, it is necessary to express $\exp H(N_o, s)$ in a manner that explicitly represents its behavior at $N_o = 0$. This can be done by defining $J(N_o, s)$ such that

$$\exp H(N_o,s) = N_o J(N_o,s) \quad (27)$$

Assuming that $\partial J/\partial N_o$ is continuous $N_o = 0$, the first order approximation to $\exp H(N_o, s)$ is

$$\exp H(N_o,s) \simeq N_o J(0,s) \quad (28)$$

Now define $\overline{N_o}$ to be the average value of $N_o$:

$$\overline{N_o} = \Sigma P(N_o)\ N_o \quad (29)$$

Then Equations 24-29 imply that

$$p(i|t)\,\Delta t = \exp[F_i + L + G(0,s)]\ [1 + \overline{N_o}\ \partial G(0,s)/\partial N_o]\ \Delta t \quad (30)$$

$$p(h|t)\,\Delta t = \overline{N_o}\ J(0,s)\ [\exp K(t)]\ \Delta t \quad (31)$$

As $\Delta t$ approaches zero, $\overline{N_o}$ approaches N, the average number of automobile driver trips per day to nonwork destinations other than home. Thus, Equations 30 and 31 can be written

$$p(i|t)\,\Delta t = \exp[F_i + L + G(0,s)]\ [1 + N\ \partial G(0,s)/\partial N_o]\ \Delta t \quad (32)$$

$$p(h|t)\,\Delta t = NJ(0,s)\ (\exp K)\ \Delta t \quad (33)$$

Because $\Delta t$ is defined to be an interval of time sufficiently short that at most one automobile driver trip can begin during t to t + $\Delta t$, $p(i|t)\Delta t$ and $p(h|t)\,\Delta t$ respectively are equal to the average number of automobile driver trips to destination i and to home that start during t to t + $\Delta t$. The average number of automobile driver trips per day to destination i and to home, $N_i$ and $N_h$ respectively, therefore can be obtained by integration:

$$N_i = \int p(i|t)\ dt \quad (34)$$

$$N_h = \int p(h|t)\ dt \quad (35)$$

where the integrals are over a day. Define

$$\exp L^* = \int [\exp L(t)]\ dt \quad (36)$$

$$\exp K^* = \int [\exp K(t)]\ dt \quad (37)$$

Then

$$N_i = \exp[F_i + L^* + G(0,s)]\ [1 + N\ \partial G(0,s)/\partial N_o] \quad (38)$$

$$N_h = NJ(O,s) \exp K^* \quad (39)$$

The terms $\exp L^*$ and $\exp K^*$ respectively can be absorbed into $G(0, s)$ and $J(0, s)$ to yield

$$N_i = \exp[F_i + G(0,s)]\ [1 + N\ \partial G(0,s)/\partial N_o] \quad (40)$$

$$N_h = NJ(0,s) \quad (41)$$

Note that

$$N = \sum_i N_i \quad (42)$$

This, together with Equations 40 and 41, implies the following system of equations:

$$N_i/N = \exp(F_i) / \sum_k \exp(F_k) \quad (43)$$

$$N_i = \exp[F_i + G(0,s)] \Big/ \Big\{1 - [\partial G(0,s)/\partial N_o] \sum_k \exp[F_k + G(0,s)]\Big\} \quad (44)$$

$$N = \sum_i \exp[F_i + G(0,s)] \Big/ \Big\{1 - [\partial G(0,s)/\partial N_o] \sum_k \exp[F_k + G(0,s)]\Big\} \quad (45)$$

$$N_h = NJ(0,s) \quad (46)$$

Equation 43 gives the probability that a nonwork automobile driver trip goes to destination i if a nonwork automobile driver trip is made. Equation 44 gives the average number of automobile driver trips per day to destination i. Equation 45 gives the average number of nonwork automobile driver trips per day to all nonwork destinations, and Equation 46 gives the average number of automobile driver trips per day from nonwork destinations to home. However, the average number of trips per day to a nonwork destination i equals the average number of sojourns per day at that destination, and the average number of trips from nonwork destinations to home equals the average number of tours from home to nonwork destinations to home. Thus, Equations 43-46 constitute a model of destination choice, sojourn frequency, and tour frequency for nonwork automobile driver travel.

Because increasing the number of trips begun at times other than t to t + $\Delta$t will tend to reduce the probability that an additional trip starts during t to t + $\Delta$t, owing to the hypothesized limitations on the travel resources available to households, it is expected that

$$\partial G(0,s)/(\partial N_o) < 0 \quad (47)$$

Thus, the denominators of Equations 44 and 45 are positive. Similarly, increasing the number of trips begun at times other than t to t + $\Delta$t will tend to increase the probability that a trip to home begins during t to t + $\Delta$t. Hence, $J(0, s)$ and $N_h$ are both positive.

Equation 43 is the disaggregate analog of a conventional trip distribution model. Equations 44 and 45 are analogous to trip generation models. However, they differ from conventional trip generation models in two important ways. First, they incorporate the concept that current travel decisions depend both on past travel decisions and future travel plans. This is a consequence of the hypothesized limitations on households' travel resources. Second, the sojourn frequencies implied by Equations 44 and 45 are sensitive to both transportation levels of service and measures of the attractiveness of potential nonwork destinations.

Equation 46 has no analog in conventional models. This is because the concept of a tour is not present in conventional models.

Equations 43-45, combined, imply that increasing the attractiveness of travel to a particular destination, other things being equal, causes average daily sojourns at that destination to increase. The increased sojourns are due, in part, to new travel and, in part, to travel diverted from other destinations. As a consequence of the new travel, the average values of total sojourns per day and tours per day both increase.

Equation 46 implies that the average number of sojourns per tour is independent of transportation level-of-service variables. This is a consequence of including only the first-order term in the Equation 28 approximation for $\exp H(N_o, s)$. If higher order terms were included in the approximation, then average sojourns per tour would not be independent of transportation level of service. However, the approximation that average sojourns per tour and transportation level of service are independent does not appear to be severe. Several studies have indicated that average sojourns per tour and level of service are either independent or only weakly related (8, 9).

## EMPIRICAL MODEL

To illustrate and test the theoretical framework, an empirical model of nonwork travel demand based on Equations 43-46 has been estimated. The data used to estimate the model were obtained from the 1968 Washington, D.C., area transportation survey. The estimation data set consisted of 890 households located in 13 traffic districts around the Washington area. The 890 households were selected randomly from the surveyed households in the 13 districts. Each household was assigned a set of destination choice alternatives consisting of the traffic zones visited for nonwork purposes by members of the households in its traffic district. The sets of destination choice alternatives thus constructed contained an average of 25 destinations per household.

Three types of exogenous variables are included in the empirical model: household characteristics, destination attraction variables, and travel times and costs. The variables specifying household characteristics are household size, number of cars owned, and income. The destination attraction variables are retail employment, service employment, and population, all according to traffic zone. The first of these variables is used to characterize attraction for shopping trips; the other two characterize attraction for nonshopping trips. The choice of attraction variables was strongly influenced by the contents of the Washington survey, zone level employment and population data being the only information pertaining to attraction contained in the survey.

The travel time and cost variables used in the model are the travel times and automobile operating costs associated with trips from home to the nonwork destinations available to households. The automobile operating cost variable excludes parking costs, as the trips in the estimation data set had free parking. Variables characterizing the travel times and costs associated with trips between nonwork destinations are not included. Although these travel times and costs presumably influence travel decisions involving multidestination tours and can be accommodated readily within the theoretical framework, their inclusion in the empirical model would have severely complicated the statistical estimation of the model. Statistical tests described later in this paper suggest that the damage done to the model by the omission of non-home-based travel times and costs may not be serious.

The functions F, G, and J in Equations 43-46 were specified as linear-in-parameters forms involving either the exogenous variables or simple transformations of the exogenous variables. The function F, which determines destination choice, was specified as

$$F_i = \sum_j c_j x_{ji} \quad (48)$$

where the c's are coefficients to be estimated and the x's represent travel time, travel costs, and destination attraction variables for a destination i. The functions G and J of Equations 44-46 were specified as follows:

$$G(N_o,s) = a_o + \sum a_i s_i - N_o(a_o' + \sum a_i' s_i) \quad (49)$$

$$J(0,s) = b_o + \sum b_i s_i \quad (50)$$

where the a's and b's are coefficients to be estimated, $N_o$ is defined as in Equation 22, and the $s_t$ represents the household characteristics variables.

Using Equations 49 and 50 and defining

$$W = \sum_i \exp(F_i) \tag{51}$$

Equations 45 and 46 respectively can be rewritten as

$$N = W[\exp(a_o + \Sigma\, a_i s_i)]/[1 + W(a_o' + \Sigma a_i' s_i) \exp(a_o + \Sigma\, a_i s_i)] \tag{52}$$

$$N_h = N(b_o + \Sigma\, b_i s_i) \tag{53}$$

Equations 52 and 53 are the specifications of the sojourn and tour frequency equations that were used in estimating the model.

The model was estimated in three stages. First, the coefficients of the function F were estimated by using Equation 43 and the method of maximum likelihood. The estimated coefficients were substituted into Equation 51, and W was computed for each household in the estimation data set. The coefficients of Equation 52 then were estimated using nonlinear regression. The resulting coefficients were substituted back into Equation 52, and N was computed for each household in the data set. Finally, the coefficients of Equation 53 were estimated by using ordinary least squares.

The estimation results for the destination choice model are shown in Table 1. All coefficients have the expected signs. The travel time, retail employment, and population coefficients are statistically significantly different from zero at the 1 percent level. The coefficient of automobile operating cost is not significantly different from zero. Moreover, a 25 percent change in automobile operating cost has roughly the same effect as a 1 percent change in travel time when the time and cost variables have their mean values. Thus, nonwork travelers appear to be relatively insensitive to variations in automobile operating costs at the cost levels represented in the estimation data set (roughly \$0.10 to \$0.30 per trip). Insensitivity of travelers to automobile operating costs has been reported previously for both nonwork and work travel (10, 11).

Estimation results for the sojourn and tour frequency models (Equations 52 and 53) are shown in Table 2. Because Equation 52 is highly nonlinear in parameters in the vicinity of the estimated values of the coefficients, it is not possible to develop useful estimates of the standard deviations of the coefficients or to perform t-tests of hypotheses concerning the coefficients (12). However, asymptotic tests of hypotheses concerning the coefficients can be performed using a chi-square test based upon ratios of error sums of squares. Let $S(\Theta^H)$ be the error sum of squares when the coefficients have their hypothesized values, $S(\Theta)$ be the unconstrained nonlinear regression estimates of the coefficients, n be the sample size, and p be the number of constrained parameters in $\Theta^H$. Then $n\ln[S(\Theta^H)/S(\Theta)]$ is asymptotically distributed as chi-square with p degrees of freedom. Using this test, all of the coefficients of the sojourn frequency model were found to be statistically significantly different from zero at the 1 percent level.

The statistical significance of the a′ coefficients in Equation 52 is particularly noteworthy. These coefficients result from the hypothesis that households have limited travel resources and, therefore, must consider both past travel decisions and future travel plans when making current travel decisions. The finding that the a′ coefficients are statistically significantly different from zero tends to confirm the hypothesis.

The positive coefficient of household size in Equation 52 indicates that increases in household size lead to increases in average sojourn frequency. The negative value of the a′ coefficient of automobile ownership indicates that increases in automobile ownership increase households' travel resources. The signs of the two automobile ownership coefficients are such that increases in automobile ownership lead to increases in sojourn frequency. The negative value of the a′ coefficient for income indicates that high-income households have greater travel resources than low-income ones. This may indicate that high-income households are better able to afford the activities that generate nonwork travel or that high-income households have more time for nonwork travel than low-income households. However, because the a coefficient for income is also negative, increases in income do not necessarily lead to increases in sojourn frequency. Rather, the sign of the effect of income on sojourn frequency depends on household characteristics. Increases in income tend to decrease sojourn frequency for households of sizes one and two and tend to increase sojourn frequency for other households.

**Table 1. Coefficients for destination choice model.**

| Variable | Coefficient | t-Statistic |
|---|---|---|
| Cost in cents divided by household income indicator[a] | -0.016 13 | -0.6272 |
| Logarithm of home-to-destination travel time in minutes | -1.207 8 | -7.649 |
| Logarithm of retail employment in destination zone | 0.077 36 | 3.027 |
| Population in destination zone | $2.801 \times 10^{-5}$ | 2.395 |
| Service employment in destination zone | $7.738 \times 10^{-5}$ | 1.042 |

Note: $\rho^2 = 0.066$, 510 observations, and 12 985 alternatives.

[a] The household income data consisted of the following indicators of annual income (1968 dollars): (1) 0-2999, (2) 3000-3999, (3) 4000-5999, (4) 6000-7999, (5) 8000-9999, (6) 10 000-11 999, (7) 12 000-14 999, (8) 15 000-19 999, (9) 20 000-24 999, (10) 25 000 or more.

**Table 2. Coefficients for sojourn ($R^2 = 0.0061$) and tour ($R^2 = 0.078$) frequency models.**

| Variable | Coefficient Type[a] | Sojourn Coefficient[b] | Tour Coefficient[b] | Tour t-Statistic |
|---|---|---|---|---|
| Constant | | | 0.7923 | 28.66 |
| Household size | a | 1.497 | | |
| Cars owned | a | 0.860 3 | | |
| Household income indicator[c] | a | -0.488 4 | | |
| Constant | $a_o'$ | 1.511 | | |
| Cars owned | a′ | -0.150 4 | | |
| Household income indicator[c] | a′ | -0.068 99 | | |

Note: 890 observations.

[a] Defined in Equation 52.

[b] Blanks are missing coefficients found not to be statistically significantly different from zero and therefore dropped from the specification.

[c] Defined in Table 1.

## ESTIMATING VEHICLE-KILOMETERS TRAVELED

Let $p_i$ be the probability that a nonwork trip goes to destination i, given that a nonwork trip is made (i.e., $p_i$ is the dependent variable in Equation 43), and let $d_i$ be the distance from home to nonwork destination i. Let D be the average distance from home to the nonwork destinations visited by a household. Then

$$D = \Sigma p_i d_i \tag{54}$$

An estimate of the average length, V, of a nonwork tour that includes S sojourns is

$$V = 2D + (S - 1)r \tag{55}$$

where r is the average distance between the destinations of multidestination tours. An estimate of a household's daily vehicle-kilometers traveled on nonwork tours (VKMT) can be formed by summing Equation 55 over all tours made by the household. If $N_h$ is the number of nonwork tours and N is the number of sojourns on these tours, then

$$VKMT = 2DN_h + r(N - N_h) \tag{56}$$

Because $p_i$, N, and $N_h$ are estimated in Equations 43, 52, and 53, Equation 56 provides an endogenous estimate of households' average daily vehicle-kilometers traveled on nonwork tours. The estimated form of this equation is

$$VKMT = 2DN_h + \underset{(9.256)}{6.061}\,(N - N_h) \qquad R^2 = 0.55 \tag{57}$$

The quantity in parentheses, 9.256, is the t-statistic.

## TESTS OF THE VALIDITY OF THE MODEL

If the empirical model and the theoretical framework upon which it is based are valid, then reestimation of the model using a data set different from the original estimation data set should produce coefficient estimates that are not statistically significantly different from the original estimates. Therefore, the model was reestimated using a data set consisting of 878 households that were selected randomly from 14 traffic districts in the Washington, D.C., area. These 14 districts were distinct from the districts represented in the original estimation data set. In the following discussion, the data set used in reestimating the model is referred to as DS-2, and the original estimation data set as DS-1.

Table 3 shows the DS-2 estimates of the coefficients of F in the destination choice Equation 43. Although the coefficients of automobile operating cost and service employment have signs that are contrary both to expectation and to the DS-1 results, none of the pairwise differences between DS-1 and DS-2 coefficients is statistically significant at the 10 percent level. Moreover, a chi-square test of the hypothesis that the complete vectors of DS-1 and DS-2 coefficients are equal accepted the hypothesis at the 10 percent significance level. These results support the validity of the destination choice model.

The DS-2 estimates of the coefficients of Equation 52 for sojourn frequency and Equation 53 for tour frequency are shown in Table 4. The values of W in Equation 52 were computed by using the DS-1 estimates of the coefficients of F. Similarly, the values of N in Equation 53 were computed by using the DS-1 estimates of the coefficients in Equation 52. These procedures were necessary to ensure that the independent variables W and N in Equations 52 and 53 respectively would have the same definitions in both the DS-1 and the DS-2 estimates of these equations, thus making comparisons of the DS-1 and DS-2 estimates of the coefficients in the two equations possible. A comparison of these estimates (Equation 52) is conditional on the correctness of Equation 43, and their comparison in Equation 53 is conditional on the correctness of Equation 52. Subject to these conditions, chi-square tests of the hypotheses that the vectors of DS-1 and DS-2 coefficients in Equation 53 are equal were performed. Both hypotheses were accepted at the 10 percent significance level. These results support the validity of the sojourn and tour frequency models.

The DS-2 estimate of Equation 56 for average daily nonwork vehicle-kilometers traveled is

$$VKMT = 2DN_h + \underset{(10.7)}{8.477}\,(N - N_h) \qquad R^2 = 0.46 \tag{58}$$

A test of the hypothesis that the DS-1 and DS-2 coefficients of $(N - N_h)$ are equal resulted in rejection of

**Table 3. Coefficients for destination choice model using DS-2 sample.**

| Variable | Coefficient | t-Statistic |
|---|---|---|
| Cost in cents divided by household income indicator[a] | 0.2870 | 1.490 |
| Logarithm of home-to-destination travel time in minutes | -1.488 | -10.68 |
| Logarithm of retail employment in destination zone | 0.1437 | 4.882 |
| Population in destination zone | $9.995 \times 10^{-6}$ | 0.736 8 |
| Service employment in destination zone | $-3.197 \times 10^{-6}$ | -0.054 82 |

Note: $\rho^2$ = 0.086, 578 observations, and 10 656 alternatives.

[a] Defined in Table 1.

**Table 4. Coefficients for sojourn ($R^2$ = 0.056) and tour ($R^2$ = 0.035) frequency models using DS-2 sample.**

| Variable | Coefficient Type[a] | Sojourn Coefficient | Tour Coefficient | Tour t-Statistic |
|---|---|---|---|---|
| Constant | | | 0.7923 | 26.96 |
| Household size | a | 0.724 5 | | |
| Cars owned | a | 1.302 | | |
| Household income indicator[b] | a | -0.370 8 | | |
| Constant | $a'_0$ | 1.039 | | |
| Cars owned | $a'$ | -0.065 88 | | |
| Household income indicator[b] | $a'$ | -0.037 04 | | |

Note: 878 observations.

[a] Defined in equation 52. [b] Defined in Table 1.

the hypothesis at the 10 percent significance level and acceptance at the 1 percent level. Thus the equality of the DS-1 and DS-2 coefficients and the statistical validity of the equation for vehicle-kilometers traveled are questionable, although they cannot be conclusively rejected. However, when Equations 52 and 53 are used to estimate N and $N_h$, the forecasts of average daily nonwork vehicle-kilometers traveled produced by the DS-1 and DS-2 versions of Equation 56 differ by less than 9 percent. Thus, the DS-1 and DS-2 versions of Equation 56 appear to be operationally equivalent, despite the possibility of their not being statistically equivalent.

## CONCLUSIONS

Nonwork travel patterns that include several visits to nonwork destinations during a tour or several nonwork tours during a day account for much of urban nonwork travel. These multidestination travel patterns are not treated well in current travel demand models. The modeling approach that has been presented here provides several encouraging, although not final steps toward the development of improved means of incorporating multidestination travel patterns into models of nonwork travel demand. A theoretical framework for dealing with nonwork travel that includes multidestination travel has been developed. An empirical model based upon the theoretical framework has been estimated. This model relates nonwork tour frequency, sojourn frequency, destination choice, and daily vehicle-kilometers traveled to variables describing household characteristics, destination characteristics, and transportation level of service. The empirical model has been found to perform well in several tests of its validity.

There is, of course, a variety of ways in which the model and its approach presented here could be further developed, for example through the addition of explanatory variables that characterize the travel times and costs associated with trips between nonwork destinations. These travel times and costs presumably influence travel decisions concerning multidestination tours. One way of incorporating them into the model would be to include in the utility function for travel to each destination a variable equal to the expected utility of travel from that destination to other destinations. The principal obstacle to doing this, and the reason for not doing it in the current empirical model, is that it would cause the function F in the destination choice Equation 43 to be nonlinear in parameters and, therefore, difficult to estimate. This obstacle might be overcome by developing a suitable linear-in-parameters approximation to the expected utility of travel between one nonwork destination and others.

Another way of improving the modeling approach would be to broaden the range of travel options it treats.

Much of this broadening could be accomplished within the current theoretical framework. For example, transit travel could be incorporated into the framework by redefining the set of available travel options so as to represent transit and automobile travel to a given destination as separate options with separate utility functions. A similar approach could be used to treat nonwork sojourns on tours that originate or terminate at work.

The approach also could be broadened by developing a means of estimating the demand for individual trip links. This might be done by constructing a matrix of probabilities of travel between specific origins and destinations. These travel probabilities could depend on variables characterizing households' daily travel patterns as well as on variables characterizing the specific origin-destination links, thereby avoiding the Markovian assumption that travelers' decisions depend only on the characteristics of individual trip links and not on the characteristics of tours or daily travel patterns.

## ACKNOWLEDGMENTS

I thank David Syskowski for his invaluable assistance in processing the data used in this study. The views expressed in this paper are mine and are not necessarily endorsed by the Environmental Protection Agency.

## REFERENCES

1. T. A. Domencich and D. McFadden. Urban Travel Demand. North Holland/American Elsevier, New York, 1975.
2. T. J. Adler and M. Ben Akiva. Joint-Choice Model for Frequency, Destination, and Travel Mode for Shopping Trips. TRB, Transportation Research Record 569, 1976, pp. 136-150.
3. Cambridge Systematics and Alan M. Voorhees and Associates. Carpool Incentives: Analysis of Transportation and Energy Impacts. Rept. prepared for the Federal Energy Administration, June 1976.
4. Urban Transportation Planning: General Information and Introduction to System 360. Federal Highway Administration, Mar. 1972.
5. T. Sasaki. Estimation of Person Trip Patterns Through Markov Chains. In Traffic Flow and Transportation (G. F. Newell, ed.), American Elsevier, New York, 1972.
6. G. Gilbert, G. L. Peterson, and J. L. Schofer. Markov Renewal Model of Linked Trip Travel Behavior. Transportation Engineering Journal, Vol. 98, Aug. 1972, pp. 691-704.
7. F. E. Horton and W. E. Wagner. A Markovian Analysis of Urban Travel Behavior: Pattern Response by Socioeconomic-Occupational Groups. HRB, Highway Research Record 283, 1974, pp. 19-29.
8. T. J. Adler. Modeling Non-Work Travel Patterns. Massachusetts Institute of Technology, Cambridge, PhD thesis, Sept. 1976.
9. J. R. Ginn. Transportation Considerations in an Individual's Activity System. Northwestern Univ., Evanston, IL, PhD thesis, Aug. 1969.
10. J. Horowitz. Effects of Travel Time and Cost on the Frequency and Structure of Automobile Travel. TRB, Transportation Research Record 592, 1976, pp. 1-5.
11. K. Train and D. McFadden. The Measurement of Urban Travel Demand II. Urban Travel Demand Forecasting Project, Univ. of California, Berkeley, Working Paper No. 7517, Sept. 1975.
Forecasting Project, Univ. of California, Berkeley, Working Paper No. 7517, Sept. 1975.
12. E. M. L. Beale. Confidence Regions in Nonlinear Estimation. Journal of the Royal Statistical Society, Series B, Vol. 22, 1960, pp. 41-76.

*Publication of this paper sponsored by Committee on Passenger Travel Demand Forecasting and Committee on Traveler Behavior and Values.*

# Modeling the Choice of Residential Location

Daniel McFadden, Department of Economics, Massachusetts Institute of Technology, Cambridge

**The problem of translating the theory of economic choice behavior into concrete models suitable for analyzing housing location is discussed. The analysis is based on the premise that the classical, economically rational consumer will choose a residential location by weighing the attributes of each available alternative and by selecting the alternative that maximizes utility. The assumption of independence in the commonly used multinomial logit model of choice is relaxed to permit a structure of perceived similarities among alternatives. In this analysis, choice is described by a multinomial logit model for aggregates of similar alternatives. Also discussed are methods for controlling the size of data collection and estimation tasks by sampling alternatives from the full set of alternatives.**

The classical, economically rational consumer will choose a residential location by weighing the attributes of each available alternative—accessibility to work place, shopping, and schools; quality of neighborhood life and availability of public services; costs, including price, taxes, and travel costs; and dwelling characteristics, such as age, number of rooms, type of appliances—and by choosing the alternative that maximizes utility.

This paper considers the problem of translating the theory of economic choice behavior into concrete models suitable for the empirical analysis of housing location. We are concerned particularly with two problems in the modeling of individual, or disaggregate, choice among residential locations. First, there may be a structure of perceived similarities among alternatives that invalidates the commonly used joint multinomial logit model of choice. We treat individual dwelling units as the basic alternatives among which choice is made. Each unit will have a list of attributes, observed and unobserved, to which the individual responds. We assume that the space of attributes, including unobserved attributes, is sufficiently rich so that each physical dwelling unit is represented by a unique point in attribute space. Of course, the individual may perceive two dwellings that are similar in some attributes as quite similar overall; it is the impact of such perceptions on choice that I wish to model.

I shall introduce a family of probabilistic choice models, of which the joint multinomial logit model is a special case, that has the property of aggregating dwelling units perceived as similar. The weight given to an aggregate of alternatives in the choice process will depend on the degree of perceived similarity.

At one extreme, the elements of the aggregate will be perceived as independent, and choice will be described by a multinomial logit model with individual dwellings as alternatives. At the other extreme, all dwellings with the same observed attributes will be perceived as virtually the same, and choice will be described by a multinomial logit model with dwelling types, which are distinguished by observed attributes, as the alternatives. The family of models introduced here permits empirical estimation of the degree of perceived similarity and tests of the two extreme cases mentioned above.

The second problem treated in this paper is that of estimation of individual choice models when the number of elemental alternatives is impractically large. The section on limiting the number of alternatives establishes that, if choice among a set of alternatives is described by a multinomial logit model, then the model can be estimated by sampling from the full set of alternatives, with appropriate adjustment in the estimation mechanism. Thus, estimation can be carried out with limited data collection and computation.

The solutions I give to the two problems above will be applied to empirical studies of housing location by Quigley (1) and Lerman (2).

## THEORY OF HOUSING LOCATION CHOICE

Assume the classical model of the rational, utility-maximizing consumer. Suppose the consumer faces a residential location decision, with a choice of communities indexed $c = 1, \ldots, C$ and dwellings indexed $n = 1, \ldots, N_c$ in community c. The consumer will have a utility $U_{cn}$ for alternative cn, which is a function of the attributes of this alternative, including accessibility, quality of public services, neighborhood and dwelling characteristics, etc., as well as a function of the consumer's characteristics, such as age, family size, and income. The consumer will choose the alternative that maximizes his utility.

Not all attributes of alternatives will be observed. The unobserved variables will have some probability distribution in the population, conditioned on the value of the observed variables. If the observer knows the form of the utility function and the probability distribution of unobserved variables, then probabilistic statements can be made about the expected distribution of choices:

$$P_{cn} = \mathrm{Prob}[U_{cn} > U_{bm} \text{ for } bm \neq cn] \tag{1}$$

where $P_{cn}$ denotes the probability of choice cn and the probability on the right side is defined with respect to the distribution of unobserved variables. The econometric approach to this problem is to specify, as a maintained hypothesis, a class of utility forms and distributions from which one member can be statistically identified.

Consider the decomposition $U_{cn} = V_{cn} + \epsilon_{cn}$ of utility into a term $V_{cn}$ that is a function specified up to a finite vector of unknown parameters, of observed variables, and a term $\epsilon_{cn}$ summarizing the contribution of unobserved variables. Hereafter, $V_{cn}$ will be called the strict utility of cn. Let $\underset{\sim}{\epsilon}$ denote the vector $(\epsilon_{11}, \ldots, \epsilon_{1N_1}, \ldots, \epsilon_{C1}, \ldots, \epsilon_{CN_C})$ and let $F(\underset{\sim}{\epsilon})$ denote the cumulative distribution function of $\underset{\sim}{\epsilon}$. Then Equation 1 can be written

$$P_{cn} = \int_{\epsilon_{cn}=-\infty}^{+\infty} F_{cn}(\langle V_{cn} + \epsilon_{cn} - V_{dm}\rangle)\,d\epsilon_{cn} \tag{2}$$

where $F_{cn}$ denotes the derivative of F with respect to its cn argument, and $\langle V_{cn} + \epsilon_{cn} - V_{dm}\rangle$ denotes a vector with components indexed by dm. An econometric model of choice is specified by choosing a parametric form for

$V_{dn}$ and a parametric distribution F.

## MULTINOMIAL LOGIT MODEL

An empirically important specialization of Equation 2 is the multinomial logit model,

$$P_{cn} = \exp(V_{cn}) \Big/ \sum_{b=1}^{C} \sum_{m=1}^{N_b} \exp(V_{bm}) \tag{3}$$

obtained by assuming the $\epsilon_{dn}$ to be independently and identically distributed with the extreme value distribution,

$$\text{Prob}\,(\epsilon_{cn} < \epsilon) = \exp(-e^{-\epsilon}) \tag{4}$$

This model was proposed as a theory of psychological choice behavior by Luce (3). Its econometric analysis has been investigated by McFadden (4, 5) and Nerlove and Press (6). A particular structural feature of this model, termed independence from irrelevant alternatives by Luce, is that the relative odds for any two alternatives are independent of the attributes, or even the availability, of any other alternative. This property is extremely useful in simplifying econometric estimation and forecasting (7) but can be shown to be implausible for choice problems where it is unreasonable to assume that the $\epsilon_{dn}$ are statistically independent (8, 9).

For later analysis, it will be useful to rewrite the joint choice Equation 3 in terms of a conditional choice probability $P_{n|c}$ for dwelling, given community, and a marginal choice probability $P_c$ for community. The strict utility $V_{cn}$ can often be expressed in an additively separable, linear-in-parameters form

$$V_{cn} = \beta' x_{cn} + \alpha' y_c \tag{5}$$

where $x_{cn}$ is a vector of observed attributes that vary with both community and dwelling (e.g., work-place accessibility), $y_c$ is a vector of observed attributes that vary only with community (e.g., availability of community recreation facilities), and $\alpha$ and $\beta$ are vectors of unknown parameters. Hereafter, we assume the structure of Equation 5. From Equations 3 and 5, one obtains the formulas

$$P_{n|c} = \exp(V_{cn}) \Big/ \sum_{m=1}^{N_c} \exp(V_{cm}) = \exp(\beta' x_{cn}) \Big/ \sum_{m=1}^{N} \exp(\beta' x_{cm}) \tag{6}$$

$$P_c = \sum_{n=1}^{N_c} \exp(V_{cn}) \Big/ \sum_{b=1}^{C} \sum_{m=1}^{N_b} \exp(V_{bm})$$

$$= \exp(\alpha' y_c) \left[\sum_{n=1}^{N_c} \exp(\beta' x_{cn})\right] \Big/ \sum_{b=1}^{C} \exp(\alpha' y_b) \left[\sum_{m=1}^{N_b} \exp(\beta' x_{bm})\right] \tag{7}$$

Define an inclusive value

$$I_c = \log\left[\sum_{n=1}^{N_c} \exp(\beta' x_{cn})\right] \tag{8}$$

Then, Equations 6 and 7 can be rewritten

$$P_{n|c} = \exp(\beta' x_{cn})/\exp(I_c) \tag{9}$$

$$P_c = \exp(\alpha' y_c + I_c) \Big/ \sum_{b=1}^{C} \exp(\alpha' y_b + I_b) \tag{10}$$

One method of estimating the joint model (Equation 3) is to first estimate the parameters $\beta$ from the conditional choice model (Equation 6). Next define $I_c$ using the log of the denominator of the estimated equation. Finally, estimate the parameters $\alpha$ from the marginal probability model (Equation 10), given $I_c$. This sequential approach to estimation economizes on the number of alternatives and the number of parameters considered at each stage of estimation, with some loss of efficiency relative to direct estimation of the joint model (Equation 3).

## NESTED LOGIT MODEL

An empirical generalization of the multinomial logit model in the form of Equations 9 and 10 is obtained by allowing the inclusive value $I_c$ in the latter to have a coefficient other than one:

$$P_c = \exp[\alpha' y_c + (1 - \sigma) I_c] \Big/ \sum_{b=1}^{C} \exp[\alpha' y_b + (1 - \sigma) I_b] \tag{11}$$

where $(1 - \sigma)$ is a parameter. The model represented by Equations 9 and 11, termed the "nested logit model," was first used with the estimation procedure described above, but with an unsatisfactory definition of inclusive value (9). Ben-Akiva has suggested the correct definition (Equation 8) of inclusive value and explored the implications of fitting the joint model or various nested models. Amemiya (10) corrects an error in the formula used in the earlier studies to compute the standard errors of estimates in the last stage of the sequential estimation procedure [see also McFadden (11)].

## GENERALIZED EXTREME VALUE MODEL

I shall now introduce a family of choice models, derived from stochastic utility maximization, that includes multinomial and nested logit. This family allows a general pattern of dependence among the unobserved attributes of alternatives and yields an analytically tractable closed form for the choice probabilities. The following result characterizes the family.

Suppose $G(y_1, \ldots, y_J)$ is a nonnegative, homogeneous-of-degree-one function of $(y_1, \ldots, y_J) \geq 0$. Suppose $G \rightarrow \infty$ if $y_i \rightarrow \infty$ for each i, and for k distinct components $i_1, \ldots, i_k$, $\partial^k G/\partial y_1 \ldots y_{i_k}$ is nonnegative if k is odd and nonpositive if k is even. Then

$$P_i = \{\exp(V_i) G_i[\exp(V_1), \ldots, \exp(V_J)]\}/G[\exp(V_1), \ldots, \exp(V_J)] \tag{12}$$

defines a probabilistic choice model from alternatives $i = 1, \ldots, J$, which is consistent with utility maximization. Further, expected maximum utility, defined by

$$\bar{U} = \int_{\underset{\sim}{\epsilon}=-\infty}^{+\infty} \max_i (V_i + \epsilon_i) f(\underset{\sim}{\epsilon})\, d\underset{\sim}{\epsilon} \tag{13}$$

(with f the density for F), satisfies

$$\bar{U} = \log G[\exp(V_1), \ldots, \exp(V_J)] + \gamma \tag{14}$$

where $\gamma = 0.57721$ is Euler's constant, and

$$P_i = \partial \bar{U}/\partial V_i \tag{15}$$

I have proved this result (11).

The special case $G(y_1, \ldots, y_J) = \sum_{j=1}^{J} y_j$ yields the multinomial logit model. An example of a more general

G function satisfying the hypotheses of the theorem is

$$G(y) = \sum_{m=1}^{M} a_m \left[ \sum_{i \in B_m} y_i^{1/(1-\sigma_m)} \right]^{1-\sigma_m} \tag{16}$$

where $B_m \subset \{1, \ldots, J\}$, $\bigcup_{m=1}^{M} B_m = \{1, \ldots, J\}$, $a_m > 0$, and $0 \leq \sigma_m < 1$.

For the bivariate case with a single class m, Equation 16 reduces to

$$G(y) = \left[ y_1^{1/(1-\sigma)} + y_2^{1/(1-\sigma)} \right]^{1-\sigma} \tag{17}$$

The bivariate extreme value distribution based on this form has been studied by Oliveira (12, 13), who shows that $\sigma$ is the product-moment correlation between the two variates. In the general case of Equation 16, $\sigma_m$ can be interpreted as an index of the similarity of the unobserved attributes $B_m$. However, the relation between the $\sigma_m$ and product-moment correlations between the alternatives is more complex.

The choice probabilities for Equation 16 satisfy

$$P_i = \sum_{m=1}^{M} P(i \mid B_m) P(B_m) \tag{18}$$

where

$$P(i \mid B_m) = \exp[V_i/(1-\sigma_m)] \Big/ \sum_{j \in B_m} \exp[V_j/(1-\sigma_m)] \tag{19}$$

if $i \in B_m$, and

$$P(i \mid B_m) = 0 \tag{20}$$

if $i \notin B_m$, with $P(i \mid B_m)$ denoting the conditional probability, and

$$P(B_m) = a_m \left\{ \sum_{j \in B_m} \exp[V_j/(1-\sigma_m)] \right\}^{1-\sigma_m} \div \sum_{n=1}^{M} a_n \left\{ \sum_{k \in B_n} \exp[V_k/(1-\sigma_n)] \right\}^{1-\sigma_n} \tag{21}$$

Choice probabilities of the form of Equation 18 were apparently first derived, for the case of three alternatives and $B_1 = \{1\}$, $B_2 = \{2, 3\}$, by Scott Cardell. For the case of disjoint $B_m$, the form of Equation 18 was treated independently by Daly and Zachary (14), Williams (15), and Ben-Akiva and Lerman (16). The demonstration by Daly and Zachary that Equation 18 is consistent with random utility maximization is noteworthy in that it permits generalization of the generalized extreme value model and provides a powerful tool for testing the consistency of choice models.

Consider an example of Equation 16,

$$G(y_1, y_2, y_3) = y_1 + \left[ y_2^{1/(1-\sigma)} + y_3^{1/(1-\sigma)} \right]^{1-\sigma} \tag{22}$$

where alternative 1 represents a dwelling in one community, and alternatives 2 and 3 represent dwellings of a similar type in a second community. Let $V_i$ be the strict utility of alternative i. The choice probabilities when the three alternatives are offered are, from Equation 18,

$$P(1 \mid 1,2,3) = \exp(V_1) \Big/ \left( \{\exp(V_1) + \exp[V_2/(1-\sigma)] + \exp[V_3/(1-\sigma)]\}^{1-\sigma} \right) \tag{23}$$

$$P(2 \mid 1,2,3) = \exp[V_2/(1-\sigma)]\{\exp[V_2/(1-\sigma)] + \exp[V_3/(1-\sigma)]\}^{-\sigma} \div \left( \exp(V_1) + \{\exp[V_2/(1-\sigma)] + \exp[V_3/(1-\sigma)]\}^{1-\sigma} \right) \tag{24}$$

where $P(i \mid A)$ denotes the probability that i is chosen from the alternatives A. If only alternatives 1 and 2 are available, then the choice probability (obtained from Equation 23 by setting $V_3 = -\infty$) has the binomial form

$$P(1 \mid 1,2) = \exp(V_1)/[\exp(V_1) + \exp(V_2)] \tag{25}$$

If only alternatives 2 and 3 are available, the choice probability again has a binomial logit form,

$$P(2 \mid 2,3) = \exp[V_2/(1-\sigma)] / \{\exp[V_2/(1-\sigma)] + \exp[V_3/(1-\sigma)]\} \tag{26}$$

Examining the choice probabilities of Equations 23 and 24 when all three alternatives are available, the value $\sigma = 0$ gives multinomial logit probabilities, while the limiting value $\sigma \to 1$ gives the probabilities

$$P(1 \mid 1,2,3) = \exp(V_1)/\{\exp(V_1) + \max[\exp(V_2), \exp(V_3)]\} \tag{27}$$

$$\begin{aligned} P(2 \mid 1,2,3) &= \exp(V_2)/[\exp(V_2) + \exp(V_3)] \quad \text{if } V_2 > V_3 \\ &= \tfrac{1}{2}\exp(V_2)/[\exp(V_2) + \exp(V_3)] \quad \text{if } V_2 = V_3 \\ &= 0 \quad \text{if } V_2 < V_3 \end{aligned} \tag{28}$$

In this extreme case, the consumer will treat two alternatives with identical strict utilities $V_2 = V_3$ as a single alternative in comparisons with alternative 1.

## RELATION BETWEEN THE NESTED LOGIT AND THE GENERALIZED EXTREME VALUE MODEL

The choice probabilities in Equation 18 can be specialized to the nested logit model given by Equations 9 and 11, as we shall now show. This result establishes that nested logit models are consistent with stochastic utility maximization and that the coefficient of inclusive value provides an estimate of the similarity of the unobserved terms in the first level of the nested model. Hence, it is possible to estimate some generalized extreme value choice models using nested logit models and inclusive values. Further, the generalized extreme value choice models provide a generalization of nested logit models and could be estimated directly to test for the presence and form of a nested (or tree) structure for similarities.

To obtain the nested logit model Equations 9 and 11 from Equation 18: replace the alternative index i with the double index cn for community c and dwelling n; replace m by c; assume the sets $B_c$ have the form $B_c = \{c1, \ldots, cN_c\}$; and assume the similarity coefficients have a common value $\sigma$. Then Equation 18 becomes

$$P_{cn} = \exp[V_{cn}/(1-\sigma)] \left\{ \sum_{m=1}^{N_c} \exp[V_{cm}/(1-\sigma)] \right\}^{-\sigma} \div \left( \sum_{b=1}^{C} \left\{ \sum_{m=1}^{N_b} \exp[V_{bm}/(1-\sigma)] \right\}^{1-\sigma} \right) \tag{29}$$

implying that

$$P_c = \sum_{n=1}^{N_c} P_{cn} = \left\{ \sum_{m=1}^{N_c} \exp[V_{cm}/(1-\sigma)] \right\}^{1-\sigma} \div \left( \sum_{b=1}^{C} \left\{ \sum_{m=1}^{N_b} \exp[V_{bm}/(1-\sigma)] \right\}^{1-\sigma} \right) \tag{30}$$

and that

$$P_{n|c} = P_{cn}/P_c = \exp[V_{cn}/(1-\sigma)] \Big/ \sum_{m=1}^{N_c} \exp[V_{cm}/(1-\sigma)] \quad (31)$$

Recalling that $V_{cn} = \beta' x_{cn} + \alpha' y_c$, these formulas can be written

$$P_c = \exp[\alpha' y_c + (1-\sigma) I_c] \Big/ \sum_{b=1}^{C} \exp[\alpha' y_b + (1-\sigma) I_b] \quad (32)$$

$$P_{n|c} = \exp[\beta' x_{cn}/(1-\sigma)] \Big/ \sum_{m=1}^{N_c} \exp[\beta' x_{cm}/(1-\sigma)] = \exp[\beta' x_{cn}/(1-\sigma)]/\exp(I_c) \quad (33)$$

$$I_c = \log \sum_{m=1}^{N_c} \exp[\beta' x_{cm}/(1-\sigma)] \quad (34)$$

Hence, the nested logit model is a specialization of the generalized extreme value model, with the coefficient $1-\sigma$ of inclusive value an index of the degree of independence of random terms for alternative dwellings in the same community.

This argument can be extended to trees of any depth. A sufficient condition for a nested logit model to be consistent with stochastic utility maximization is that the coefficient of each inclusive value lie in the unit interval.

## LIMITING THE NUMBER OF ALTERNATIVES CONSIDERED

Consider application of the joint multinomial logit model Equation 3 to the demand for housing, with alternatives indexed by community and by dwelling within the community. Ideally, the functional form of the model is appropriate for describing choice among the full set of alternatives available to consumers, and it is practical in terms of data collection and statistical analysis to study decision behavior at this level.

In practice, the number of available alternatives at the most disaggregate level often imposes infeasible data-processing requirements and strains the plausibility of the independence from irrelevant alternatives property of the multinomial logit functional form, as in the example of similar dwellings in the same community that are likely to have similar unobserved attributes.

Consider first the problem where enumeration of all alternatives is impractical but where data on selected disaggregate alternatives can be observed. If the multinomial logit functional form is valid, we shall establish the result that consistent estimates of the parameters of the strict utility function can be obtained from a fixed or random sample of alternatives from the full choice set.

Let C denote the full choice set. We shall assume it does not vary over the sample; however, this is inessential and can easily be generalized. Let $P(i|C, z, \theta^*)$ denote the true selection probabilities, where $\theta$ is a vector of parameters, and z is a vector of explanatory variables. We assume the choice probabilities satisfy the independence from irrelevant alternatives assumption:

$$i \in D \subseteq C \rightarrow P(i|C,z,\theta) = P(i|D,z,\theta) \sum_{j \in D} P(j|C,z,\theta) \quad (35)$$

which characterizes the multinomial logit model.

Now suppose for each case that a subset D is drawn from the set C according to a probability distribution $\pi(D|i, z)$, which may but need not be conditioned on the observed choice i. The observed choice may be either in or out of the set D. Examples of $\pi$ distributions are (a) choose a fixed subset D of C independent of the observed choice, (b) choose a random subset D of C containing the observed choice, and (c) choose a subset D of C consisting of the observed choice i and one or more other alternatives, selected randomly.

We give two examples of distributions of type (c):

1. (c-1): Suppose D is always selected to be a two-element set containing i and one other alternative selected at random. If J is the number of alternatives in C, then

$$\pi(D|i,z) = 1/(J-1) \qquad \text{if } D = [i,j] \text{ and } j \neq i \quad (36)$$

or zero otherwise.

2. (c-2): Suppose C is partitioned into sets $\{C_1, \ldots, C_M\}$, with $J_m$ elements in $C_m$, and suppose D is formed by choosing i (from the partition set $C_n$) and one randomly selected alternative from each remaining partition set. Then

$$\pi(D|i,z) = J_n / \prod_{m=1}^{M} J_m \qquad \text{if } i \in D,\ M = \#(D) \quad (37)$$

and $D \cap C_m \neq \phi$ for m = 1, ..., M, or zero otherwise.

The $\pi$ distributions of the types (a), (b), and (c-1) and (c-2) all satisfy the following basic property, which guarantees that, if an alternative j appears in an assigned set D, then it has the logical possibility of being an observed choice from the set D, in the sense that the assignment mechanism could assign the set D if a choice of j is observed.

### Positive Conditioning Property

If $j \in D \subseteq C$ and $\pi(D|i, z) > 0$, then $\pi(D|j, z) > 0$.

The $\pi$ distributions (a), (b), and (c-1) but not (c-2) satisfy a stronger condition.

### Uniform Conditioning Property

If $i, j \in D \subset C$, then $\pi(D|i, z) = \pi(D|j, z)$.

Consider a sample n = 1, ..., N, with the alternative chosen on case n denoted $i_n$, and $D_n$ denoting the choice set assigned to this case from the distribution $\pi(D|i_n, z_n)$. Observations with an observed choice not in the assigned set of alternatives are assumed to be excluded from the sample. Write the multinomial logit model in the form

$$P(i|C,z,\theta) = \exp[V_i(z,\theta)] \Big/ \sum_{j \in C} \exp[V_j(z,\theta)] \quad (38)$$

where $V_i(z, \theta)$ is the strict utility of alternative i.

If $\pi(D|i, z)$ satisfies the positive conditioning property and the choice model is multinomial logit, then maximization of the modified likelihood function

$$L_N = (1/N) \sum_{n=1}^{N} \log \left\{ \exp[V_{i_n}(z_n,\theta) + \log \pi(D_n|i_n,z_n)] \div \sum_{j \in D} \exp[V_j(z_n,\theta) + \log \pi(D_n|j,z_n)] \right\} \quad (39)$$

yields, under normal regularity conditions, consistent estimates of $\theta^*$. When $\pi(D|i, z)$ satisfies the uniform conditioning property, then Equation 39 reduces to the standard likelihood function,

$$L_N = (1/N) \sum_{n=1}^{N} \log \left\{ \exp[(V_{i_n}(z,\theta)] \Big/ \sum_{j\epsilon D} \exp[V_j(z_n,\theta)] \right\} \quad (40)$$

A proof is given by McFadden (17).

In conclusion, analysis of housing location can be carried out with a limited number of alternatives, which facilitates data collection and processing, provided the choice process is described by the multinomial logit model. If a mechanism such as (c-2) is used to select alternatives, the likelihood function should be modified to the form of Equation 39 to obtain consistent estimates of all parameters. If a non-modified likelihood function is used, estimation can still be carried out satisfactorily provided the effect of the selection mechanism for alternatives is absorbed by class-specific parameters. Caution is required in this case in verifying that the configuration of class-specific variables in the model is adequate to accommodate the selection mechanism effects, and in interpreting the estimates of class-specific parameters.

## AGGREGATION OF ALTERNATIVES AND THE TREATMENT OF SIMILARITIES

The preceding section has shown that, when the multinomial logit functional form is valid, estimation can be carried out by using randomly selected "representative" alternatives from each "class" of elemental alternatives, where the classes are defined by the analyst. Community and dwelling type were classification criteria mentioned in the earlier examples. Analysis of choice among classes by identifying them with "representative" members can be viewed as a method of aggregation of alternatives.

We shall now consider alternative methods of aggregation that can be employed when the multinomial logit form fails because of dependence between unobserved attributes of different alternatives within a class.

Again consider a consumer faced with a choice of housing locations in $c = 1, \ldots, C$ communities, with $n = 1, \ldots, N_c$ dwellings in community c, all of which have common unobserved community attributes. This introduces a dependence that conflicts with the assumptions of the joint multinomial logit model. To represent this dependence we shall assume that the choice probabilities have the nested logit structure of Equations 32-34, with $\sigma$ a measure of the degree to which dwellings within a class c are perceived as similar. When $\sigma = 0$, Equation 32 reduces to the multinomial logit model, and in the limit when $\sigma = 1$, it reduces to

$$P_c = \exp(\alpha' y_c + \max \beta' x_{cn}) \Big/ \sum_{b=1}^{C} \exp(\alpha' y_b + \max \beta' x_{bn}) \quad (41)$$

An analysis of housing demand by Quigley (1) using Pittsburgh data employs a model of the form of Equation 41. In Quigley's model, the nesting of community and housing type is reversed, with c denoting housing type, and n denoting specific dwelling, identified by community and location. Quigley assumes a sufficient structure on location choice so that the term $\max_n \beta' x_{cn}$ can be computed prior to parameter estimation. Then Equation 41 can be treated as an ordinary multinomial logit model.

In an analysis of neighborhood choice using Washington, D.C., data, Lerman (2) estimates a model of the form

$$P_c = \exp[\alpha' y_c + X_c^* + (1-\sigma)\log N_c] \div \sum_{b=1}^{C} \exp[\alpha' y_b + X_b^* + (1-\sigma)\log N_b] \quad (42)$$

where c indexes census tracts and $X_c^*$ is an "average" of the utility terms $\beta' x_{cn}$ of the dwellings in tract c. He notes that log $N_c$ is

> the measure of tract size required to correct for the fact that a census tract is actually a group of housing units. Other conditions being equal, a very large tract (i.e., one with a large number of housing units) would have a higher probability of being selected than a very small one, since the number of disaggregate opportunities is greater in the former than the latter. If all units of a particular type in a given zone are relatively homogeneous and the [joint multinomial] logit model applies to each individual unit, then the appropriate term to correct for tract size is the natural logarithm of the number of units [with] a coefficient of one.

Noting the model (Equation 41) as a second extreme case, Lerman concludes that "if the assumptions of the [joint multinomial] logit model are violated, the coefficient may differ from one." Lerman estimates the coefficient of log $N_c$ to be $1 - \sigma = 0.492$, with a standard error of 0.094. Hence, $\sigma$ satisfies the hypotheses of theorem 1 and is significantly different from both zero and one.

In the nested logit model (Equations 32 and 34), the inclusive value can be rewritten

$$I_c = [X_c^*/(1-\sigma)] + \log N_c + \log 1/N_c \sum_{m=1}^{N_c} \times \exp[(\beta' x_{cm} - X_c^*)/(1-\sigma)] \quad (43)$$

If a tract c is homogeneous in terms of observed variables so that $\beta' x_{cn} = X_c^*$, then the last term in Equation 43 vanishes, and the choice probability for the nested logit model (Equation 32) is exactly the Lerman model (Equation 42). This establishes the consistency of the Lerman model with stochastic utility maximization and supports his conclusion that the coefficient of log $N_c$ indexes the degree of independence of the alternatives within a tract. The same argument can be used to interpret Quigley's model, with $X_c^* = \max \beta' x_{cn}$.

When $X_c^*$ is the mean of $\beta' x_{cn}$, and not all $\beta' x_{cn} = X_c^*$, the convexity of the exponential implies

$$1/N_c \sum_{m=1}^{N_c} \exp[(\beta' x_{cm} - X_c^*)/(1-\sigma)] > 1 \quad (44)$$

and hence $I_c \geq [X_c^*/(1-\sigma)] + \log N_c$, with the difference of the two sides of the inequality depending on the variance of $\beta' x_{cn}$. One limiting case of Equation 43 that is of interest occurs when the number of dwellings within a tract is large and the $x_{cn}$ behave as if they were independently identically normally distributed with mean $X_c^*$. Let $\omega_c^2$ denote the variance of $\beta' x_{cn}$. If $N_c = r_c N$, with $r_c$ fixed and $N \to \infty$, then

$$P_c \to \frac{\exp\{[\alpha' y_c + \beta' x_c^* + (1-\sigma)\log r_c + \tfrac{1}{2}\omega_c^2]/(1-\sigma)\}}{\sum_{b=1}^{C} \exp\{[\alpha' y_b + \beta' x_b + (1-\sigma)\log r_b + \tfrac{1}{2}\omega_b^2]/(1-\sigma)\}} \quad (45)$$

When the disaggregate data $x_{cn}$ are not observed, but their distribution can be approximated or estimated, and $\omega_c$ is known, then Equation 45 can be used with standard multinomial logit estimation programs to provide estimates of $\sigma$ and $\beta$. If $r_o$ is unobserved, then it can be estimated when $\omega_o$ is known; when $y_o$ contains a tract-specific dummy variable, however, the tract-specific coefficient and $r_o$ are unidentified. This suggests one interpretation of tract-specific coefficients as indicating in part the number of equivalent disaggregate alternatives contained in the tract.

When $\omega_o$ is not known, but is known to have the structure $\omega_o^2 = \beta' \Omega_o \beta$, and the variables $x_{on}$ are multivariate normal with covariance matrix $\Omega_o$, direct estimation of $\beta$, $\sigma$, and $\alpha$ is possible. A modification of standard multinomial logit programs to handle nonlinear constraints on $\beta$ would be required for full maximum likelihood estimation. Alternately, consistent estimators could be obtained by writing out the terms in the quadratic form $\beta' \Omega_o \beta$ as independent parameters and ignoring constraints.

## CONCLUSION

This paper has considered the problem of modeling disaggregate choice of housing location when the number of disaggregate alternatives is impractically large and when the presence of a structure of similarities between alternatives invalidates the commonly used joint multinomial logit choice model. Theorems on sampling from the full set of alternatives and on generalizations of the multinomial logit model structure to accommodate similarities provide methods for circumventing these problems. Studies of housing demand by Quigley (1) and Lerman (2) motivate the analysis and illustrate its applicability.

## ACKNOWLEDGMENTS

This research was motivated by, and has benefitted from, discussions with Moshe Ben-Akiva, Steven Lerman, Charles Manski, and William Tye, and comments by Anthony E. Smith and Folke Snickars. I am indebted to the National Science Foundation for research support. This paper abstracts a more complete research report.

## REFERENCES


1. J. M. Quigley. Housing Demand in the Short-Run: An Analysis of Polytomous Choice. Explorations in Economic Research, Vol. 3, No. 1, Winter 1976, pp. 76-102.
2. S. R. Lerman. Location, Housing, Automobile Ownership, and Mode to Work: A Joint Choice Model. TRB, Transportation Research Record 610, 1977, pp. 6-11.
3. R. D. Luce. Individual Choice Behavior. Wiley, New York, 1959.
4. D. McFadden. Conditional Logit Analysis of Qualitative Choice Behavior. In Frontiers in Econometrics (P. Zarembka, ed.), Academic Press, New York, 1973.
5. D. McFadden. Quantal Choice Analysis: A Survey. Annals of Economic and Social Measurement, Vol. 5, No. 4, 1976, pp. 363-390.
6. M. Nerlove and J. Press. Univariate and Multivariate Log-Linear and Logistic Models. RAND, Rept. No. R-1306-EDA/NIH, 1973.
7. D. McFadden, W. Tye, and K. Train. Diagnostic Tests for the Independence From Irrelevant Alternatives Property of the Multinomial Logit Model. Paper presented at the 57th Annual Meeting, TRB, 1978.
8. G. Debreu. Review of R. Luce, Individual Choice Behavior. American Economic Review, Vol. 50, 1960, pp. 186-188.
9. T. Domencich and D. McFadden. Urban Travel Demand: A Behavioral Analysis. North-Holland, Amsterdam, 1975.
10. T. Amemiya. Specification and Estimation of a Multinomial Logit Model. Institute of Mathematical Studies in the Social Sciences, Stanford Univ., Stanford, CA, Technical Rept. No. 211, 1976.
11. D. McFadden. Econometric Models of Probabilistic Choice. In Econometric Analysis of Discrete Data (C. Manski and D. McFadden, eds.), MIT Press, Cambridge, MA, 1979.
12. J. T. de Oliveira. Extremal Distributions. Revista de Faculdada du Ciencia, Lisboa, Serie A, Vol. 7, 1958, pp. 215-227.
13. J. T. de Oliveira. La Representation des distributions extrémales bivariés. Bulletin of the Intional Statistical Institute, Vol. 33, 1961, pp. 477-480.
14. A. Daly and S. Zachary. Improved Multiple Choice Models. Planning and Transport Research and Computation (International), London, 1976.
15. H. C. L. Williams. On the Formation of Travel Demand Models and Economic Evaluation Measures of User Benefit. Environment and Planning, Vol. A.9, 1977, pp. 285-344.
16. M. Ben-Akiva and S. Lerman. Disaggregate Travel and Mobility Choice Models and Measures of Accessibility. Paper presented at the 3rd International Conference on Behavioral Travel Modeling, Tanenda, Australia, 1977.
17. D. McFadden. Modelling the Choice of Residential Location. In Spatial Interaction Theory and Planning Models (A. Karlqvist, L. Lundqvist, F. Snikars, and J. Weibull, eds.), North-Holland, Amsterdam, 1978.


*Publication of this paper sponsored by Committee on Passenger Travel Demand Forecasting and Committee on Traveler Behavior and Values.*

# Using Functional Measurement to Identify the Form of Utility Functions in Travel Demand Models

Steven R. Lerman,* Department of Civil Engineering, Massachusetts Institute of Technology, Cambridge
Jordan J. Louviere,* Departments of Marketing and Geography, University of Iowa, Ames

**Existing statistical procedures for model estimation that use data on observed behavior focus principally on the problem of estimating the parameters of the model, given the functional form. In contrast, methods for measuring function, which are psychological measurement procedures that use laboratory or interview data, provide a powerful set of tools for diagnosing the functional form of behavioral relationships. This paper explores the potential of a synthesis of these approaches in which functional measurement is used to guide travel demand model specification. A case study on choice of residential location by rural workers provides evidence that a model form based on functional measurement gives a better specification than the typical alternative functional forms used in travel demand models. Although it is relatively limited in scope, the case study strongly suggests that functional measurement methods can improve demand model specification.**

Estimating a travel demand model can be viewed as two interrelated problems: first, the development of a specification, or functional form, that describes the process of interest; second, the estimation of the parameters of that function. However, as Gaver and Geisel (1) point out, the existing literature on model estimation is oriented heavily toward the problem of estimating a set of parameters given a functional specification and offers only limited insight into how to select the appropriate specification in the first place. Available statistical approaches such as those suggested by Box and Cox (2) and those reviewed by Ramsey (3) rely either on statistical tests of goodness of fit or on a more general functional form that has other, simpler forms as restricted versions.

For example, Box and Cox suggest a transformation of the form

$$f(y) = \begin{cases} (y^\lambda - 1)/\lambda, \text{ if } \lambda \neq 0 \\ \log y, \text{ if } \lambda = 0 \end{cases} \qquad (1)$$

One can show that in the case of $\lambda = 0$, the expression $(y\lambda - 1)/\lambda$ converges to log y. Clearly, the case $\lambda = 1$ corresponds to a linear transformation. By estimating $\lambda$ (as well as the other parameters of the model), one can obtain insight into whether linear, logarithmic, or other transformations are appropriate and test whether the value of $\lambda$ is significantly different from any given value.

Unfortunately, these transformations usually complicate the computational problems of model estimation greatly, and it is generally infeasible to test the functional form of every single variable and all possible combinations of variables. Most travel demand models, because of the availability of relatively inexpensive model estimation procedures, have been restricted to specifications that have linear parameters, i.e., models of the form

$$\begin{aligned} f(x) &= x\beta \\ &= x_1\beta_1 + x_2\beta_2 + x_3\beta_3 + \ldots \end{aligned} \qquad (2)$$

This restriction is not particularly burdensome if one already knows that a particular nonlinear specification is appropriate, since, by judicious use of piece-wise linear forms and nonlinear transformations of the dependent and independent variables, one can approximate most nonlinear parameter functions fairly well. However, lacking guidance as to the appropriate functional form, and given, with existing techniques, the virtually infinite number of candidate transformations, choosing among specifications on goodness-of-fit criteria is far more likely to lead to one of the numerous incorrect specifications than to the correct one.

In contrast to the existing literature in econometrics, work in the areas of functional measurement, information integration theory, conjoint analysis, and direct utility assessment has been deeply concerned with questions of functional form. Studies that rely on these theories typically use laboratory-based experiments in which subjects are asked to make judgments about hypothetical alternatives. For example, subjects are confronted with a number of possible modes, each with an associated travel time, travel cost, etc. They are then asked to select a most preferred alternative, to rank the alternatives, or to associate some level of utility with each option.

Because a given subject can be asked to make a fairly large number of judgments in a single interview, the designers of these experiments can explore how individuals' responses are affected as a single independent variable changes while all other variables are held constant. This capability allows for a much more detailed assessment of the functional form of peoples' preferences, since, in an intuitive sense, the designer of the experiment can trace the shape of peoples' responses along each variable.

The fairly extensive experience with one of these psychological techniques, functional measurement, indicates that for any particular decision the functional form of peoples' preferences tends to be fairly stable across the population, even though the parameters of the function may vary widely (4, 5, 6, 7, 8). A key unresolved question, however, is whether the functional form derived from laboratory-based experiments is also a relevant model for actual decision making. If this is the case, then it would seem reasonable to develop a demand model estimation strategy that synthesizes the best features of both econometric estimation of revealed preferences (actual behavior) and function measurement (or some other, related technique) of laboratory or interview data. One could begin any demand model estimation by first performing a functional measurement experiment on a small sample; then, by using the resulting functional form as the starting point, one could estimate a demand model on data from real-world decisions.

This paper explores the potential of this two-phase demand model development technique in a small but fairly suggestive case study. The next section describes the general methodology of the case study, followed by a discussion of the theory of functional measurement and its application in the case study respectively. The next two sections are parallel to the second and third in that they describe theory and empirical results respectively, except that they examine models estimated on revealed preference data. Since most readers are likely to be far more familiar with travel demand modeling methods based on revealed preferences than functional measurement, the theory of the latter technique is developed in far greater detail than that of the former.

The conclusion is an evaluation of the implications of the study and discusses the potential of the two-phase model-building technique for improving travel demand models.

## METHOD OF ANALYSIS

In order to test the two-phase demand model formulation procedure, it was necessary to obtain two sets of data, one in which the behavior of interest was examined in a controlled experimental format and a second in which the corresponding real-world behavior was measured.

The data used represent the spatial choices of nonlocal workers in the rural West. The actual real-world behavior data are on nonlocal workers who take employment in the rural West and must select some town near the place of employment as a place of residence. These data were taken from surveys of power plant employees at six sites in Wyoming, North Dakota, and Montana. In all but one case, a complete record of each employee's residence was available; in the one exception, a 20 percent sample of employees was used. In total, approximately 9000-10 000 employees were included in the sample. The data used in the functional measurement experiment were based on surveys administered to a nonrepresentative sample of students, staff, and faculty at the University of Wyoming.

This case study examines a two-phase analysis in which (a) the functional form of a utility model is assessed in a controlled experiment, and (b) the parameters of this form are re-estimated by using data from actual decisions and are compared to more traditional linear parameter forms of logit choice models.

A major limitation of the study is that it deals with a relatively simple, two-variable model of behavior and a limited case study population. The results should be viewed, therefore, as a pilot test of a methodology rather than as a demonstration of a substantive result.

It is necessary to assume that individuals share a common form for their utility functions but that the parameters of those functions may vary widely. Empirical evidence available in psychological studies of judgment and decision making supports the existence of a common utility function for individual decision makers (4, 5, 6, 7, 8, 9, 10, 11, 12, 13). Thus, this assumption does not appear unreasonable. If such a function exists, it should be possible to infer (at least approximately) its functional form from relatively small, nonrepresentative samples. The parameters of these functions will differ, of course, from those of the population at large, but the form should be relatively invariant across subpopulations.

If the utility function derived from a laboratory study has any relevance to actual behavior, then a model estimated on data from actual decisions, or revealed preferences, should have the same functional form. In theory, if people behave as if they actually use the laboratory-derived functional form, then that specification should fit the actual data better than alternative specifications.

We shall demonstrate this to be approximately true in the case study to be detailed by showing that the functional measurement model is consistently superior, in terms of goodness of fit, to all functions usually assumed in existing choice models. A comparison of $R^2$-values adjusted for degrees of freedom and the standard errors of the models is made for alternative models.

While it would be better to compare alternative models based on how well they forecast behavior in an independent sample, data necessary for such a comparison were unavailable for this study.

## OVERVIEW OF FUNCTIONAL MEASUREMENT

As used in this context, the term "functional measurement" describes an approach to modeling individual behavior that is characterized by two aspects: (a) functional measurement based on an explicit theory of how people reach decisions and (b) use of laboratory experimental measurement methods to estimate models rather than observations on peoples' revealed preferences.

Functional measurement is based on theoretical and empirical research in mathematical psychology and related fields, where there is extensive support for the following assumptions.

$$x_{ki} = f_{ki}(X_{ki}) \tag{3}$$

$$U_i = g_i(x_{1i}, x_{2i}, \ldots, x_{Ki}) \tag{4}$$

$$b = h(U) \tag{5}$$

where

- $X_{ki}$ = physically measurable attributes of the alternative under study,
- $x_{ki}$ = values of $X_{ki}$ as perceived by individuals,
- $U_i$ = some level of response (such as numerical judgments, rankings, or choices) observed in an experimental context for alternative i (for the purpose of this paper, we shall refer to this response as utility),
- U = vector $U_1, \ldots, U_I$,
- B = an actual choice or behavior in a nonexperimental situation,
- I = number of available alternatives, and
- K = number of variables.

In many cases, $X_{ki}$ may include factors for which corresponding $x_{ki}$ are difficult to measure or not well understood. For example, automobile safety may affect a person's choice of auto type, but its physical referrents are not well known. Such factors are treated in our theory as distinct, qualitative variables that are part of $x_{ki}$.

As developed above, this theory allows for responses, perceptions, and behavior over any set of discrete alternatives, indexed as i = 1, ..., I. For example, one might be interested in mode-choice behavior, in which there are different factors influencing the desirability of driving alone, carpooling, taking transit, etc. In many situations, however, the behavior of interest is continuous and involves only one alternative. In these instances, the theory can often be reduced to the case I = 1, and the i subscript can be deleted. However, because in this case study we are concerned with peoples' choices among discrete alternatives, we will retain the full notation except as noted.

Each of these assumptions is restated more formally below, and the case of additive and multiplicative utilities is explored in detail.

Assumption 1

For any observed travel behavior there exists a set of independent factors that are functionally connected to its occurrence or the magnitude of its occurrence. Each factor may be either quantitative or qualitative in nature. We shall denote the set of J quantitative factors by $S_i = (S_{1i}, S_{2i}, \ldots, S_J)$ and the set of L quantitative factors by $Q_i = (Q_{1i}, Q_{2i}, \ldots, Q_L)$; J + L = K. The entire vector $X_i$ is simply $S_i$ and $Q_i$.

Assumption 2

Associated with each quantitative and qualitative factor is a corresponding value or quantity of its magnitude that may be obtained by one of several psychological measurement procedures. We shall let the utility of this quantity provided by one or a group of subjects be $(u_{1i}, u_{2i}, \ldots, u_{Ki})$. Because there may be K different values or corresponding utilities for each of the K factors, we may represent the utilities as $u_{ki}$. Formally, we postulate that

$$u_{ki} = f_{ki}(x_{ki}) \tag{6}$$

Assumption 3

In an experimental context we observe a response to a combination of $(S_{1i}, S_{2i}, \ldots, S_{Ji}, Q_{1i}, \ldots, Q_{Li})$ on a psychological measurement scale. We assume that this response measure is connected to the utility of the experimental factors according to some algebraic combination rule. If we agree to let $U_i$ represent the response to the i th alternative,

$$U_i = g_i(x_{1i}, x_{2i}, \ldots, x_{Ki}) \tag{7}$$

The vector of responses (U) is connected to the observed travel behavior by means of some algebraic function. Hence, if we agree to call the observed behavior B, then we can write

$$B = h(U) \tag{8}$$

Then by substitution

$$\begin{aligned} B &= h(U) \\ &= h[g(x)] \\ &= h\{g[f(S,Q)]\} \end{aligned} \tag{9}$$

This is too general a formulation for modeling purposes. In a practical application, one must make explicit assumptions about f, g, and h and deduce their consequences. The results lead to a general paradigm for the analysis of travel behavior that has growing empirical support (14, 15).

THEORY DEVELOPMENT

The critical component of this theory for the purposes of developing appropriate functional forms for travel demand models is the specification of $U_i = g_i(x_{1i}, x_{2i}, \ldots, x_{Ki})$. Analysis of variance provides a straightforward means of implementing the theory and diagnosing and/or testing alternative functional forms. In this study, we will consider both the linear and multiplicative cases.

There are two key conditions involved in the application of analysis of variance that must be satisfied. First, the pattern of the statistical significance (or nonsignificance) of the utility responses to various combinations of the independent variables must be of a specific nature so as to permit inference (diagnosis) or testing of model form. Second, corresponding graphical evidence must support the inference or test.

Consider the hypothesis that individuals in the experiment outlined above will trade off distance (or travel time) and town size (or amenities) independently of one another. That is, they will combine the effects of these two variables linearly. This hypothesis may be tested directly by an analysis of variance. If for clarity we suppress the subscript i and write

$$U_{mn} = U_m^1 + U_n^2 + \epsilon_{mn} \tag{10}$$

where

$U_m^1$ = utility values assigned to the m th level of the first factor (say, distance) in a factorial experimental plan,
$U_n^2$ = utility values assigned to the n th level of the second factor (say, town size),
$U_{mn}$ = overall utility assigned by individuals to combinations of levels of factors one and two, and
$\epsilon_{mn}$ = random error term with zero mean.

The test for independence of the two effects (distance and town size) corresponds to the test of the significance of the interaction effect of $U_m^1 \times U_n^2$. In an analysis of variance, this is a global test for any and all interaction effects between distance and town size. If the interaction is not significant (i.e., the hypothesis that $U_m^1$ and $U_n^2$ combine linearly cannot be rejected), then the linear form may be accepted. If the interaction is significant, it signals that some form other than a simple linear combination is appropriate.

This test is accompanied by a graphical plot of the interaction. If the hypothesis of linearity is correct, the data should plot as a series of parallel lines when plotted against either $U_m^1$ or $U_n^2$ values on the abscissa.

To see why, assume the linear form to be correct and consider the effect of subtracting level 1 from level 2 of the first factor. This yields

$$\begin{aligned} U_{2n} - U_{1n} &= (U_2^1 + U_n^2) - (U_1^1 + U_n^2) + (\epsilon_{2n} - \epsilon_{1n}) \\ &= U_2^1 - U_1^1 + (\epsilon_{2n} - \epsilon_{1n}) \end{aligned} \tag{11}$$

where $U_1^1$ and $U_2^1$ are the utility values assigned to levels 1 and 2 of factor one, respectively. Thus, the difference between the points when $U_n^2$ takes on any value is always a constant $U_2^1 - U_1^1$ (except for disturbances). Hence, the graph should yield a series of parallel lines.

Note that this is true regardless of the forms we assume for the marginal relationships [i.e., $U_m^1 = f_1(X_m)$ and $U_n^2 = f_2(Z_n)$]. It can be demonstrated that a measure of the average effect or utility (the so-called marginal utilities) of each of the two variables is given by their marginal means. We now demonstrate that this is true for any multilinear utility model, thereby confirming that it holds for any more restricted form such as simple addition or multiplication.

If the data were obtained from a factorial design in which factor one is the row factor (subscripted m) and factor two is the column factor (subscripted n), we may write the most general multilinear form as

$$U_{mn} = k_0 + k_1 U_m^1 + k_2 U_n^2 + k_3 U_m^1 \times U_n^2 + \epsilon_{mn} \tag{12}$$

where all terms are as defined previously and k are scaling constants. Additional factors simply add additional one-way, two-way, three-way, and higher terms. Now, if we average the factorial data over the second subscript n (i.e., the column factor), we would have

$$U_{m.} = k_0 + k_1 U_m^1 + k_2 \overline{U_n^2} + k_3 U_m^1 \times \overline{U_n^2} + \epsilon_{m.} \qquad (13)$$

where $\overline{U_n^2}$ is the average over-the-column factor. Thus, Equation 13 reduces to

$$U_{m.} = K_0 + K_1 U_m^1 + \epsilon_{m.} \qquad (14)$$

where K is collected terms.

Equation 14 demonstrates that the marginal row means (in general, the marginal means for any subscript) are equal to the marginal utilities up to a linear transformation. Hence, they are as good as any other estimate measured on an interval scale. Equation 14 is important because it demonstrates that an estimate of the marginal utility for any factor may be obtained by manipulating that factor as part of a factorial or fractional factorial design so long as any multilinear utility function can be assumed to have generated the data.

Returning to the reduced, strictly additive form, it may also be demonstrated that these marginal means relate to the overall utility value of cell m,n as follows (5, 6, 7, 8):

$$U_{mn} = U_{m.} + U_{.n} - U_{..} + \epsilon_{mn} \qquad (15)$$

where U.. is the grand average utility (mean). Similarly, for a strictly multiplicative form, it may be demonstrated that the following is true (5, 6, 7, 8).

$$U_{mn} = k + [(U_{m.} - k)(U_{.n} - k)/(U_{..} - k)] + \epsilon_{mn} \qquad (16)$$

where all terms are as defined in Equation 15 except k, which is a scaling constant that represents the arbitrary zero point on the utility scale.

Now, on the assumption that Equation 14 is true, we may write the following expressions by assigning levels of distance to the rows and levels of town size to the columns.

$$U_{m.} = f_1(\text{distance}_m) \qquad (17)$$

$$U_{.n} = f_2(\text{town size}_n) \qquad (18)$$

because the only source of variation in $U_{m.}$ and $U_{.n}$ is that due to the levels of distance and town size and error. Thus,

$$U_{mn} = f_1(\text{distance}_m) + f_2(\text{town size}_n) - U_{..} + \epsilon_{mn} \qquad (19)$$

if the two factors combine additively, or

$$U_{mn} = \{[f_1(\text{distance}_m) \times f_2(\text{town size}_n)]/U_{..}\} + \epsilon_{mn} \qquad (20)$$

if the factors combine multiplicatively and we assume that k in Equation 16 equals zero as a first hypothesis.

Following our previous logic, Equation 20 is testable statistically and graphically. In particular, Equation 20 requires that all interaction effects be statistically significantly different from zero and that the graph of the interaction must consist of a series of diverging curves. An exact statistical test may be obtained by using the marginal means as the independent values, estimating k (usually done by iterative methods) and performing the following linear regression.

$$\ln(U_{mn} - k) = \ln(U_{m.} - k) + \ln(U_{.n} - k) - \ln(U_{..} - k) \qquad (21)$$

If Equation 20 is true, the coefficients of the distance and town size terms should not be significantly different from 1.0.

Thus, we have demonstrated that an algebraic and statistical theory for diagnosing and testing any multilinear utility form does exist. In order to derive a model in the units of the original variables (e.g., miles, minutes, population), it is necessary to first diagnose the overall form of Equation 12 and then make assumptions about the functions in Equations 19 and 20 (or a more general form given by Equation 12, if appropriate).

In the next section we shall demonstrate the application of this theory and methodology to the problem of choosing a town in which to live, given that one has taken a job at a plant or a mine located in a rural area. We then demonstrate that knowledge of the functional form of the utility expression provides quite accurate recovery of real-world data in an analogous choice situation. We then compare the derived function form to a large number of linear parameter forms that might typically be fit in a logit analysis.

## RESULTS OF FUNCTIONAL MEASUREMENT EXPERIMENT FOR TOWN CHOICE

In order to develop a specification for a utility function for town choice, a functional measurement experiment designed to reflect the employment and residence situations for isolated plants in Wyoming, North Dakota, and Montana was developed. To maintain a fairly simple structure, only two variables, distance to work and size of town, were considered.

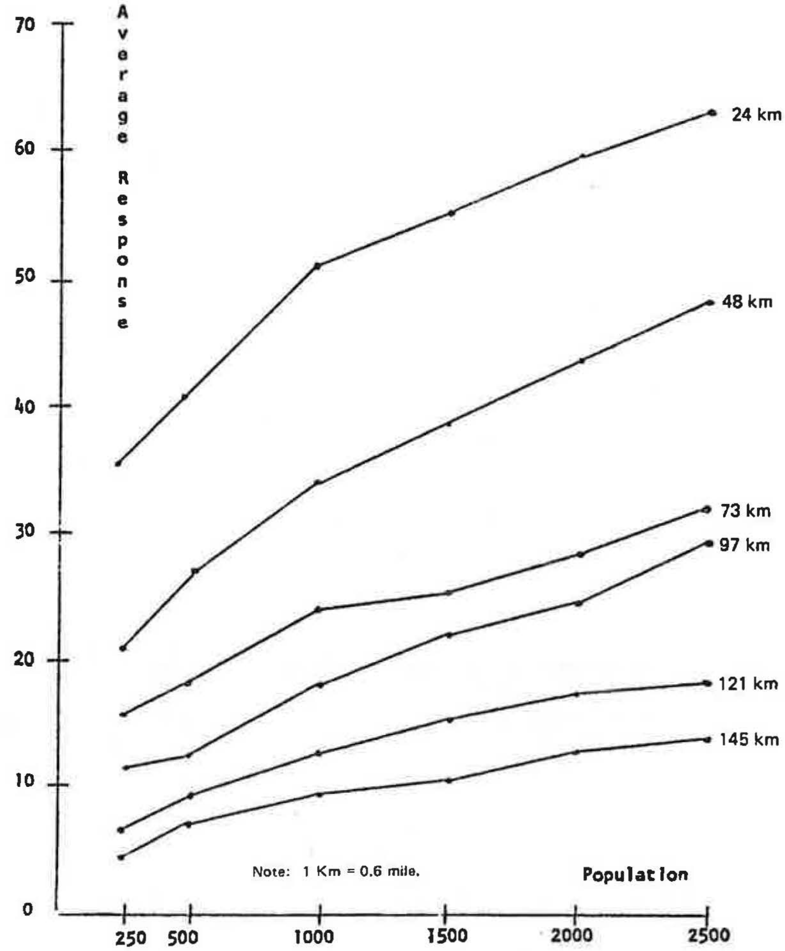
Sets of hypothetical classes of towns were constructed by developing linear regression functions relating population to the number of each of ten types of facilities such as bars, grocery stores, restaurants, and churches. The number of expected functions in each class was predicted from population sizes of 250, 500, 1000, 1500, 2000, and 2500.

This procedure is based on empirical research in central place theory (16). Thus, there are six levels of the composite stimulus (town size and facilities). The six levels of driving distance, chosen by examining actual commuting distances of plant workers (17), were 24, 48, 73, 97, 121, and 145 km (15, 30, 45, 60, 75, and 90 miles). All combinations of towns and distances yield a 6 × 6 factorial design. This design was printed in five different random orders. In addition, four filler combinations more extreme than the design combinations were inserted. Thus, the experiment involved 40 distinct combinations that were presented to respondents on sheets.

Filler combinations are used to transfer response bias away from the experimental combination extremes. Subjects respond more extremely to the fillers that they quickly learn are the best and worst combinations in the design. To test the effects of the order of the combinations in the questionnaires, five different orders were prepared by random draw. Sixty usable questionnaires were obtained from students, faculty, and staff at the University of Wyoming who volunteered to participate. These subjects were randomly assigned to the five order conditions. Thus the experiment is a 6 × 6 × 5 × 12 factorial design (town times distances times orders times subjects). It should be recalled that this sample should be as good as any other for estimating functional form, although the parameter values will be biased for the population as a whole.

Subjects were asked to estimate a numerical value for their degree of preference for each combination by assigning a number (interpreted as a utility measure) between zero (absolutely the worst combination imaginable) and 100 (the best imaginable). Subjects were shown combinations that pretesting had revealed to be very undesirable and very desirable. They were told that all items to be evaluated were considered less ex-

Figure 1. Utility responses by population and distance.




treme than these and that they were to use these combinations to anchor their numerical judgments of preference. All subjects, therefore, completed all combinations.

Data were first analyzed by means of analysis of variance. The results showed that the effect of order was not significant, while that for distance and town size was, as was their interaction. The traditional gravity model of trip distribution would indicate a multiplicative relationship between these factors that would yield a significant interaction; this is confirmed by the analysis. The form of this interaction, however, is critical. Hence, graphical evidence of multiplication is necessary to bolster the statistical evidence of a significant interaction.

As discussed earlier, if the data are graphed as a function of increasing (or decreasing) column values, the difference between each row must increase (or decrease). Thus, the data plot as divergent (convergent) curves. This is approximately true in Figure 1. We can therefore tentatively accept a multiplicative combination rule as a reasonable approximation to the decision process for this experiment.

As demonstrated in Equations 13 and 14, the marginal row and column means are interval scale estimates of the utility values corresponding to the levels of the experimental factors, measured in the units of the dependent variable. These effects or utilities can only arise from variation in the experimental values. Hence,

$$U^1_m = f_1(\text{town size}) + \epsilon_m \tag{22}$$

$$U^2_n = f_2(\text{distance}) + \epsilon_n \tag{23}$$

We may assume specific function forms for $f_1$ and $f_2$; two likely candidates consistent with both psychological and utility theory are

$$U^1_m = a_1 + b_1 \text{ town size}_m^{c_1} + \epsilon_m \tag{24}$$

$$U^2_n = a_2 e^{-b_2 \text{ distance}_n} + \epsilon_n \tag{25}$$

where $U_m$ is measured by the marginal row mean and $U_n$ is measured by the marginal column mean derived from the factorial design. Because these are sums or averages of random variables, it is reasonable to assume them to be normally distributed. The experimental factor values are fixed, so we have a classical fixed-effects regression case and can estimate the desired parameters via least-squares.

If we assume the multiplicative hypothesis encouraged by Figure 1 and the results of the analysis of variance, we can write (assuming Equations 16 and 20 and letting k = 0)

$$U_{mn} = (U_{m.})(U_{.n})/(U_{..}) + \epsilon_{mn} \tag{26}$$

where $U_{m.}$ and $U_{.n}$ are marginal means and $U_{..}$ is the grand mean. Equation 26 is always true if there is a true multiplicative rule underlying the data. Substituting Equations 24 and 25 into 26 yields

$$U_{mn} = [1/(U_{..})]\,(a_1 + b_1 \text{ town size}_m^{c_1})(a_2 e^{-b_2 \text{ distance}_n}) + \epsilon_{mn} \tag{27}$$

By expanding Equation 27 and combining constants, we have the expectation of the following equation.

$$U_{mn} = k_0 + k_1 e^{-b_2 \text{ distance}_n} + k_2(\text{town size}_m^{c_1} \times e^{-b_2 \text{ distance}_n}) + \epsilon_{mn}. \quad (28)$$

The parameters of this equation were estimated via an iterative, least-squares procedure that yielded the following utility expression for town size and distance.

$$U_{mn} = 24.76(e^{-0.021\,41 \text{ distance}_n}) + 1.313(\text{town size}_m^{0.562}) \times (e^{-0.021\,42 \text{ distance}_n}) + \epsilon_{mn} \quad (29)$$

This equation accounts for 98.6 percent of the variance in the experimental design cell means. Both terms are highly significant, as is the overall equation ($F = 1154.8$; $df = 2.33$). Hence, we tentatively retain Equation 29 as a reasonable approximation to the utility function employed in the experiment.

## ESTIMATION OF CHOICE MODEL FROM REVEALED PREFERENCES

Given that the functional form developed in the previous section using functional measurement adequately describes respondents' expressed preferences under hypothetical conditions, the next logical step is to demonstrate that the same functional form usefully describes actual choice of residential location. In order to do so, we must first postulate a model of decision making. Because of measurement errors, omitted variables, and the use of proxy individuals that associate with choice alternatives, we cannot reasonably expect to describe choices perfectly. However, by assuming a distribution of the random elements, we can model the probability with which any alternative is selected.

For this study we have chosen the multinomial logit model. For the sake of brevity, we shall forego the derivation of the logit model (18, 19, 20, 21). In terms of the simple model of location decisions under consideration, this model can be expressed as

$$P(i|A_t) = [e^{V_t(P_i,D_{it})}] \Big/ \left[\sum_{j \in A_t} e^{V_t(P_j,D_{jt})}\right] \quad (30)$$

where

$P(i|A_t)$ = denotation of the probability that town i is selected by person t from some feasible set of towns $A_t$,

$V_t(P_i, D_{it})$ = denotation of the representative utility (i.e., the nonrandom portion of the total utility) for person t,

$P_i$ = population of the ith town, and

$D_{it}$ = distance between the ith town and person t's work place.

The parameters of this model are generally estimated by maximum likelihood. However, Berkson (22) and Theil (23) describe the use of least-squares estimation for binary and multinomial logit respectively when the observations have many repeated entries. For multinomial logit, this can be done by noting that

$$\{[P(i|A_t)]/[P(k|A_t)]\} = \{[e^{V_t(P_i,D_{it})}]/[\sum_j e^{V_t(P_j,D_{jt})}]\} \div \{[e^{V_t(P_k,D_{kt})}]/[\sum_j e^{V_t(P_j,D_{jt})}]\} = e^{V_t(P_i,D_{it}) - V_t(P_k,D_{kt})} \quad (31)$$

Hence

$$\ln[P(i|A_t)/P(k|A_t)] = V_t(P_i,D_{it}) - V_t(P_j,D_{jt}) \quad (32)$$

If we now impose the restriction that $V_t$ is linear in its parameters, the above equation reduces to

$$\ln[P(i|A_t)/P(k|A_t)] = \beta(X_{it} - X_{kt}) \quad (33)$$

where $\beta$ is a vector of parameters and $X_{it}$ and $X_{kt}$ are the vectors of variables characterizing alternatives i and k respectively.

In this technique, each observation represents the proportion of people with common values of X and common choice set $A_t$ who choose alternative i relative to the proportion choosing the base alternative k for that group. Thus, while the actual sample may have a great number of responses, the data for estimation are grouped. For example, the original data used for this study were gathered from a 100 percent sample of workers at six power plants and from a 20 percent sample at one other. Since only population and distance are used as independent variables, each person at a single power plant has the same set of available towns, $A_t$, and the same values of the independent variables characterizing that set. The proportion of people at a power plant choosing a town i for a residence, denoted as $f(i|A_t)$, is a consistent, unbiased estimate for $P(i|A_t)$. Thus, $n[f(i|A_t)/f(k|A_t)]$ is a consistent (though biased) estimate for $\ln[P(i|A_t)/P(k|A_t)]$.

In a Monte Carlo simulation study Domencich and McFadden (21) demonstrated that, when the number of repetitions for each alternative is reasonably large, the bias in least-squares estimation is small and the Berkson-Theil procedure is to be preferred over maximum likelihood on the grounds of computational efficiency.

The use of the Berkson-Theil procedure requires selecting one alternative location to act as the base (denoted by k above) for each power plant. In the results reported below, the town with the median share of the workers was used as a base location for each of the groups of workers employed at the seven power plants.

In our data, the total number of towns used in the models was 46, although each power plant had a different subset of towns available. Some towns were eliminated because they had unusual characteristics that did not match the range of independent variables used in the functional measurement experiment. Given that a base alternative is used for each town, the actual number of observations as input to the estimation program was 39, the original 46 towns minus the 7 used as base alternatives.

Another way to view the base alternative is to recognize that the full set of probabilities $P(i|A_t)$ is not independent; the probabilities must sum to unity. Hence, if there are I alternatives in any choice set $A_t$, only I - 1 of them convey any information; the Ith is redundant.

## RESULTS OF ESTIMATION FROM REVEALED PREFERENCES

In order to test the usefulness of the functional form derived above, the same specification was estimated on the town choice data by nonlinear least squares. A large number of linear parameter specifications using ordinary least squares via the Berkson-Theil method then were estimated. If the functional form derived from the functional measurement experiments is indeed a useful description of actual decision making, then one would expect it to fit the available town choice data better than more commonly applied linear parameter forms.

The results of the estimation for the specification in Equation 29 are summarized below in Equation 34.

Figure 2. Summary of linear parameter models.

| model # | # of parameters | P | D | $\ell_n D$ | $\ell_n P$ | 1/P | 1/D | $\varepsilon^P$ | $\varepsilon^D$ | $D^2$ | $P^2$ | PD | $\ell_n PD$ | $R^2$ | $\bar{R}^2$ | s |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 2 | X | X | | | | | | | | | | | .5225 | .5096 | 1.014 |
| 2 | 2 | | | X | X | | | | | | | | | .5915 | .5805 | .941 |
| 3 | 2 | | | | | X | X | | | | | | | .2249 | .2040 | 1.30 |
| 4 | 2 | | | | | | | X | X | | | | | .3022 | .2833 | 1.23 |
| 5 | 2 | | | | X | | | | X | | | | | .5819 | .5708 | .952 |
| 6 | 2 | | | | X | | X | | | | | | | .2267 | .2058 | 1.29 |
| 7 | 2 | | | X | | | | X | | | | | | .4292 | .4138 | 1.11 |
| 8 | 2 | | | | | | X | X | | | | | | .2082 | .1868 | 1.31 |
| 9 | 2 | | | X | | X | | | | | | | | .5339 | .5213 | 1.01 |
| 10 | 2 | | | | | X | | X | | | | | | .3862 | .3696 | 1.15 |
| 11 | 3 | X | X | | | | | | | | | X | | .6207 | .5996 | .919 |
| 12 | 3 | | | X | X | | | | | | | | X | .6351 | .6148 | .902 |
| 13 | 4 | X | X | | | | | | | X | X | | | .7221 | .6983 | .798 |
| 14 | 4 | X | X | X | X | | | | | | | | | .7468 | .7251 | .762 |
| 15 | 4 | X | X | | | X | X | | | | | | | .6356 | .6044 | .914 |
| 16 | 4 | X | X | | | | | X | X | | | | | .6242 | .5920 | .928 |
| 17 | 4 | X | X | | | | | | | X | | X | | .7127 | .6881 | .811 |
| 18 | 4 | X | X | | | | | | | X | | | X | .5785 | .5424 | .983 |
| 19 | 4 | X | X | | | | | | | | X | X | | .7220 | .6982 | .798 |
| 20 | 4 | X | X | | | | | | | | X | | X | .7191 | .6950 | .802 |
| 21 | 5 | X | X | | | | | | | X | X | X | | .7447 | .7147 | .776 |
| 22 | 5 | X | X | | | | | | | X | X | | X | .7221 | .6894 | .810 |
| 23 | 5 | X | X | X | X | | | | | | | X | | .7581 | .7296 | .755 |
| 24 | 5 | X | X | X | X | | | | | | | | X | .7656 | .7380 | .744 |
| 25 | 6 | X | X | X | X | | | | | X | X | | | .7592 | .7248 | .763 |
| 26 | 7 | X | X | X | X | | | | | X | X | X | | .7752 | .7331 | .751 |
| 27 | 8 | X | X | X | X | | | | | X | X | X | X | .8042 | .7600 | .711 |

$$V_t(P_i,D_{it}) = \underset{(0.091)}{-0.141} \; \exp(\underset{(0.003\,66)}{-0.0128}.D_i) \; \underset{(0.044)}{P_{it}^{0.165}} + \underset{(0.799)}{1.74} \; \exp(\underset{(0.003\,66)}{-0.0128}.D_i) \; P_{it} \quad (34)$$

where

$R^2 = 0.8052$,
$\bar{R}^2 = 0.7885$, and
$s = 0.668$.

These estimates were derived by using a nonlinear least-squares procedure incorporated in Time Series Processor (TSP), an econometric software package developed at Massachusetts Institute of Technology. The numbers in parentheses below the parameter estimates are their estimated standard errors.

Three statistics also are reported. The first, $R^2$, is the percentage of explained variance; $\bar{R}^2$ is the value of $R^2$ corrected for the number of parameters estimated. The standard error of the regression, s, is an estimate of the standard error of the disturbance in the model. Note that in interpreting these measures, one must recall that the dependent variable is $\ln[P(i|A_t)/P(k|A_t)]$, not $P(i|A_t)$.

In order to determine whether the specification derived from the functional measurement was in any sense "better" than more usual functional forms, a series of linear parameter forms was estimated. The first set of these runs used distance and population in linear, exponential, inverse, and logarithmic form. These specifications, however, have only two parameters, while the form in Equation 34 has four. Therefore, a wide range of combinations of the variables, including quadratic forms of the first two, was also estimated. Some of these forms were simply curve-fitting efforts; it is difficult to see how one would arrive at them from any behavioral argument. Others are extended forms of gravity models of spatial interaction.

Figure 2 summarizes these models and is structured so that the columns denote the various independent variables used and the rows correspond to different linear-in-parameters function forms. In each model, an x indicates that a particular variable was used. The figure also summarizes the number of parameters in the model and the values of $R^2$, $\bar{R}^2$, and s.

The most important feature of these results is that, even without correcting for the number of parameters estimated, none of the linear models fits these data better than the specification in Equation 34. The last regression, with eight parameters, comes close in terms of $R^2$-values to the form derived from the functional measurement experiment before adjustment, but it is significantly worse after the degrees of freedom are accounted for.

A second significant point is that the behavioral underpinnings for the form in Equation 34 are relatively clear (after the analysis is performed), but the specifications involving many parameters in Figure 2 are somewhat obscure. For this reason it is unlikely that someone using the revealed preference data would actually choose many of these models. Furthermore, many of the parameter estimates for the linear models with five or more coefficients are statistically insignificant.

A side result not reported in the figure is that many of the coefficients from at least the two-, three-, and four-parameter linear models were statistically more significant (as measured by their t-statistics) than those from the functional measurement form. It is difficult to interpret this result other than to note that, if the specification of either (or both) model(s) is incorrect, then the estimated coefficients and their corresponding t-statistics are in general inconsistent; the higher t-statistics in the linear form may be meaningless and may simply reflect the outcome of a particular form of misspecification.

## CONCLUSION

It is interesting—at least as a mental exercise—to ask the question, What form would a reasonable modeler select if he or she were developing a model from the revealed preference data and the functional measurement experiment had been infeasible? Obviously the answer to such a question depends on the criteria for model selection used by the analyst, but reasonable answers might include simple two-parameter forms such as model 2 or 5 in Figure 2 or quadratic forms such as 13 or 23. It is most unlikely that the form in Equation 34 would ever be considered, particularly since its estimation requires the use of a fairly expensive nonlinear estimation procedure. Without some prior evidence, such as an equation developed in a functional measurement experiment, that such a functional form might be useful, most modelers would never even consider nonlinear forms unless no reasonable linear model could be developed.

Obviously, the empirical evidence presented in this paper is extremely limited, and it would be premature to suggest that functional measurement or some other, similar technique should be a major component of a demand model estimation strategy. However, the relatively low cost of laboratory experiments does make it appear to be an attractive way to analyze functional form. The two-step procedure proposed in this paper offers at least one feasible approach to improving travel demand models and may actually reduce the cost of model estimation by restricting the class of model specifications with which travel demand modelers need be concerned.

## REFERENCES


1. K. M. Gaver and M. S. Geisel. Discriminating Among Alternative Models: Bayesian and Non-Bayesian Methods. In Frontiers in Econometrics (P. Zarembka, ed.), Academic Press, New York, 1974.
2. G. E. P. Box and D. R. Cox. An Analysis of Transformations. Journal of the Royal Statistical Society, Series B, 1964.
3. J. Ramsey. Classical Model Selection Through Specification Error Tests. In Frontiers in Econometrics (P. Zarembka, ed.), Academic Press, New York, 1974.
4. J. J. Louviere, L. L. Beavers, K. L. Norman, and F. Stetzer. Theory, Methodology, and Findings in Mode Choice Behavior. Institute of Urban and Regional Research, Univ. of Iowa, Iowa City, Working Paper No. 11, 1973.
5. N. H. Anderson. Information Integration Theory: A Brief Survey. In Contemporary Developments in Mathematical Psychology 2 (D. H. Krantz, R. C. Atkinson, R. D. Luce, and P. Suppes, eds.), W. H. Freeman, San Francisco, 1974.
6. N. H. Anderson. Algebraic Models in Perception. In Handbook of Perception 2 (E. C. Carterette and M. P. Friedman, eds.), Academic Press, New York, 1974.
7. N. H. Anderson. Social Perception and Cognition. Center for Human Information Processing, Univ. of California at San Diego, La Jolla, Technical Rept. CHIP 62, June 1976.
8. N. H. Anderson. How Functional Measurement Can Yield Validated Interval Scales of Mental Quantities. Journal of Applied Psychology, in press.
9. J. J. Louviere. Psychological Measurement of Travel Attributes. In Determinants of Travel Choice (D. A. Hensher and M. Q. Dalvi, eds.), Teakfield, Farnborough (Saxon House Studies), London, 1978.
10. K. L. Norman. Attributes in Bus Transportation: Importance Depends on Trip Purpose. Journal of Applied Psychology, Vol. 62, No. 2, 1977.
11. I. P. Levin. The Development of Attitudinal Modelling Approaches in Transportation Research. Paper presented to the workshop on Attitudes, Attitudinal Measurement, and the Relationship Between Behavior and Attitudes, Third International Conference on Behavioral Travel Modelling, Barossa Valley, Australia, April 1977.
12. I. P. Levin and M. J. Gray. Analysis of Human Judgment in Transportation. Paper presented to the Special Sessions on Human Judgment and Spatial Behavior, Great Plains/Rocky Mountains AAG Meetings, Manhattan, KS, Oct. 1976.
13. J. Shanteau and R. Phelps. Livestock Judges: How Much Information Can the Experts Use? Dept. of Psychology, Kansas State Univ., Manhattan, Technical Rept. No. (KSU-HIPI), Rept. No. 76-15, Dec. 1976.
14. J. M. Piccolo and J. J. Louviere. Information Integration Theory Applied to Real-World Choice Behavior: Validational Experiments Involving Shopping and Residential Choice. Paper presented to the

Special Session on Human Judgment and Spatial Behavior, Great Plains/Rocky Mountains AAG Meetings, Manhattan, KS, Oct. 1976.
15. J. J. Louviere, J. M. Piccolo, R. J. Meyer, and W. Duston. Theory and Empirical Evidence in Real-World Studies of Human Judgment: Three Shopping Behavior Examples. Center for Behavioral Studies, Institute for Policy Research, Univ. of Wyoming, Laramie, Research Paper No. 1, March 1977.
16. B. J. L. Berry. Geography of Market Centers and Retail Distribution. Prentice-Hall, Englewood Cliffs, NJ, 1967.
17. Old West Commission. Construction Worker Profile. Billings, MT, 1975.
18. R. D. Luce and P. Suppes. Preference, Utility and Subjective Probability. In Handbook of Mathematical Psychology, Vol. 3 (R. D. Luce, R. R. Bush, and E. Galanter, ed.), Wiley, New York, 1965.
19. D. McFadden. Conditional Logit Analysis of Qualitative Choice Behavior. In Frontiers in Econometrics (P. Zarembka, ed.), Academic Press, New York, 1974.
20. M. G. Richards and M. E. Ben-Akiva. A Disaggregate Travel Demand Model. Saxon House, D.C. Heath Ltd., Westmead, England, 1975.
21. T. Domencich and D. McFadden. Urban Travel Demand: A Behavioral Analysis. Elsevier, New York, 1975.
22. J. Berkson. A Statistically Precise and Relatively Simple Method of Estimating the Bioassay with Quantal Response, Based on the Logistic Function. Journal of the American Statistical Association, Vol. 48, 1953, pp. 565-599.
23. H. Theil. A Multinomial Extension of the Linear Logit Model. International Economic Review, Vol. 10, 1969, pp. 251-259.

*Publication of this paper sponsored by Committee on Passenger Travel Demand Characteristics and Committee on Traveler Behavior and Values.*

**The work described in this paper was performed while both authors were at Cambridge Systematics, Inc.*

# Effects of Employment and Residential Location Choices on Urban Structure: A Dynamic Stochastic Simulation

Timothy J. Tardiff, Tenny N. Lam, and Brian F. Odell, Department of Civil Engineering, University of California, Davis

**The pattern of home-to-work linkages in urban areas is affected by household mobility decisions. This paper describes a dynamic stochastic simulation model designed to illustrate the effects of mobility decisions on urban structure. The major feature of the model is the representation of household changes in employment or residential locations or both. The sequential process of the decision to move and the search for and selection of new location is specified. The role of accessibility in this process is an important consideration in the model, which allows quick execution of simulation experiments. Experiments consist of alternative input assumptions involving factors such as city size, numbers and locations of job centers and dwelling units, initial pattern of home-to-job linkages, moving rate, and importance of accessibility in selecting new locations. Two types of experiments are presented. The first examines the dynamic properties of the model by varying the initial pattern of home-to-job linkages and the mobility rate. The major conclusion is that there appears to be an equilibrium pattern of home-to-job linkages that is independent of the initial configuration and mobility rate. The second type of experiment involves the variation of the importance of accessibility in the mobility decision process. The results show that the pattern of home-to-job linkages varies in the expected way with changes in the decision process.**

The pattern of linkages between residential and employment locations in urban areas and the changes in this pattern over time are the result of many complex economic and social processes. To isolate the contribution of each by looking at the total patterns through time series or cross-sectional analyses is difficult, if not impossible. An alternate approach is to study theoretical models that deal with a small number of processes, or a small part of the problem, at a time. Data and theory must finally agree, of course. In attempting to model the effects of the processes on urban structure, one must make simplifying assumptions in order to make the problem analytically tractable. The nature of the assumptions is dictated by the purposes of a particular modeling approach.

The purpose of the present modeling effort is to examine the effects of accessibility-based household decision rules on the changing structure of prototypical urban areas. The emphasis on accessibility is consistent with numerous previous studies and also facilitates an examination of the transportation requirements for urban areas. The model is a dynamic stochastic simulation model that processes household location decisions sequentially within a given time period. Each time period represents a fraction of the population making intraurban mobility decisions, i.e., residential location changes, employment location changes, or simultaneous home-job changes. An additional objective of the modeling effort is to generate a basis for conducting controlled theoretical experiments, the purpose of which is to examine questions such as stability, statistical variabilities, and observability of urban relationships. These issues are important for linking theoretical results with empirical data.

Much simplification of the socioeconomic details such as the distributions of household and dwelling unit types over the urban areas is allowed in order to sharpen the model's focus on the accessibility-based choice decisions. The central issue is how household decision rules affect the overall state of the hypothetical urban system. Simplification allows the model to be processed over many time periods at fairly low computational expenses and thus facilitates examination of the dynamic aspects of urban areas.

A primary concern of the modeling effort is the existence of constraints at both the household and urban system levels. In regard to the former, the dynamic aspects of the model implicitly incorporate three constraints: those on the search for new residences; those on the available choices, which are limited by the number of vacancies in a given time period; and those implicit in the competition among households for the available dwelling units. All of these constraints potentially lessen the pure effects of accessibility in the household decision rule. There appear to be dynamic equilibria at the systemwide level that are determined, in part, by the nature of the household location decision rule. This implies that planned changes in the urban structure that are inconsistent with the underlying decision rules are unstable. Therefore, effective planning policies seem to be constrained by basic household behavior.

## BACKGROUND

Since the emphasis of the present study differs from that of much previous work, the methodological approach also differs. Of the two major categories of previous work, the first emphasizes the description and analysis of the spatial patterns of urban areas, usually in an atemporal manner. Studies of this nature often develop and apply spatial interaction concepts such as accessibility (1, 2, 3, 4) or intervening opportunities (5, 6). Although many of these models yield fairly rich spatial detail for particular urban areas, they usually do not explicitly consider the household decision processes that generate the macrolevel patterns.

The second category of previous work focuses on explaining household spatial choice processes. Of particular interest for this study is the examination of intraurban location choices. Studies of this nature have used various methodologies and have arrived at conclusions that differ according to the disciplinary perspective of the study. Some studies, such as Lerman's (7), hypothesize an economic explanation of location choice. In these studies, home-to-work accessibility is implicit in transportation costs. Other studies focus on the search process involved in moving. In these studies the key accessibility concept is often the accessibility of potential new homes to the old home, rather than home-to-work accessibility (8, 9). A third type of study emphasizes the effects of household characteristics and/or reported reasons for moving on the actual decision to move (10, 11, 12, 13). These studies seem to indicate that accessibility is a relatively unimportant reason for moving. Household, dwelling unit, and neighborhood characteristics are cited much more frequently. However, there is evidence that actual behavior of households is consistent with an accessibility hypothesis (14, 15).

Recently there have been a few notable efforts to explicitly incorporate household choice behavior into models that explain spatial structure to create a hybrid of the two categories (16, 17). Often these efforts are consistent with economic explanations of spatial choice. Further, there is some attempt to dynamically represent the urban structure. However, such models have had fairly detailed spatial descriptions, have been specific to a particular metropolitan region, and have not been explicitly concerned with the sensitivity of spatial patterns to changes in behavioral assumptions.

In addition to the importance of accessibility, two other important issues are encountered in previous work. First, it is possible to distinguish between optimal versus nonoptimal decision processes. At the household level optimality is implicit in the conventional utility maximization assumption of economic models. In contrast to this are explanations involving a three-phase decision process: decision to move, search, and selection (18, 19). These explanations do not rely on optimality assumptions in the sense that there are usually constraints in the search phase that inhibit the selection of the optimal location. Systemwide models allocate activities based upon some optimality criterion such as welfare maximization (8, 16, 20, 21) or some other planning goal such as systemwide travel minimization (22, 23, 24). In contrast to the optimality at a systemwide level are models using an accessibility allocation rule that is not explicitly optimal (1, 4, 25, 26).

The second issue involves the role of home-to-job linkages in location choices. Consistent with monocentric spatial models (27), most economic models assume that job location selection occurs before residential location choice and that the latter depends on the former. On the other hand, job location is not explicitly considered in explanations that hypothesize search processes dependent on the old home. Both explanations ignore the possibility of changes in job location without changes in residential location. Recent empirical evidence suggests that such a location decision may also be important (28, 29).

The various possibilities are summarized in a classification scheme suggested in a review article by Senior (30). The entire population in an urban area in a given time period can be classified as nonmovers, job changers with fixed residences, residential movers with fixed job locations, and simultaneous residential and job movers. It is apparent that the usual employment-centered residential choice assumption ignores some types of moves.

## SIMULATION MODEL

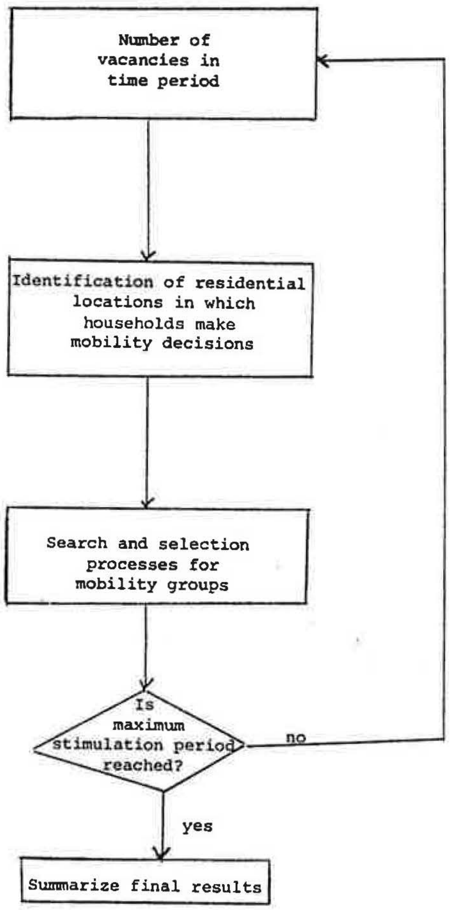
The basic model represents a highly idealized prototypical urban area. No attempt is made to classify households or dwelling units; therefore, zonal demographic and individual characteristics remain unspecified in the simulation process. The random component of the household location choice process allows potential effects of household behavioral and characteristic differences to be realized without explicit specification.

The urban area is represented by a rectangular array containing housing and job center locations. In general, the relative spatial separation between the cells of the array may be manipulated through topological transformations to reflect actual spatial patterns and/or transportation systems. Each residential location contains a small number of identical dwelling units. The distribution of dwelling units among locations can represent any desired density, although most of the simulations accomplished to date have used uniform densities. Similarly, the size and location of job centers can be specified in any manner, although there is no differentiation by job classification.

At any given time the structure of the urban area is represented by the distribution of job locations for residents of each residential location. Using this basic information, it is possible to develop various systemwide summary statistics such as the distribution and moments of the distribution of home-to-work distance.

The dynamic properties of the model are a result of the representation of changes in employment and residential locations over time. The essentials of these processes are represented in Figure 1.

Figure 1. Major components of the simulation model.




In addition to accounts of the temporary and final states of the urban system, the dynamic model contains three major components. The first of these is the generation of the number of movers for a particular time period. This number is selected randomly from a Poisson distribution with expected value equal to a fixed proportion of the population. This proportion can be interpreted as a mobility rate for a given time period. There are two interpretations of this mobility rate. First, higher or lower rates may represent higher or lower mobility propensities for a time period of fixed length. Alternatively, increases or decreases in the mobility rate may represent decreases or increases, respectively, in the length of the basic time period. Operationally, the two interpretations are indistinguishable.

The remaining two major components capture the essentials of the hypothesized household location choice processes. These processes are consistent with previous work suggesting a sequential decision process that involves a decision to move, a search process, and a selection process rather than the individual or systemwide optimization procedure.

The representation of the decision to move is the first of these components. Consistent with some of the survey research findings on decisions to move (10, 11, 12, 13) and also with the Morrison concept of the existence of a hypermobile population (31), the decision to move is assumed to be determined by household or environmental characteristics and to be independent of home-to-work accessibility considerations. This strategy is also consistent with that used in other recent dynamic models (16, 17). Since the present model does not distinguish among these types of characteristics, the location of households making mobility decisions is identified randomly as suggested by Harsman and Snickars (20).

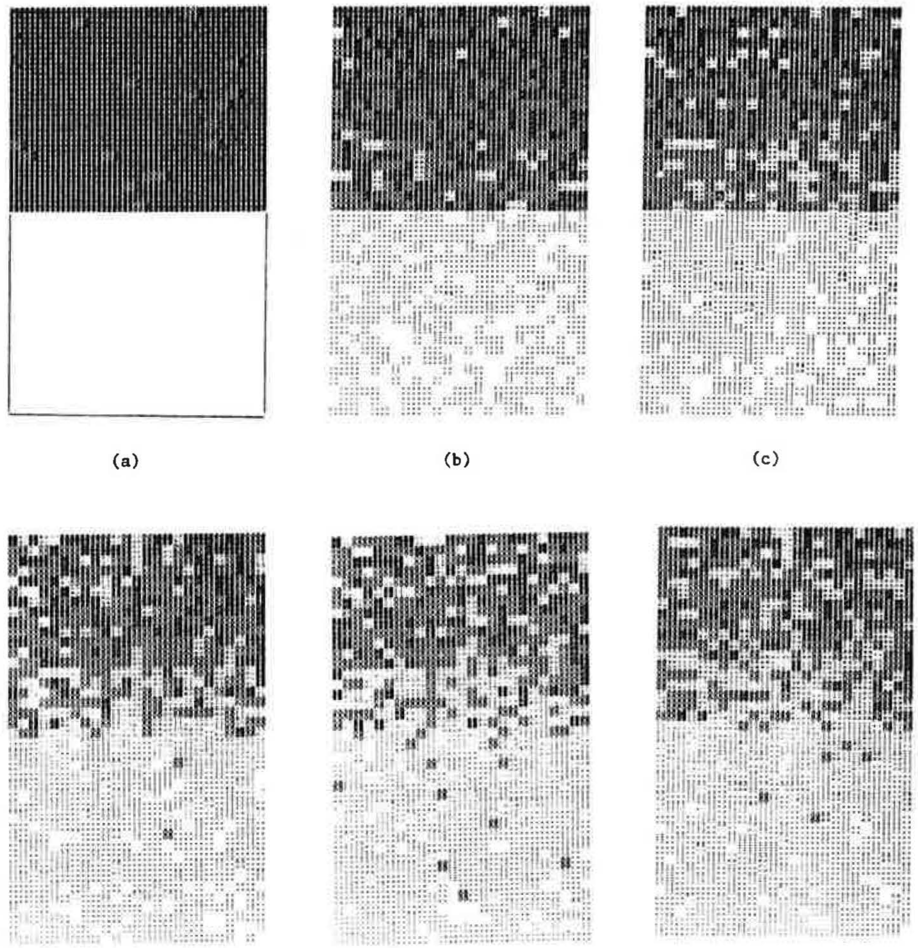
The identification process involves two random processes. First, the location of the household is selected randomly from all possible locations. Then, the household is classified probabilistically by the type of mobility decision: job change only, residential change only, or simultaneous job and residence change. The classification probabilities are established as input parameters. This procedure continues until the previously established households have been identified and classified. Movers are accumulated by mobility type in a regional pool to await subsequent processing.

The last major operational component represents the search and selection processes. There are two levels of sequential processing. First, individual households are randomly selected from the mover pool and randomly assigned to either available house or available job locations. The procedure is somewhat similar to that used by Mason (17), although this process is random while Mason's process prioritized households by income. Second, individual households drawn from the mobility pool and assigned to initial locations sequentially encounter housing or job opportunities. They make an acceptance or rejection decision based on a preselected choice function that usually specifies that the probability of acceptance decreases with an increase in the city block distance between the potential opportunity and the relevant home or job location. The role of accessibility in the choice function is consistent with economic explanations of location choice and with some survey research findings suggesting that, while accessibility might not be important in the decision to move, it may be an important consideration in the selection of the new location (10). By varying the importance of the random component of the choice function, it is possible to test various hypotheses that give greater or less weight to the importance of accessibility.

The search and selection processes are similar to actual or proposed strategies in other dynamic models but differ in important ways. First, the random nature of the search process has been suggested by Okabe (32) and Harsman and Snickars (20). However, both of those studies hypothesize a simultaneous deterministic choice among a prespecified number of possible vacancies rather than the sequential probabilistic choice process in the present model. The sequential processing of households contrasts with the models incorporating systemwide optimization rules (16, 32).

The details of the search and selection processes are as follows. First, a mobility type is selected probabilistically (without replacement) from the list of households generated by the previous component. The probabilities are proportional to the actual numbers of households in each mobility category at the time the particular household is processed. If the mobility type is a job change only, a residential location is selected randomly from those locations with job-changing residents. Next, a job location is selected probabilistically (without replacement). The probabilities are proportional to the number of job vacancies at each job center. The accessibility-based choice function is then used to probabilistically accept or reject the job location. Since rejected job locations are not replaced, the household must accept the last location if all others have been exhausted. As all jobs at a given location are identical, this strategy is consistent with the assumption that rejection of a job implies rejection of all job opportunities at the location for that time period.

For the residential change only, a job location is selected probabilistically in proportion to the number of



(a) (b) (c)

(d) (e) (f)

Figure 2. Pattern of home-to-work linkages over time (initial pattern concentrated at time periods 0, 10, 20, 30, 40, and 50).

residence-changing job holders at each location at the time the household is processed. A residential location is then selected randomly (with replacement) from those locations with housing vacancies. The accessibility-based decision rule is then used to accept or reject the residential location. The process continues sequentially until a residential location is selected or until a prespecified maximum search length is encountered, in which case the last location is selected.

Households simultaneously changing jobs and residences are first probabilistically classified into two groups: those who select the residence first and those who select the job first. The classification probabilities are prespecified. Households subclassified into the first category are then processed in the same way as job changers and those in the second subclassification are processed as residential changers.

The complete processing of all moving households marks the end of the time period. Iterative processing through subsequent time periods, or moving cycles, continues until a terminal period is encountered.

## ILLUSTRATIVE RESULTS

The purpose of the simulation is to study the effects of household mobility processes and various physical characteristics of urban systems on the home-to-work transportation requirements. By sacrificing detail, a wide range of simulation experiments can be examined.

Each simulation experiment involves variation of one or more of the following inputs: (a) the dimensions of the rectangular array; (b) the number, sizes, and location of job centers; (c) the distribution of dwelling units within the array; (d) the probabilistic accessibility-based location choice rule; (e) the expected proportions of mobility types; (f) the mobility rates; (g) the initial pattern of home-to-job linkages; and (h) the level of service on transportation links.

The experiments performed to date can be classified by the essential purpose of the experiment. All experiments have used the same expected proportions for mobility types: one-third job changers, one-third residence changers, and one-third simultaneous changers. In the last category, half are expected to select the residence first. Almost all the experiments have used a uniform distribution of ten dwelling units per location and have assumed uniform level of service on transportation links, although some preliminary results modifying these assumptions are available.

Four major categories of experiments have been performed. First, the dynamic properties of urban systems have been examined by varying the initial pattern of home-to-job linkages and the mobility rates. Second, the effects of the accessibility-based decision rule on home-job patterns have been examined by varying the choice function. Third, the effects of urban structure on transportation requirements have been examined by varying the dimensions of the array and the number, size, and location of job centers. Finally, the effects of variations in transportation level of service and the density distribution of dwelling units were examined.

Over 40 simulation experiments have been run. Because the purpose of this paper is to illustrate methodology, an exhaustive description of the results will not be given. Rather, general findings from the first two categories of experiments will be described. These categories contain the most easily interpretable results and the most definitive conclusions at this time.

Some results involving dynamic properties have been described elsewhere (33) and will only be summarized here. The major conclusion is that there appears to be an equilibrium pattern in home-to-job linkages aggregated over distance that is independent of the initial configuration and the mobility rate. However, the time required to reach equilibrium varies with the initial configuration and the mobility rates.

This conclusion is based on five experiments involving a 40 x 25 array with ten dwelling units per location. Two job centers located in the middle of the array in the short dimension and one-quarter of the length of the array from the edge in the long direction both contained 5000 jobs. Three percent of the dwelling units were vacant. The accessibility function was linear with a maximum probability of 95 percent at the minimum distance (1 block) and 5 percent at the maximum distance (42 blocks).

Three initial patterns and three mobility rates were used. The three patterns combined into a uniform pattern (roughly equal distribution of home-to-job linkages for each residential location), a concentrated pattern (every resident working at the closest job center), and the equilibrium pattern. The last pattern resulted from starting with all residences and jobs vacant and assigning the entire population in one time period. By necessity, all movers were simultaneous job and residential changers with probability of subclassifications equal to one-half. The three mobility rates were 0.5, 1, and 4 percent.

The apparent equilibrium is that obtained with the equilibrium initial pattern. The dynamic aspects of the path to equilibrium are illustrated in Figure 2. This figure depicts the distribution of households working in one job center for each residential location. Darker areas represent higher concentrations of workers. The initial configuration is the concentrated one, and the mobility rate is 4 percent. The pattern changes from one in which everyone in one-half of the city and no one in the other half is working at the job center to the equilibrium position in which the split is roughly two-thirds and one-third.

There are two major implications. First, for experiments concerned with equilibrium rather than dynamic properties, the equilibrium initial pattern can be generated directly, with substantial savings in computation costs. Second, the existence of equilibria seems to indicate that initial patterns inconsistent with the choice behavior resulting in the equilibria are unstable. For example, the situation involving the concentrated pattern may represent the outcome of the frequently advo-

Figure 3. Alternative accessibility-based location choice functions.

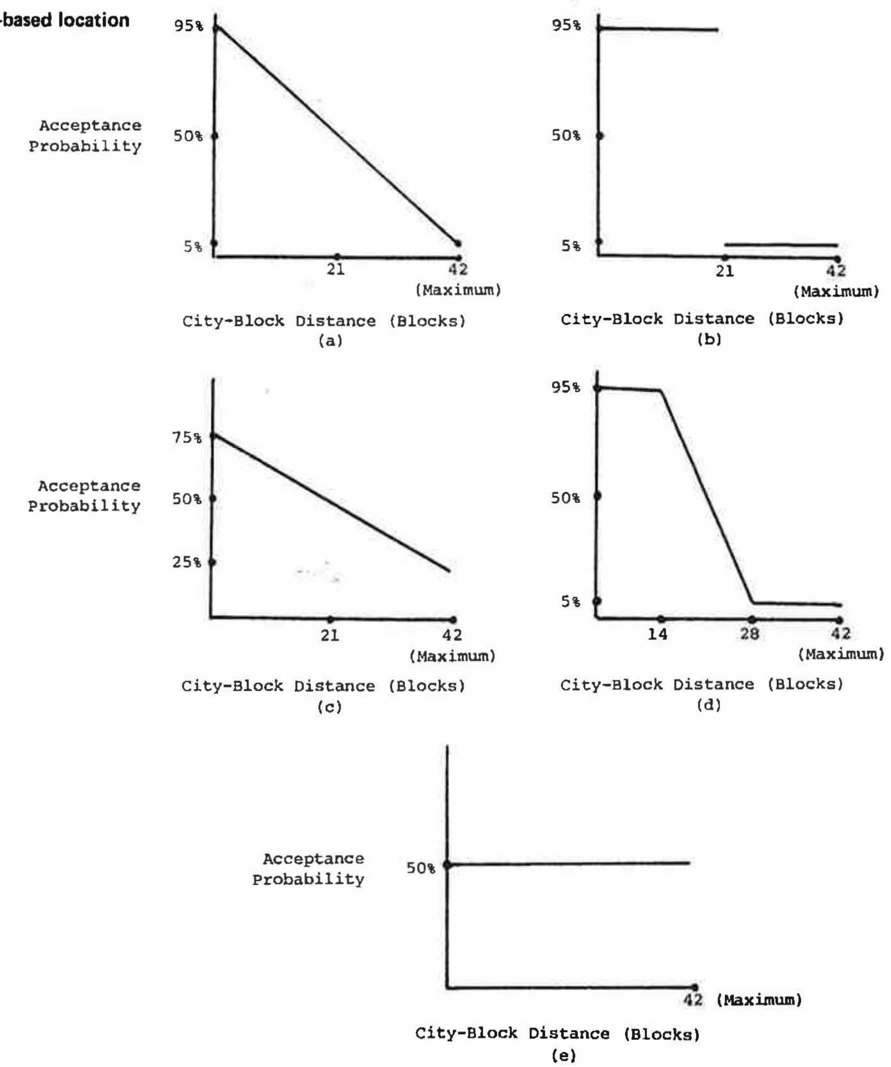


Table 1. Means and variances of home-to-work distance distribution for various location choice functions.

| Item | Choice Function a | b | c | d | e | Concentrated Initial Assignment | Uniform Initial Assignment |
|---|---|---|---|---|---|---|---|
| Mean | 15.89 | 12.86 | 17.26 | 13.33 | 18.92 | 11.25 | 19.00 |
| Variance | 74.18 | 35.79 | 85.86 | 42.99 | 94.14 | 21.49 | 93.68 |

Table 2. Frequency distribution of the number of residential locations containing N workers at the first job center for various location choice functions.

| N | Choice Function a | b | c | d | e | Concentrated Initial Assignment | Uniform Initial Assignment |
|---|---|---|---|---|---|---|---|
| 10 | 16 | 187 | 2 | 109 | 2 | 381 | 0 |
| 9 | 38 | 109 | 28 | 116 | 4 | 100 | 0 |
| 8 | 88 | 54 | 48 | 78 | 37 | 17 | 0 |
| 7 | 129 | 46 | 109 | 77 | 102 | 2 | 0 |
| 6 | 137 | 72 | 183 | 73 | 195 | 0 | 0 |
| 5 | 145 | 65 | 200 | 77 | 242 | 0 | 854 |
| 4 | 129 | 44 | 183 | 72 | 229 | 0 | 134 |
| 3 | 130 | 44 | 125 | 65 | 122 | 0 | 12 |
| 2 | 108 | 26 | 103 | 68 | 49 | 0 | 0 |
| 1 | 63 | 120 | 14 | 112 | 13 | 0 | 0 |
| 0 | 17 | 233 | 5 | 153 | 5 | 500 | 0 |

cated new town planning strategy of minimizing home-to-work distance.

The effects of alternative assumptions about the importance of accessibility in the location choice process can be examined by changing the accessibility-based location choice function. Five alternative functions, which are represented graphically in Figure 3, were examined. The dimensions of the array, the density distribution of dwelling units, and the number, size, and location of job centers were the same as in the previous sets of experiments. All experiments were the result of generating the equilibrium initial pattern.

The mean probability of acceptance, unweighted by the number of opportunities at each distance, is one-half for all alternative functions. However, the unweighted variances differ substantially. Conceptually, the alternative functions allow accessibility and other factors represented by the probabilistic nature of the functions to have varying importance. For example, function (e) in Figure 3 is consistent with the hypothesis that accessibility is of no importance. On the other hand, function (b) represents a situation where accessibility is crucial, i.e., close locations are accepted with near certainty and further locations are rejected, also with near certainty. Function (b) also represents the case in which decisions are made with respect to a maximum accessibility threshold.

Tables 1 and 2 present data that summarize the results of the experiments according to the choice functions a-e in Figure 3. The first table presents the first and second moments of the home-to-work distance distribution for the urban area, while the second table lists the frequency distributions of the concentration of workers at the first job center over the 1000 residential locations. In addition, both tables contain the corresponding information for the concentrated and uniform initial patterns used in the previous set of experiments. These latter results delimit the range of possibilities; i.e., the concentrated pattern has accessibility as totally deterministic, while the uniform pattern is essentially independent of accessibility considerations.

Qualitatively, the results in the tables are not surprising. The most important accessibility is in the location choice function, the shorter the average home-to-work distance. Further, great importance assigned to accessibility appears to result in small variations around the average distance and, consequently, more homogeneity within individual residential locations. That is, Table 2 shows that the functions that correspond to greater emphasis on accessibility (b, d, concentrated) lead to situations where a large majority of residential locations have most of their workers at a single job center. Conversely, the remaining functions (a, c, e, uniform) yield more heterogeneous results. That is, the majority of residential locations have a fairly even distribution of workers at the two job centers.

Preliminary investigation using the results of these and other experiments indicates that these qualitative findings can be strengthened by relating input parameters to model outputs. For example, there appears to be a strong linear relationship between the unweighted second moment of the location choice function and the second moment of the home-to-work distance distribution. Verification of this relationship and further attempts to relate other inputs and outputs in a rigorous fashion are an important component of future work and will be instrumental in demonstrating the ultimate potential of the methodological approach.

## SUMMARY AND CONCLUSIONS

The simulation methodology just described has the potential for yielding insights into the manner in which household choice processes and the locations of spatial opportunities result in patterns of home-to-work linkages. Current results are preliminary and await further research that will define relationships between model inputs and outputs more rigorously.

The modeling system is intentionally abstract. Consequently, it is not appropriate for examining the effects of a wide range of urban policies, such as policies involving housing and/or particular subgroups of the population. In addition, since the model does not necessarily represent any particular urban area, the conclusions that do emerge from particular types of experiments are likely to be general in nature.

These limitations contribute to the particular strength of the methodology, however. Rather than applying to only one particular urban area, as do most more detailed models, the current methodology allows the quick examination of conditions for a wide range of actual or potential urban systems. Consequently, any general findings are more likely to describe the class of urban systems rather than a particular urban area. In this way, the approach lies between the two major approaches used in previous studies of location choice. That is, rather than attempting to study particular location choice behavioral processes on the one hand or attempting to describe spatial regularity at an aggregate level on the other, this methodology has the potential for examining, in detail, the consequences of location behavior on the nature of urban systems in general.

Despite the abstract nature of the methodology, the existing model has been developed with careful attention to existing knowledge. This is reflected in the dynamic capabilities of the model, in the separation of the location choice process into the decision to move and search and select components, in the division of mobility decisions into job change and residence change and simultaneous change components, and in the fact that the location choice function contains both accessibility and other factors subsumed under the random component of the function.

These four features allow a wide range of tests of competing hypotheses. In addition, the existing model can easily be changed to incorporate other hypotheses. For example, the dimensions of the urban area may expand over time, representing growth. Experiments of this nature have already been performed. Similarly, a simultaneous decision process among a number of alternative locations instead of the sequential binary process of acceptance or rejection in the current model can be easily represented within the simulation structure.

Since the model is abstract, it will ultimately be necessary to determine the extent to which it offers insights into real urban systems. At one extreme it may be only the mathematical system represented by the computer code, and at the other extreme it may be adapted to offer insights into particular urban areas. Preliminary evidence indicates that the model does, indeed, result in a reasonable representation of particular areas as indicated by a comparison of empirical and model home-to-work distance distributions (34).

In conclusion, a methodology has been developed that has the potential for yielding insights into the equilibrium and dynamic properties of urban systems. Such insights will be useful in identifying the extent to which the outcomes of planning decisions, such as the location of employment centers and transportation facilities, are constrained by household choice behavior and also by the reciprocal constraints of the existing urban opportunity structure on household choice.

## ACKNOWLEDGMENTS

Research for this paper was supported in part by a National Science Foundation grant on research initiation and urban structure and transportation requirements. Additional support was received from the Rockefeller Fund in Environmental Studies at the University of California, Davis.

## REFERENCES

1. W. Goldner. The Lowry Model Heritage. American Institute of Planners Journal, Vol. 37, 1971, pp. 100-110.
2. W. G. Hansen. How Accessibility Shapes Land Use. American Institute of Planners Journal, Vol. 25, 1959, pp. 73-76.
3. T. R. Lakshmanan and W. G. Hansen. A Retail Potential Model. American Institute of Planners Journal, Vol. 31, 1965, pp. 134-143.
4. I. S. Lowry. Seven Models of Urban Development: A Structural Comparison. HRB, Special Rept. 97, 1967, pp. 121-146.
5. G. T. Lathrop and J. R. Hamburg. An Opportunity-Accessibility Model for Allocating Regional Growth. American Institute of Planners Journal, Vol. 31, 1965, pp. 95-102.
6. S. A. Stouffer. Intervening Opportunities and Competing Migrants. Journal of Regional Science, Vol. 2, 1960, pp. 1-26.
7. S. R. Lerman. Location, Housing, Automobile Ownership, and Mode to Work: A Joint Choice Model. TRB, Transportation Research Record 610, 1976, pp. 6-11.
8. L. A. Brown, F. E. Horton, and R. I. Wittick. On Place Utility and the Normative Allocation of Intraurban Migrants. Demography, Vol. 7, 1970, pp. 173-183.
9. W. A. V. Clark. Measurement and Explanation in Intraurban Residential Mobility. Tijdschrift voor Economiste en Sociale Geografie, Vol. 61, 1970, pp. 49-57.
10. E. W. Butler, F. S. Chapin, Jr., G. C. Hemmens, E. J. Kaiser, M. A. Stegman, and S. F. Weiss. Moving Behavior and Residential Choice. NCHRP, Rept. 81, 1969.
11. G. Duncan and S. Newman. People as Planners: The Fulfillment of Residential Mobility Expectations. In Five Thousand American Families - Patterns of Economic Progress (G. J. Duncan and J. N. Morgan, eds.), Institute for Social Research, Ann Arbor, Vol. 3, 1975.
12. J. B. Lansing, E. Mueller, and N. Barth. Residential Location and Urban Mobility. Survey Research Center, Univ. of Michigan, 1964.
13. E. Roistacher. Residential Mobility: Planners, Movers, and Multiple Movers. In Five Thousand American Families - Patterns of Economic Progress (G. J. Duncan and J. N. Morgan, eds.), Institute for Social Research, Ann Arbor, 1975.
14. H. J. Brown. Changes in Workplace and Residential Location. American Institute of Planners Journal, Vol. 41, 1975, pp. 32-39.
15. O. Hochma, G. Fishelson, and D. Pines. Intraurban Spatial Association Between Places of Work and Places of Residence. Environment and Planning, Vol. 7, 1975, pp. 273-278.
16. G. K. Ingram, J. F. Kain, and J. R. Ginn. The Detroit Prototype of the NBER Urban Simulation Model. National Bureau of Economic Research, New York, 1972.
17. R. P. Mason. ASMUG—Simulation Model of Urban Growth. ASCE, Journal of the Urban Planning and Development Division, Vol. 103, 1977, pp. 13-25.
18. L. A. Brown and E. G. Moore. The Intraurban Migration Process: A Perspective. Geografiska Annaler, Vol. 52, Series B, 1970, pp. 1-13.
19. J. Wolpert. Behavioral Aspects of the Decision to Migrate. Papers of the Regional Science Association, Vol. 15, 1965, pp. 159-169.
20. B. Harsman and F. Snickars. Disaggregated Housing Demand Models: Some Theoretical Approaches. Papers of the Regional Science Association, Vol. 34, 1975, pp. 121-143.
21. J. D. Herbert and B. H. Stevens. A Model for the Distribution of Residential Activities in Urban Areas. Journal of Regional Science, Vol. 2, 1960, pp. 21-36.
22. J. Berechman. A Design Model for Optional Land Use Activity Allocation. Paper presented at the Meetings of the Operations Research Society of America, Las Vegas, Nov. 1975.
23. P. Gordon and W. K. MacReynolds. Optimal Urban Forms. Journal of Regional Science, Vol. 14, 1974, pp. 217-231.
24. G. C. Hemmens. Experiments in Urban Form and Structure. HRB, Highway Research Record 207, 1967, pp. 32-42.
25. J. L. Edwards and J. L. Schofer. Relationships Between Transportation Energy Consumption and Urban Structure: Results of Simulation Studies. TRB, Transportation Research Record 599, 1976, pp. 52-59.
26. A. G. Wilson. The Use of Entropy Maximizing Models. Journal of Transport Economics and Policy, Vol. 3, 1969, pp. 108-126.
27. W. Alonso. Location and Land Use. Harvard Univ. Press, Cambridge, MA, 1964.
28. M. E. Beeseley and M. Q. Dalvi. Spatial Equilibrium and Journey to Work. Journal of Transport Economics and Policy, Vol. 8, 1974, pp. 197-222.

29. J. Siegal. Intermetropolitan Migration: A Simultaneous Model of Employment and Residential Location of White and Black Households. Journal of Urban Economics, Vol. 2, 1975, pp. 29-47.
30. M. L. Senior. Approaches to Residential Location Modeling 1: Urban Ecological and Spatial Interaction Models (A Review). Environment and Planning, Vol. 5, 1973, pp. 165-197.
31. P. A. Morrison. Movers and Stayers: An Analysis Based on Two Longitudinal Data Files. Rand Corp. Paper P-4409, 1970.
32. A. Okabe. Formulation of the Intervening Opportunities Model for Housing Location Choice Behavior. Journal of Regional Science, Vol. 17, 1977, pp. 31-40.
33. T. Lam, B. Odell, and T. Tardiff. A Stochastic Model of Residential Mobility and Urban Structure. Proc., 1975 ICS Winter Simulation Conference, 1976, pp. 631-637.
34. T. N. Lam and T. J. Tardiff. Urban Structure and Transportation Requirements. Proc., 1978 ICS Summer Simulation Conference, 1978, in press.

*Publication of this paper sponsored by Committee on Passenger Travel Demand Forecasting and Committee on Traveler Behavior and Values.*

# Spatial Aggregation of Disaggregate Choice Models: Areawide Urban Travel Demand Sketch-Planning Model

Thawat Watanatada, International Bank for Reconstruction and Development, Washington, D.C.
Moshe E. Ben-Akiva, Center for Transportation Studies, Massachusetts Institute of Technology, Cambridge

**This paper describes an aggregate urban travel demand model designed for areawide transportation policy evaluation with limited preparation of input data and fast response times. It does not include supply models but it can be used by itself for travel demand predictions with exogenously specified transportation level-of-service changes or it can be incorporated in the framework of the TRANS model. The methodology is generally applicable to urban transportation sketch-planning situations in which large geographic units are used. Aggregation is performed over spatial travel alternatives and spatially distributed individuals to produce required aggregate travel demand forecasts. An efficient solution method for spatial aggregation was developed that employs mathematical functions, expressed in terms of coordinates in the urban space, to describe the spatial choice process and to represent the geographic distribution of behavioral units, spatial alternatives, level-of-service characteristics, and locational attributes. This allows the spatial aggregation problem to be solved efficiently, by integrating the travel demand models over the urbanized area. Monte Carlo simulation techniques are employed and the procedure entails (a) generation of a sample of representative households distributed over the urban area using available census data, (b) generation of a sample of destinations for each trip purpose for each household, (c) computation of travel demand forecasts for each household based on the sampled destinations using a system of disaggregate travel demand models, and (d) accumulation and expansion of disaggregate predictions to produce aggregate forecasts.**

This paper describes the methodology and application of an aggregate model of urban travel behavior. The model is designed to be applied at a high level of geographic aggregation—the entire urban area—for quick assessment of urban transportation policies. The underlying methodology is applicable to a wider range of sketch-planning analyses that are characterized by the use of readily available input data (for example, from the U.S. Census) and fast response times. These features are essential for a successful integration of technical analysis and transportation decision-making processes.

Typically, the travel demand models employed in existing sketch-planning packages offer little policy sensitivity and require separate calibration for different levels of spatial aggregation or zone sizes.

The basic premise of this study is that the use of disaggregate travel behavior models (1, 2, 3, 4) in sketch-planning applications with appropriate aggregation procedures would remove these shortcomings. Although disaggregate choice models have many advantages over conventional aggregate travel demand models in general, their distinct advantages from the sketch-planning standpoint are that once estimated they can be applied to any desired level of geographic aggregation and that they have the potential of being transferred from one urban area to another.

The aggregate demand model developed in this study represents an extreme level of spatial aggregation since it treats an entire urban area as a single analysis unit. It is suited to metropolitan transportation planning studies in which impacts on specific areas are not required. When incorporated into a complete supply-and-demand analysis system, it can be used to determine scale and composition of transportation investments on an areawide basis, involving approximate funding allocations to modes and facility types, to construction and maintenance expenditures, etc. It can also be used to analyze aggregate impacts of pricing and operating policies such as fuel price change, parking cost surcharge, and areawide transit improvements.

The model can be used in the framework of the multimodal national urban transportation policy planning model known as TRANS (5, 6, 7, 8), which is referred to as MIT-TRANS in this paper.

## SPATIAL AGGREGATION OF DISAGGREGATE TRAVEL DEMAND MODELS

Aggregate forecasting cannot, in general, be performed by substitution of average values of the independent variables in a disaggregate model. Therefore, aggregate forecasting requires the application of an aggregation procedure that employs information about the distributions of the variables (9). An efficient aggregation procedure is particularly important in sketch-planning models designed for large spatial analysis units (10). This section describes the basic concepts and methods of aggregating disaggregate travel demand models over individuals and spatially distributed alternatives.

### Basic Definitions

Consider an origin zone as a group of individual behavioral units and a destination zone as a group of elemental spatial alternatives. The elemental spatial alternatives—such as housing units (in the choice of residential location) and jobs (in the choice of work place)—are assumed to be mutually exclusive in that one and only one of the available alternatives will be chosen.

Consider a prediction of the number of trips (given purpose) from an origin i to a destination j (both i and j are equal to the entire urbanized area as used for the level of aggregation in the MIT-TRANS model). First, for each individual t in the origin i predict the probability that the individual will choose each spatial alternative $\epsilon$ in destination j. Denote this probability $P_t(s)$. Then sum the probability for all spatial alternatives in the destination to produce the probabilities for all spatial alternatives in the destination and call this step aggregation of alternatives. Thus

$$P_t(j) = \sum_{s \epsilon j} P_t(s) \tag{1}$$

Finally, sum the probabilities for all individuals t in origin i to get the expected number of trips from origin i to destination j and call this step aggregation of individuals.

$$T_{ij} = \sum_{t \epsilon i} P_t(j) = \sum_{t \epsilon i} \sum_{s \epsilon j} P_t(s) \tag{2}$$

To illustrate the relationship between spatial and nonspatial alternatives, let $P_t(m \mid s)$ be the probability of trip maker t choosing mode m given that the trip maker has chosen spatial alternative s. By similar argument, the expected number of trips from origin i to destination j by mode m is given by

$$T_{ijm} = \sum_{t \epsilon i} \sum_{s \epsilon j} P_t(m \mid s) P_t(s) \tag{3}$$

The above example illustrates the basic idea of spatial aggregation that consists of two basic steps, the aggregation of spatial alternatives for each individual and the aggregation of spatially distributed individuals. These steps are conceptually applicable to any desired traffic zone size and to both intrazonal and interzonal trips.

### Spatial Aggregation Using Continuous Functions

Obviously, the discrete summation form used in the above example is too microscopic for actual prediction, since complete enumeration, as the example implies, would require astronomical amounts of data and computation. Furthermore, even if computational costs were not a barrier, it is still infeasible to describe the detailed characteristics of each spatial alternative and behavioral unit. In order to develop an operational spatial aggregation procedure, some degree of abstraction of spatial alternatives and behavioral units is necessary.

There are many possible ways to represent spatial distributions, depending on the level of detail desired. To generalize the definition of spatial aggregation, assume that mathematical functions expressed in terms of two-dimensional coordinates represent the spatial distributions. This type of representation can be used as a basis for comparing different spatial aggregation methods (9, 11, 12, 13, 14, 15).

Suppose that the attributes of spatial alternatives and the distributions of spatial alternatives and behavioral units could be expressed in terms of coordinates of the urban space. Define a spatial choice density function, denoted $G_k(p, q \mid x, y)$, as the probability of a behavioral unit type k located at point (x, y) choosing a spatial alternative located at point (p, q) for a specific purpose such as shopping destination (16). This is a unique surface for individual type k located at (x, y) that is also a function of the distribution of attributes of spatial opportunities and their transportation level of service for origin point (x, y) and socioeconomic variables of individual type k. Define spatial distribution functions for spatial alternatives and behavioral units as M(p, q) equals the number of elemental spatial alternatives per unit area at point (p, q), and $H_k(x, y)$ equals the number of behavioral units type k per unit area at point (x, y). The number of trips from zone i to zone j can now be derived as (a) aggregation over spatial alternatives to obtain the probability of behavioral unit type k located at (x, y) choosing an alternative in zone j:

$$P_k(j \mid x,y) = \iint_{\text{zone } j} G_k(p, q \mid x, y)\, M(p, q)\, dp\,dq \tag{4}$$

or (b) aggregation over behavioral units to obtain the expected number of behavioral units type k located in zone i who travel to zone j:

$$T_{kij} = \iint_{\text{zone } i} P_k(j \mid x, y)\, H_k(x, y)\, dx\,dy \tag{5}$$

The total number of trips is

$$T_{ij} = \sum_k T_{kij}$$

$$= \sum_k \iint_{\text{zone } i} \iint_{\text{zone } j} G_k(p, q \mid x, y)\, M(p, q)\, H_k(x, y)\, dp\,dq\,dx\,dy \tag{6}$$

It is also possible to repeat the above steps to derive other travel demand predictions. For example, let D(p, q; x, y) be the distance traveled between points (p, q) and (x, y). Then, the expected kilometers of travel for zone pair (i, j) is given by

$$MT_{ij} = \sum_k \iint_{\text{zone } i} \iint_{\text{zone } j} D(p, q; x, y) G_k(p, q \mid x, y) M(p, q) \times H_k(x, y)\, dp\,dq\,dx\,dy \tag{7}$$

## Numerical Techniques for Spatial Aggregation

The spatial choice density function can be expressed in terms of spatial coordinates via the distribution over space of the independent variables of a spatial choice model. The independent variables that enter the utility functions of choice models of travel behavior are

$L(p, q;x, y)$ = transportation level-of-service attributes by different modes, times, and facilities between origin point $(x, y)$ and destination point $(p, q)$;
$A(p, q)$ = location attributes, or attraction variables, of the relevant elemental spatial alternatives at point $(p, q)$; and
$S_k$ = socioeconomic characteristics of an individual of type k.

Thus, the required input data include spatial distributions for transportation level of service variables (L), locational attributes (A), and spatial density functions for elemental spatial alternatives (M) and behavioral units ($H_k$) of different types.

There are two broad approaches in which the spatial distribution of the input data can be applied to carry the spatial aggregation—a direct and an indirect approach.

### Direct Approach

The mathematical functions $L(p, q;x, y)$, $A(p, q)$, $M(p, q)$, and $H_k(x, y)$ are expressed explicitly in terms of the coordinates $(p, q)$ and $(x, y)$. Furthermore, the spatial integration is also carried directly.

Define the following probability density function:

$$PDF_{kij}(p, q;x,y) = \begin{cases} \dfrac{M(p,q)H_k(x,y)}{M_jH_{ki}} & \text{for } (x,y) \in \text{zone } i \\ & \text{for } (p,q) \in \text{zone } j \\ 0, \quad \text{otherwise} \end{cases} \tag{8}$$

where

$M_j$ = the number of elemental alternatives in zone j, or
$M_j = \int\int_j M(p, q)dpdq$; and
$H_{ki}$ = the number of behavioral units type k in zone i, or
$H_{ki} = \int\int_i H_k(x, y)dxdy$.

The integral for the number of trips by individuals type k from i to j, as an example, can now be rewritten as

$$T_{kij} = M_jH_{ki} \int\int_i \int\int_j C_k(p, q \mid x, y) PDF_{kij}(p, y;x, y)\, dpdqdxdy$$
$$= M_jH_{ki} \cdot \bar{G}_{kij} \tag{9}$$

where $\bar{G}_{kij}$ is the expected value of the spatial choice density function for individuals type k in zone i and elemental spatial alternatives in zone j. Thus, the expectation of the spatial choice density function is taken over the distribution of $(p, q)$ and $(x, y)$ defined by function $PDF_{kij}$.

This is the approach taken in the development of the MIT-TRANS model. Two broad classes of numerical integration methods are possible: mechanical or approximate quadrature techniques such as described in Davis and Rabinowitz (17) and Monte Carlo simulation techniques such as described in Hammersley and Handscomb (18).

If the functions $G_k(p, q;x, y)$, $L(p, q;x, y)$, $A(p, q)$, $M(p,q)$, and $H_k(x, y)$ are relatively well behaved (e.g., having a continuous first derivative) and if zones i and j assume a simple shape (e.g., a circle or rectangle), mechanical quadrature techniques seem appropriate. Otherwise, Monte Carlo simulation seems to be the only feasible approach.

In the MIT-TRANS model, because of the difficulty in setting the bounds of the integrals for the complex shape of the urban area, mechanical quadrature techniques were ruled out as infeasible, and the Monte Carlo approach was chosen. Although Monte Carlo methods are generally less accurate than mechanical quadrature methods, they are appropriate to travel demand forecasting applications because great precision is not required and we are interested in predicting changes.

In urban transportation planning applications, the Monte Carlo approval offers the following advantages:

1. No aggregation bias (in the context defined by Koppelman, 1975) is produced by Monte Carlo simulation;
2. Forecast error measures are available immediately as a by-product of the Monte Carlo simulation process;
3. Errors in Monte Carlo forecasts can be expressed as a function of sample size, which is directly proportional to computation effort; hence, the errors can be parametrically controlled by making a direct trade-off between accuracy and cost;
4. The Monte Carlo approach can be applied to any type of mathematical representation; and
5. It is possible to stratify Monte Carlo forecasts by socioeconomic or other groups. The prediction errors for each group can be controlled separately.

The major disadvantage of the Monte Carlo approach is that, for a given Monte Carlo procedure, the magnitude of random error is inversely proportional to the square root of sample size. Thus, to reduce the magnitude of error by one-half, the sample size must be quadrupled. Although this weakness can be serious in applications where a high degree of accuracy is required, it is not so in travel demand forecasting applications, since relative errors in the range of 10-20 percent or even greater are normally quite acceptable for decision-making purposes. Furthermore, a number of techniques, such as the stratified sampling and importance sampling used in the MIT-TRANS model, can be employed to substantially reduce error without increasing the sample size.

### Indirect Approach

The distribution of the coordinates $(p, q)$ and $(x, y)$ is first transformed into that of the level-of-service and attraction variables. Integration is then performed over the distribution of the latter variables.

The spatial choice density function is a function of the level-of-service variables (L) and the locational attributes (A), which are, in turn, functions of $(p, q)$ and $(x, y)$. Therefore, $\bar{G}_{kij}$ can also be expressed as the expectation of $G_k(L, A)$ over distribution of L and A, $PDF_{kij}(L, A)$:

$$\bar{G}_{kij} = \int_A \int_L G_k(L,A)\, PDF_{kij}(L,A)\, dL\, dA \tag{10}$$

Therefore, the key problem is to find the distribution of

L and A. Conceptually, if the density functions for spatial alternatives and behavioral units by type are known, the probability density function of the variables L and A, $PDF_{kij}(L, A)$, can be derived, either analytically or empirically.

In practice, however, the probability density function can be derived in closed form only under highly simplifying assumptions. For example, in the case of intrazonal trips (zones i - j), if we assume that the spatial alternatives and behavioral units are uniformly distributed in a circular zone and that the travel time is proportional to the airline distance between (p, q) and (x, y), then a closed-form expression exists for the probability density function of travel time, as derived by Kendall and Moran (19).

Generally, it appears feasible in practice to obtain only the first few moments of the distribution of the variables L and A. Parametric models can be developed empirically to express the distribution moments (e.g., the means and variances of travel times and costs) in terms of the zone configurations (e.g., zone areas), the transportation supply and level-of-service characteristics, the traffic volumes, and the land-use distribution pattern.

Two different efforts have previously been made to develop such parametric models. However, neither used spatial choice probabilities explicitly as the basis for model development. In modal split prediction, Dunbar (20) directly assumed that the distribution of trip distance between a zone pair has a trapezoidal shape and is parametrically related to the zone sizes. From this assumption and by assuming that the travel speed and cost per unit distance were constant for a given zone pair, Dunbar was able to obtain the means and variances of travel time and cost expressed in terms of the zone areas.

For the access portion of a trip, Talvitie and others (21, 22) conducted a large number of simulation experiments in which the zone configurations, the transportation supply characteristics, and the trip-end densities were varied as independent variables. With the simulation results, multiple regression was used to relate the means and variances of access travel times and costs as parametric functions of the above independent variables.

The distribution required to carry out the aggregation over alternatives and individuals could be represented in several different forms (9):

1. Parametric distribution functions with analytical or numerical (including Monte Carlo) integration techniques (11, 12);
2. Moments of distribution with the statistical differentials method that expresses the aggregate quantity in terms of moments of the distributions using a Taylor series expansion (11) (this approach proved to be unstable and is not employed in this study but is included for completeness);
3. Classification (or categorical representation of the distributions) with each class, which is a group of individuals or a group of alternatives, represented by its average values and the aggregate quantity, and is a weighted average of the choice probabilities for the classes; if overall average values are used for the independent variables—a one-class discrete distribution—it becomes the so-called "naive method" that ignores the aggregation problem entirely (9); and
4. Sample enumeration where aggregation is carried by summation of choice probabilities for a sample of individuals and a sample of elemental or homogeneous groups of spatial alternatives for each individual (this approach is identical to Monte Carlo integration except that the samples are drawn from actual observed data instead of from parametric distributions).

In the MIT-TRANS model a sample enumeration approach is used for the socioeconomic characteristics of households. The spatial distribution of behavioral units $H_k(x, y)$ is represented by classification and parametric functions within each class. All other distribution information is represented by parametric functions in terms of coordinates.

In conventional urban transportation model systems, the basic method used is classification. The population and spatial alternatives are classified into traffic zones with size varying according to the urban area and the purpose of the analysis. For each zone only average values are generally given.

The short-range travel demand prediction system developed by Ben-Akiva and Atherton (23) employs a sample enumeration approach to represent behavioral units, and a sample of aggregate traffic zones is used to represent destinations.

The method of statistical differentials was developed by Talvitie (11) and applied by Difiglio and Reed (24), Liou (25), and Dunbar (20).

## BASIC OPERATIONS OF THE MIT-TRANS MODEL

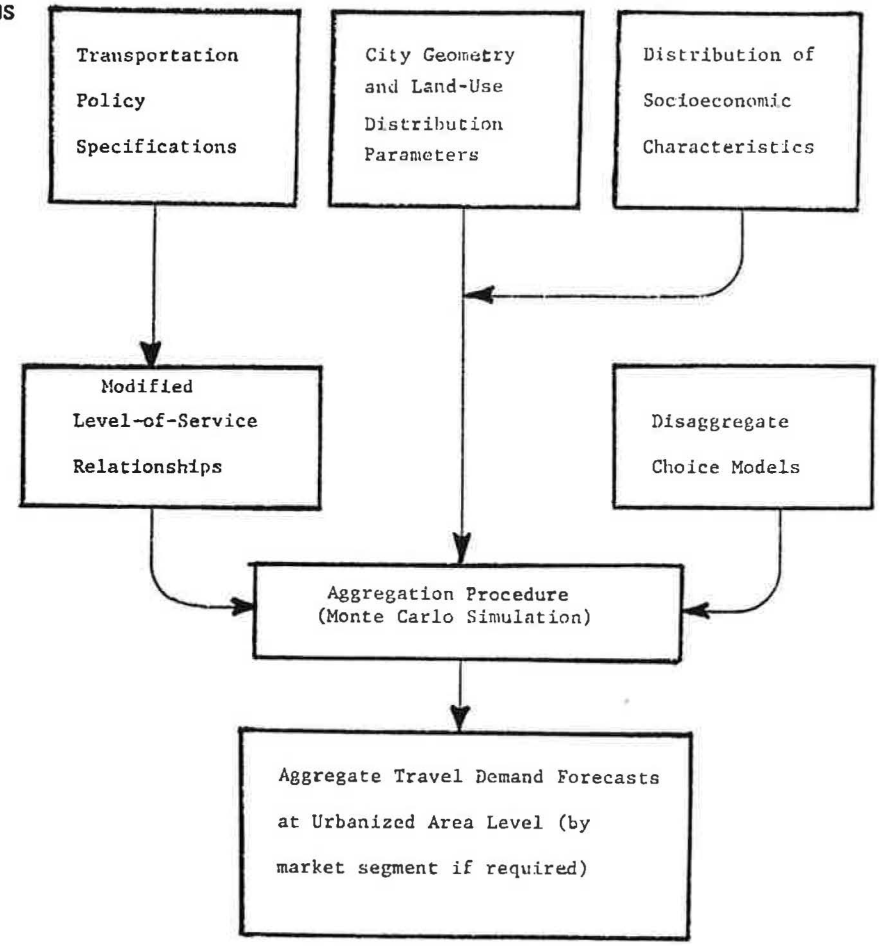
MIT-TRANS represents an extreme form of test for the feasibility and validity of the spatial aggregation methodology developed in this study, since it treats an entire urban area as a single traffic zone with almost all trips being internal. The model is based on the application of Monte Carlo simulation using synthetic socioeconomic and land-use distribution data to forecast trip generation, distribution, and modal split, with a system of disaggregate travel demand models. There are seven disaggregate models for both work and nonwork trips that have been linked together; i.e., outputs from one model become inputs to lower-hierarchy models (15). Examples of predictions are the number of trips made, mode shares, person-kilometers of travel, vehicle-kilometers, average vehicle occupancy rates for work and nonwork trips, number of automobiles per family, and so on. These predictions are policy sensitive as reflected in the elastic travel demand models for the choices of work place, auto ownership, mode to work, nonwork travel frequency, destination, and mode.

It should be noted that the MIT-TRANS model in its present form represents only the demand component of the TRANS overall policy evaluation package, which also includes the supply component and evaluation procedures. Future extensions of the MIT-TRANS model will include the development of network abstract transportation supply and traffic assignment models and the integration of these models and the aggregation procedure into an equilibrium framework. In lieu of a complete supply-demand equilibrium framework, a set of level-of-service relationships describing spatial distribution of the equilibrium conditions of an existing transportation system with externally specified parameters is being used in the current MIT-TRANS model.

The existing MIT-TRANS model can be employed to analyze a broad range of areawide transportation operating and pricing options—those policies that are not expected to significantly alter congestion on the transportation system. The results of some transportation policy alternatives are reported in the summary of empirical tests.

As summarized schematically in Figure 1, the operations of the MIT-TRANS model require three sets of inputs: (a) the aggregate city geometry and land-use

Figure 1. Basic operations of the MIT-TRANS model.




distribution parameters, (b) the urban area's socioeconomic characteristics, and (c) the specifications of a transportation policy alternative. These policy specifications are used to modify the level-of-service relationships that have been calibrated for the base conditions. The aggregation procedure, a Monte Carlo simulation, operates on these inputs, the disaggregate choice models, and the modified level-of-service relationship to produce aggregate travel demand forecasts for the urban area. The forecasts can be disaggregated by market segment, such as by income group.

In the context of the spatial aggregation concept, the MIT-TRANS model uses the direct approach to the integration problem. The urban area is modeled as a quasi-circular shape with the origins (home ends of trips) and destinations (nonhome ends) defined by sets of coordinates $(R, \lambda)$, and $(r, \phi)$ or $(L, \Theta | R, \lambda)$, respectively, as depicted in Figure 2.

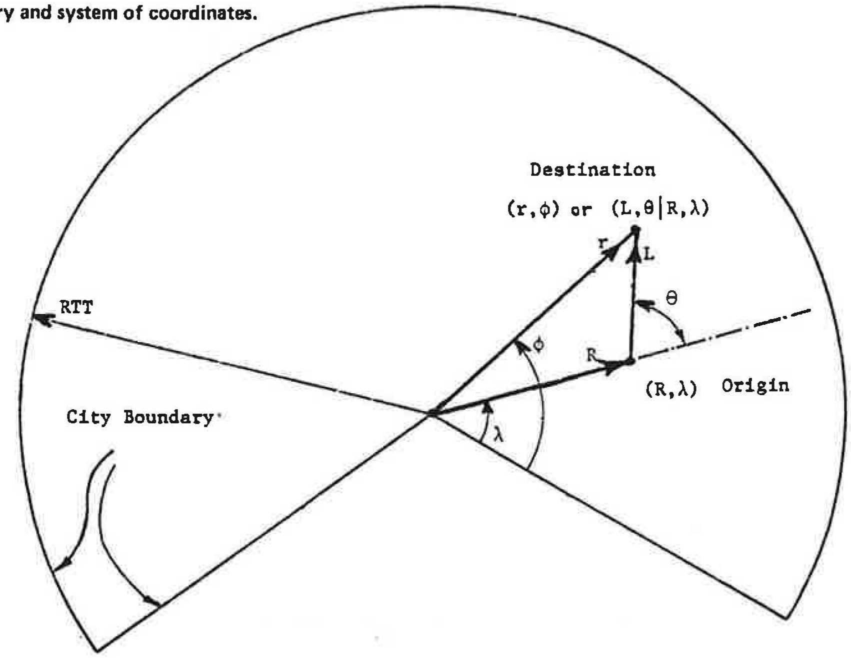
For each of three income classes the household density function is assumed to have a negative exponential shape. The spatial alternatives—jobs, shopping destinations, and social recreational facilities—are also represented by negative exponential employment density functions and functions describing location characteristics. The parameters of these density functions can be easily obtained from total counts of population and employment for an inner ring and the entire urbanized area. The transportation level-of-service functions by mode and time of day are expressed in terms of trip geometry variables, which are, in turn, functions of the coordinates of the trip ends.

MIT-TRANS also includes a procedure similar to the one used by Duguay, Jung, and McFadden (26) to obtain the distribution of socioeconomic characteristics of the urban area population by generating a sample of households from available data. The procedure is based on a sample of disaggregate observations from the U.S. Census public-use sample, or any other household survey, and available aggregate data from surveys or published sources for past years or from forecasts or staged scenarios for future years.

Monte Carlo simulation techniques are employed in all steps of the aggregation process: aggregation of spatial alternatives for a behavioral unit and aggregation of spatially distributed behavioral units. The operations of the Monte Carlo aggregation procedure include the following basic steps:

1. Determining household sample size,
2. Generating sample of households for forecast year [each household characterized by $(L, \Theta)$ location and a set of socioeconomic attributes],
3. Determining sample size of spatial alternatives by purpose,
4. Generating sample of spatial alternatives by purpose for each household in the sample [each destination defined by $(r, \phi)$ coordinates],
5. Modifying appropriate attributes of the alternatives for policy analysis,
6. Applying linked demand models for each household in the sample,
7. Expanding sample forecasts to population market segments, and
8. Comparing forecasts against base case for policy analysis.

Figure 2. City geometry and system of coordinates.




SUMMARY OF EMPIRICAL TESTS

The MIT-TRANS model was programmed in Fortran for an IBM370/168 computer. It requires about 0.6 min of computer-use time per policy run with a standard error of about 1 percent of predicted average passenger kilometers of travel. This demonstrates the computational feasibility of the Monte Carlo simulation approach for urban transportation sketch-planning applications.

The model was calibrated for the 1968 Metropolitan Washington, D.C., area and was then used to forecast 1975 conditions as a validation test. Between 1968 and 1975 there were substantial changes in this area.

The main differences in land use included a large increase in the fraction of high-income households and a corresponding decline in low-income households, plus a large population growth in the suburbs together with a sharp decline in population in the inner ring. Household income increased as did the number of workers, while the average household size declined. The transit system experienced almost no change at all, while, for drivers, auto ownership cost decreased 15 percent and auto running cost increased 9 percent in real terms.

The combination of population growth in the suburbs and almost no change in transit coverage meant fewer households had available transit in 1975. The major changes in travel behavior between 1968 and 1975 included increased auto ownership, increased vehicle-kilometers of travel per household, and decreased transit patronage. All the forecast changes given below agree with these trends.

| Household Transit Factor | Observed Changes (%) | Predicted Changes (%) |
|---|---|---|
| Auto ownership per household | | |
| 0 autos | -23.2 | -23.6 |
| 1 auto | -14.0 | 4.8 |
| 2 autos | 34.1 | 11.6 |
| Average autos per household | 15.1 | 8.9 |
| Average transit trips per household | -16.7 | -14.5 |
| Average auto-kilometers per household | 15.9 | 16.1 |

Apart from Washington, D.C., the model was also calibrated for the Minneapolis-St. Paul area, which has two central business districts. The elasticities produced by the model for both the Washington, D.C., and the Twin Cities areas are comparable to before-and-after empirical evidence and forecasts obtained from other studies.

The MIT-TRANS model's range of applications and policy sensitivity were tested in the following areawide transportation pricing and operating policies.

1. Transit policies: 50 percent fare increase, free transit, 10 percent transit coverage reduction, 100 percent transit coverage;
2. Auto policies: doubling running cost (in testing auto pricing policies the existing value of kilometers per liter is taken as a constant), annual $400 tax on ownership;
3. Carpool policy: carpool-only highway lanes and preferential parking sufficient to achieve a 20 percent reduction in out-of-vehicle travel time and a 30 percent reduction in in-vehicle travel time.

Detailed results compiled in Watanatada and Ben-Akiva (15) show that auto ownership is affected most by policies that either reduce the need for an auto or make ownership more expensive. It is less affected by policies shifting the relative travel costs by transit and auto.

Auto-kilometers traveled is predicted to change by -13 percent for Washington and -11 percent for the Twin Cities because of the doubling of the price of gasoline. Smaller impacts (-7 percent for Washington) are predicted from providing transit within a kilometer of every household and business in the urban area. No other policy—including carpooling incentives—results in more than a 3 percent reduction in auto-kilometers traveled. Both the number of trips and the mode shares of transit can be increased by either pricing or coverage policy. Reducing transit coverage by 10 percent has approximately half the impact of a 50 percent fare increase. Free transit increases transit trips more than complete coverage, where the impact on nonwork trips is very different, nonwork trips being much more price sensitive.

In addition to the policies shown and discussed, other policy options can be analyzed by using the model. Parking costs may be varied throughout the city. In-vehicle and out-of-vehicle times for all three modes may be modified for both peak and nonpeak conditions.

Several Monte Carlo sampling experiments were conducted to investigate the statistical properties of the model. It was found empirically that a small sample of destinations results in minimal bias and optimal efficiency. The results of sensitivity tests of major input parameters show the importance of the distribution of transit route coverage.

The empirical results have led to the basic conclusion supporting the applicability of disaggregate travel demand models and Monte Carlo aggregation for sketch planning. The travel demand forecasting methodology proposed operates with readily available aggregate input data while still maintaining the full degree of policy sensitivity available in recently developed systems of disaggregate models. The most important future extensions of the methodology are the incorporation of supply and traffic assignment models (15) and the development of a version of MIT-TRANS for multiple zones of varying sizes.

## ACKNOWLEDGMENTS

The work reported in this paper was performed at the MIT Center for Transportation Studies as part of the project on the development of an aggregate model of urbanized area travel behavior for the Office of the Secretary of Transportation and the Federal Highway Administration. We acknowledge the significant contributions of Edward Weiner, Helen Doo, and David Gendell. Josef Bain, Peter Furth, John Nordin, and Patrick O'Keefe have significantly contributed to the conduct of this study. In addition, we benefited from the advice of Frank Koppelman, Steve Lerman, Charles Manski, and Paul Roberts.

## REFERENCES

1. T. Domencich and D. McFadden. Urban Travel Demand: A Behavioral Analysis. North Holland, Amsterdam, 1975.
2. M. G. Richards and M. E. Ben-Akiva. A Disaggregate Travel Demand Model. Saxon House, D.C. Heath, England, 1975.
3. M. Ben-Akiva, S. R. Lerman, and M. L. Manheim. Disaggregate Models: An Overview of Some Recent Research Results and Practical Applications. In Transportation Models, Proc., Transportation Research Council, 1976.
4. B. D. Spear. A Study of Individual Choice Models: Applications of New Travel Demand Forecasting Techniques to Transportation Planning. Office of Highway Planning, U.S. Department of Transportation, Federal Highway Administration, 1977.
5. H. Kassoff and D. S. Gendell. An Approach to Multiregional Urban Transportation Policy Planning. HRB, Highway Research Record 348, 1971, pp. 76-93.
6. E. Weiner, H. Kassoff, and D. S. Gendell. Multimodal National Urban Transportation Policy Planning Model. HRB, Highway Research Record 458, 1973, pp. 31-34.
7. D. S. Gendell, J. J. Hillegass, and H. Kassoff. Effects of Varying Policies and Assumptions on National Highway Requirements. HRB, Highway Research Record 458, 1973, pp. 21-30.
8. 1972 National Highway Needs Report. U.S. Department of Transportation, Federal Highway Administration, 1972.
9. F. S. Koppelman. Travel Prediction With Models of Individual Travel Behavior. Transportation Systems Division, Department of Civil Engineering, MIT, Cambridge, Ph.D. thesis, 1975.
10. T. Watanatada. Application of Disaggregate Choice Models to Urban Transportation Sketch Planning. Transportation Systems Division, Department of Civil Engineering, MIT, Cambridge, Sc.D. thesis, Apr. 1977.
11. A. Talvitie. Aggregate Travel Demand Analysis With Disaggregate or Aggregate Travel Demand Models. Transportation Research Forum Proceedings, Vol. 14, No. 1, 1973.
12. R. B. Westin. Prediction From Binary Choice Models. Journal of Econometrics, Apr. 1974.
13. D. McFadden and F. Reid. Aggregate Travel Demand Forecasting From Disaggregated Behavioral Models. TRB, Transportation Research Record 534, 1975, pp. 24-37.
14. U. Landau. Sketch Planning Models in Transportation Systems. Transportation Systems Division, Department of Civil Engineering, MIT, Cambridge, Ph.D. thesis, 1976.
15. T. Watanatada and M. Ben-Akiva. Development of an Aggregate Model of Urbanized Area Travel Behavior. Rept. prepared for U.S. Department of Transportation, Center for Transportation Studies, MIT, Cambridge, July 1977.
16. D. McFadden. The Mathematical Theory of Demand Models. In Behavioral Travel-Demand Models (Stopher and Meyburg, eds.), Lexington Books, Lexington, MA, 1976.
17. P. J. Davis and P. Rabinowitz. Methods of Numerical Integration. Academic Press, NY, 1975.
18. J. M. Hammersley and D. C. Handscomb. Monte Carlo Methods. Methuen, London, 1965.
19. M. G. Kendall and P. A. P. Moran. Geometrical Probability. Charles Griffin, London, 1963.
20. F. C. Dunbar. Quick Policy Evaluation With Behavioral Demand Models. TRB, Transportation Research Record 610, 1976, pp. 47-49.
21. A. Talvitie and N. Hilsen. An Aggregate Access Supply Model. Transportation Research Forum Proceedings, Vol. 15, No. 1, 1974, pp. 336-346.
22. A. Talvitie and T. Leung. Parametric Access Network Model. TRB, Transportation Research Record 592, 1976, pp. 45-49.
23. M. Ben-Akiva and T. Atherton. Methodology for Short Range Travel Demand Predictions. Journal of Transport Economics and Policy, Vol. 9, No. 3, Sept. 1977.
24. C. Difiglio and M. F. Reed, Jr. Transit Sketch Planning Procedures. TRB, Transportation Research Record 569, 1975, pp. 1-11.
25. P. S. Liou, G. S. Cohen, and D. Hartgen. Application of Disaggregate Modal-Choice Models Travel Demand Forecasting for Urban Transit Systems. TRB, Transportation Research Record 534, 1975, pp. 52-62.
26. G. Duguay, W. Jung, and D. McFadden. SYNSAM: A Methodology for Synthesizing Household Transportation Survey Data. Travel Demand Forecasting Project, Institute of Transportation Studies, Univ. of California, Berkeley, Working Paper No. 7618, Sept. 1976.

*Publication of this paper sponsored by Committee on Passenger Travel Demand Forecasting and Committee on Traveler Behavior and Values.*

# Aggregate Prediction With Disaggregate Models: Behavior of the Aggregation Bias

Uzi Landau, Transportation Research Institute, Technion-Israel Institute of Technology, Haifa

**Disaggregate travel demand models are being increasingly applied to predictions of aggregate demand. This is usually done by applying these models to zonal aggregated data, but it causes a bias (the aggregation bias) in the predictions obtained. The purpose of this paper is to study empirically the characteristics of the bias for a variety of conditions. The empirical analysis focuses on the bias in aggregate predictions of mode choice for work trips in Washington, D.C. Two main factors of bias are identified as the geographic aggregation level and the level of detail by which the distribution of explanatory variables is represented. Both the magnitude and the behavior of the aggregation bias are examined for a wide range of geographic aggregation levels, for several approximate representations of the distribution of explanatory variables, and for two different transportation options. The simplest aggregate prediction method uses average zonal variable values in the disaggregate model. The results of this study indicate that, by applying this method, substantially biased predictions may result. Applying more accurate distribution representation reduces this bias significantly but does not ensure its complete elimination. A residual bias of significant magnitude still remained in many of the situations examined. The implication is that sophisticated methods for bias reduction should be developed in order to make aggregate predictions with disaggregate models a more reliable analysis tool.**

The prediction of aggregate travel behavior is an indispensable element of the transportation planning process. Over the past 20 years demand for travel has been estimated by aggregate models, but, due to their limitations (1, 2), more research efforts have been directed to the development of methods for applying disaggregate models of individual choice behavior to aggregate travel predictions. These models have a number of advantages over the aggregate models: they are more policy sensitive; they require relatively little data; and they are more likely to be transferable (3). A detailed analysis of the disaggregate models is given by Ben-Akiva (4) and Charles River Associates (1).

As is the case with all models, certain errors are also involved with the disaggregate model aggregate predictions. Koppelman (5, 6, 7) presents a comprehensive analysis of the sources of these errors. This paper focuses on the behavior and magnitude of a specific source of error—the aggregation bias—which usually appears in aggregate predictions made by disaggregate models.

## APPROXIMATE AGGREGATION PROCEDURES

The process of predicting aggregate behavior with disaggregate models consists of three components (7): (a) disaggregate choice model; (b) representation of the distribution of explanatory variables of this disaggregate model; and (c) aggregation procedure, which operates on the above two components to obtain the aggregate prediction.

Several aggregation procedures have been discussed in the literature (6, 8, 9, 10, 11); some are more accurate than others. The more accurate ones apply the disaggregate models directly to disaggregate data or to the exact joint probability distribution of the explanatory variables. Other procedures that apply the disaggregate models to forms of aggregated data are less accurate (7). However, the more accurate procedures are usually less practical, while the less accurate ones are more convenient and in fact more widely used (12, 13, 14).

The aggregation bias is one of several contributors to total aggregate prediction error, which, in certain situations discussed later, is of considerable magnitude. This paper concentrates on those characteristics of the bias associated with the more popular procedures such as estimating future aggregate demand by applying the disaggregate model in its exact functional form to zonal aggregate data.

## MAIN CAUSES OF THE AGGREGATION BIAS

Aggregation is grouping individuals into zones and representing them as one group with common characteristics. The data of the new zone system are an aggregate representation of the real underlying distribution of the data from the detailed (individual) level. There are several ways to represent the aggregate zonal data.

The most common way is the representation of the data with their means, such as average zonal income and mean travel time. However, every aggregate representation of the underlying detailed data results in a loss of information. The specific nature of data variation at the detailed level is lost in the process of aggregation.

If the disaggregate model is nonlinear, then this loss of information causes a bias in the predictions obtained that is known as the aggregation bias (7). The main factors that contribute to it are the following.

1. Geographic aggregation level (GAL): Aggregation of zones simply increases within-group variability in the zonal and interzonal distributions of the data. Going to higher levels of zone aggregation is identical to saying that the zonal and interzonal related data distributions have larger variances. Representing these distributions with few measures (traditionally, only one measure, the mean) results in a loss of some variational characteristics of the data at the considered aggregation level. Therefore, the more we aggregate, the more information we lose, and hence the greater the aggregation bias.
2. Distribution representation method (DRM): The representation of the underlying distribution of explanatory variables at the considered level of aggregation is termed here "distribution representation method" (DRM). The more we aggregate, the greater is the information loss and, consequently, the greater the aggregation error. However, given a certain aggregation level, the more accurate the distribution representation method, the smaller the loss and, hence, the better the aggregate predictions.

Ideally, we would like to represent the data by their multivariate joint distribution and apply summation-integration procedures (7). In this case, no information is lost, and zone aggregation would not cause a bias. However, establishing this joint distribution is an intractable problem. Therefore, less accurate methods are used to represent the data (6, 8). They are imperfect compared to the multivariate distribution representation, but they are a more accurate representation compared to the means. They bring about a reduction in the aggregation bias, but some error still remains.

The traditional DRM in aggregate analysis is the use of weighted averages. That is, the aggregate prediction is made by making each variable in the disaggregate model equal to the weighted average of its disaggregate values.

Classification procedures are more accurate DRMs than the weighted averages method. They consist of approximating the distribution of a variable with a histogram of few classes. The expected demand for each class is estimated by using average values of all variables for this class. The overall demand is determined as a weighted sum of the individual classes.

## ANALYSIS OF BIAS IN AGGREGATE PREDICTION: PURPOSE AND METHODOLOGY

The purpose of the analysis is to study empirically the behavior of the aggregation bias, specifically to find

1. How the bias changes over a wide range of different GALs,
2. What the magnitude of the bias caused by using weighted average procedures is,
3. How well approximate DRMs of the explanatory variables reduce the bias, and
4. Whether different transportation alternatives produce similar (or different) biases.

A method for identifying the value of the aggregation bias was developed and tested in an applied prediction context. The method is illustrated by an empirical study of mode choice by work trip makers in the Washington, D.C., metropolitan area. The approach is to make multiple aggregate predictions of choice shares with a single disaggregate choice model for two different transportation options, for several GALs, and for several DRMs, each applied to all GALs and transportation options.

The predictions made with the disaggregate model by the complete enumeration procedure involve no bias (i.e., estimating expected shares by averaging the choice probabilities, calculated for each member of the population). They serve here as the no-bias reference level. The aggregation bias of the predictions obtained by applying the disaggregate model to a selected DRM, GAL, and transportation option, is determined by comparing these predictions with those obtained by complete enumeration.

## DEMAND MODEL, DATA, AND TRANSPORTATION OPTIONS

The demand model chosen for the analysis is of the N-dimensional logit form, developed by Peat, Marwick, Mitchell and Company (PMM) for San Diego (15). The model is developed to forecast central business district- (CBD-) oriented work trip makers' choice among three modes: transit passenger, automobile driver, and automobile passenger. The experiment focuses on the error in the aggregate prediction of transit share. The following variables appear in the model.

IN = household income,
TRNT = transit travel time, which is in-vehicle time plus transit transfer time,
TEXS = transit excess time, which is the walk to and from transit time plus first wait for transit,
FARE = transit fare,
HWTT = auto highway driving time,
COST = auto operating cost,
PARK = half of the auto parking cost, and
AEXS = auto excess time, which is the walk to and from the auto.

The transit share for individual t, $v_t$, is

$$v_t = [\exp(u_1)]/[\exp(u_1) + \exp(u_2 + G)] \quad (1)$$

where

$$u_1 = 1.1635 - 0.05625 \times \mathrm{TRNT} - 0.09157 \times \mathrm{TEXS} - 0.0106 \times \mathrm{FARE},$$
$$u_2 = -1.4809 - 0.05625 \times \mathrm{HWTT} - 0.09157 \times \mathrm{AEXS} - 0.01062\ (\mathrm{COST} + \mathrm{PARK}), \text{ and}$$
$$G = 2.952\ [1 - \exp(-0.035 \times \mathrm{IN})].$$

The model was reported to fit the San Diego data very well. Its predictions were also tested by PMM in the San Francisco and Boston transportation systems and were found to be satisfactory (15).

Three reasons underlie the choice of this particular model. First, it seemed desirable to test a realistic demand model. Similar models in terms of variables and value of parameters are likely to occur in many cases. Second, the model is sensitive to several level-of-service variables. Third, the model is nonlinear not only in its general structure but also in its utility function. Such a complex nonlinear form is potentially a major contributor to the aggregation bias (6).

The PMM model was applied to the Washington, D.C., data for predictions. The traffic corridor from the CBD northbound, through Silver Spring, Maryland, and beyond I-495 (the Capital Beltway) was chosen. It is a 144-traffic-zone segment of the entire metropolitan area (1207 traffic zones). This 144-zone system serves as our entire disaggregate population (144 × 144 pairs of transit share predictions).

Six superzone systems, each of a different GAL, were defined along the study corridor. The basis for the definition of a GAL in this experiment is the number of traffic zones in one superzone. The superzones were defined such that, for a given GAL, each had exactly the same number of traffic zones as the others. Consequently, the units by which the level of aggregation is quantitatively expressed are elementary (traffic) zones per superzone, in shorter notation EZ/SZ. The six aggregate systems have 72, 36, 18, 9, 3, and 1 superzones. The number of EZ/SZ in each is 2, 4, 8, 16, 48, and 144, respectively. The one-superzone system, for example, is simply the total study corridor, treated as one big zone.

Data files were prepared to represent the variables by their DRMs for the six GALs. The DRMs were

1. Weighted averages: that is, representation of the distribution of each variable by one measure (= one class), the mean;
2. Classification with 2, 3, and 4 classes: that is, approximating the distribution of a variable with a frequency histogram of 2, 3, and 4 classes (for example, in the two-class case the distribution is split at its median and is represented by the means of the two sec-

tions, each given a weight of ½; in the three-class case the distribution is divided at its tertiles and is represented by the means of the three classes, each given a weight of ⅓; the four-class approximation is analogous); and

3. Marginal distributions: that is, each variable is represented by its exact marginal frequency histogram.

The approximate distributions applied consisted of the representation of few variables (one, two, or three) with their marginal distribution or approximated frequency histogram (i.e., classification), while all the rest of the variables were at their means.

Two transport alternatives were chosen for the analysis. Both concerned the routing of the public transit vehicles. The first, alternative 1, consisted of the bus services, as provided in 1968, where transfers were made by individuals for certain origin-destination pairs. The second alternative, alternative 2, cut all the transfers for the same trip interchanges assuming 5 min/transfer.

The PMM disaggregate model was applied to 292 and 128 different combinations of GAL and DRM for alternatives 1 and 2, respectively. The behavior of the aggregation bias was examined for all these combinations. The numerical results illustrated in Figures 1, 2, 3, 4, and 5 provide a representative sample of those obtained in the entire study. A more detailed description of the analysis and results is given elsewhere (8).

## MEASURES OF THE AGGREGATION BIAS

The aggregation bias is expressed by the following three error measures: average interzonal percentage error, total area percentage error, and root-mean-square error.

### Average Interzonal Percentage Error

For average interzonal percentage error ($P_{AE}$), let $p_{ij}$ be the basic error measure in transit share prediction between superzones i and j, such that

$$p_{ij} = (V'_{ij} - V_{ij})/V_{ij} \tag{2}$$

where

$V_{ij}$ = unbiased reference level aggregate prediction for superzone pair (i, j), and
$V'_{ij}$ = aggregate prediction by approximate aggregation procedure for superzone pair (i, j).

Equation 2 expresses the magnitude of the bias as a proportion of the unbiased prediction.

The average level of the interzonal prediction error is calculated by averaging out all interzonal errors:

$$P_{AE} = \sum_{i,j} w_{ij}\, p_{ij} \tag{3}$$

where $w_{ij}$ is the weight of trips of superzone pair (i, j) with respect to all superzone pair trips.

### Total Area Percentage Error

The purpose of total area percentage error ($P_{total}$) is to capture the bias of the entire study area prediction for transit ridership. It is analogous to $P_{AE}$ and is defined by

$$P_{total} = (V'_{..} - V_{..})/V_{..} \tag{4}$$

where $V_{..}$ and $V'_{..}$ are the unbiased and approximate entire area aggregate predictions for transit share, respectively.

### Root-Mean-Square Error

Root-mean-square error (RMSE) is defined by

$$\mathrm{RMSE} = \left[\sum_{i,j} w_{ij}\,(V'_{ij} - V_{ij})^2\right]^{1/2} \tag{5}$$

## ANALYSIS OF RESULTS

Several characteristics of the aggregation bias were examined, as they varied for several GALs and DRMs. The bias was estimated by a variety of measures, all of which indicated similar bias behavior. Four more measures have been applied (8) with similar results. The empirical findings follow.

### Magnitude of the Aggregation Bias

The results obtained in this study show that the use of weighted averages in the disaggregate model causes significant bias. The values of the biases ranged from 16 to 40 percent error, depending on the GAL (Figures 1 and 3). These are much greater biases, compared to the 7 percent average bias reported by Koppelman (7).

This means that the application of the weighted averages method for aggregate prediction with disaggregate models may result in substantial biases, unless special procedures for bias reduction are applied.

The application of more accurate representation of the distribution of explanatory variables reduced the bias significantly, especially for high GALs. This finding agrees with Koppelman's conclusion about the performance of classification procedures. However, the residual bias in this study (13-35 percent) is substantially greater than the 1.4 percent average bias reported by Koppelman (7).

### Flattening of Bias With Higher Levels of Aggregation

Although the bias monotonically increases with aggregation, it does so at a decreasing rate until it reaches a certain point at which the rate almost does not change. This characteristic is clearly illustrated by Figures 1 and 3.

An interesting phenomenon is that the bias increases very rapidly at very low levels of aggregation. Moving from the original system of 144 traffic zones (GAL = 1 EZ/SZ) to a 72-superzone system (GAL = 2 EZ/SZ) and using the weighted averages method, an error jump of 15.8 percent is made. The next move to the 36-superzone system (GAL = 4 EZ/SZ) produces an additional error of only 2.8 percent, which continues to decrease for higher levels of aggregation.

This characteristic has an important practical implication. Aggregation will have to be pursued in many analysis situations. Since most of the error is already made for low GALs, it may be cost effective to go into much higher levels of aggregation. The additional small error is traded off for large savings in computational costs.

Figure 1. Aggregation bias for different GALs and DRMs for alternative 1.

| GAL EZ/SZ | 2 | 4 | 8 | 16 | 48 | 144 |
|---|---|---|---|---|---|---|
| WEIGHTED AVERAGES | | | | | | |
| $P_{total}$ | -15.8 | -18.6 | -21.6 | -23.4 | -26.6 | -40.4 |
| RMSE | 4.2 | 4.8 | 5.3 | 5.6 | 6.3 | 9.4 |
| $P_{AE}$ | -19.9 | -23.6 | -27.2 | -29.6 | -32.3 | -40.4 |
| TEXS BY MARGINAL DISTRIBUTION [(a)] | | | | | | |
| $P_{total}$ | -14.8 | -16.8 | -19.2 | -20.7 | -23.2 | -34.2 |
| RMSE | 3.9 | 4.3 | 4.7 | 5.0 | 5.5 | 8.0 |
| $P_{AE}$ | -18.0 | -20.7 | -23.5 | -25.4 | -27.7 | -34.2 |
| TRNT BY MARGINAL DISTRUBUTION [(a)] | | | | | | |
| $P_{total}$ | -14.6 | -16.2 | -17.7 | -18.6 | -19.3 | -23.1 |
| RMSE | 3.8 | 4.2 | 4.4 | 4.5 | 4.6 | 5.4 |
| $P_{AE}$ | -17.8 | -19.6 | -21.4 | -22.3 | -22.2 | -23.1 |
| TEXS AND TRNT BY MARGINAL DISTRIBUTION [(a)] | | | | | | |
| $P_{total}$ | -13.6 | -14.6 | -15.4 | -15.9 | -16.1 | -18.1 |
| RMSE | 3.6 | 3.8 | 3.9 | 3.9 | 3.9 | 4.2 |
| $P_{AE}$ | -15.8 | -16.6 | -17.5 | -18.0 | -17.5 | -18.1 |
| TEXS AND TRNT EACH BY 3 CLASSES [(a)] | | | | | | |
| $P_{total}$ | -14.1 | -15.4 | -16.5 | -17.1 | -17.9 | -21.4 |
| RMSE | 3.7 | 4.0 | 4.1 | 4.2 | 4.30 | 5.0 |
| $P_{AE}$ | -16.7 | -18.3 | -19.5 | -19.9 | -20.1 | -21.4 |

(a) All other variables are at their means.

Figure 2. Bias for two transportation alternatives in total area percentage error.

| GAL [(a)] / ALTERNATIVE | 2 | 4 | 8 | 16 | 48 | 144 |
|---|---|---|---|---|---|---|
| WEIGHTED AVERAGES | | | | | | |
| 1 | -15.8 | -18.6 | -21.6 | -23.4 | -26.6 | -40.4 |
| 2 | -14.7 | -16.8 | -18.8 | -20.0 | -22.6 | -33.9 |
| TEXS BY MARGINAL DISTRIBUTION [(b)] | | | | | | |
| 1 | -14.8 | -16.8 | -19.2 | -20.7 | -23.2 | -34.2 |
| 2 | -13.7 | -15.1 | -16.5 | -17.3 | -19.2 | -27.9 |
| TRNT BY MARGINAL DISTRIBUTION [(b)] | | | | | | |
| 1 | -14.6 | -16.2 | -17.7 | -18.6 | -19.3 | -23.1 |
| 2 | -14.0 | -15.4 | -16.5 | -17.0 | -17.9 | -21.3 |
| TEXS AND TRNT BY MARGINAL DISTRIBUTION [(b)] | | | | | | |
| 1 | -13.6 | -14.6 | -15.4 | -15.9 | -16.1 | -18.1 |
| 2 | -13.0 | -13.7 | -14.2 | -14.4 | -14.8 | -16.6 |
| TEXS AND TRNT BY 3 CLASS HISTOGRAM [(b)] | | | | | | |
| 1 | -14.1 | -15.4 | -16.5 | -17.1 | -17.9 | -21.4 |
| 2 | -13.4 | -14.4 | -15.1 | -15.3 | - | - |

(a) In EZ/SZ

(b) All other variables are at their means.

Figure 3. Aggregation bias for different GALs for alternative 1.

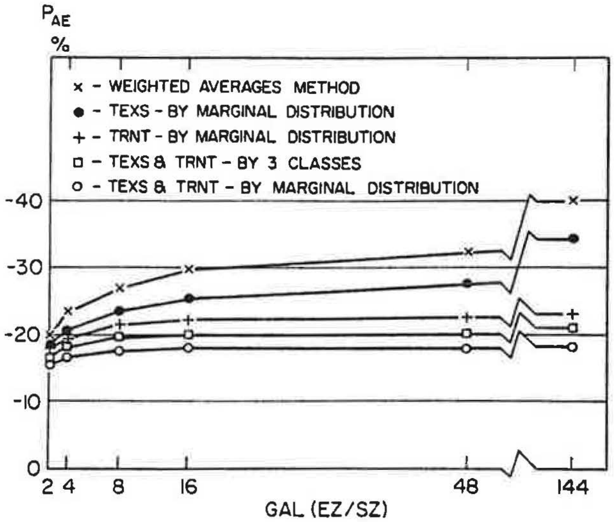
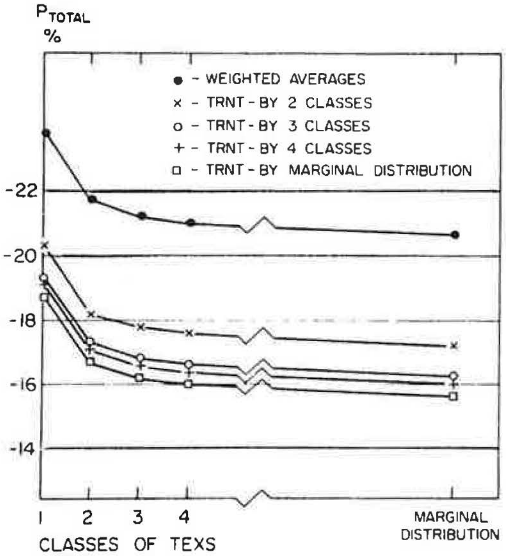



Figure 4. Areawide modal split percentage error for alternative 1.

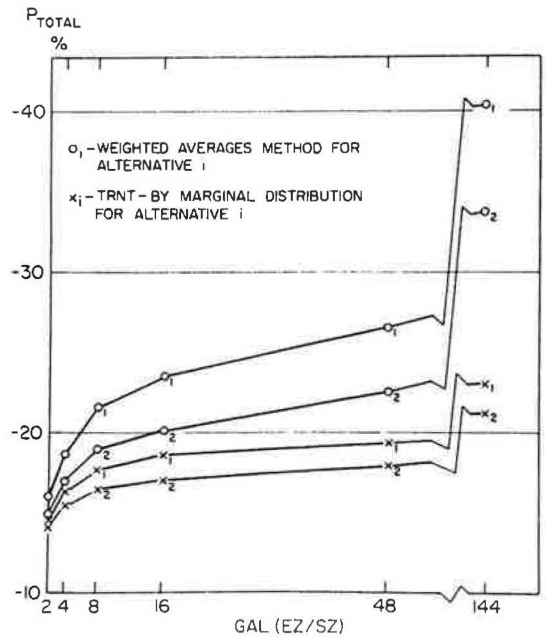



## Usefulness of More Accurate Variable Representation

Figure 4 illustrates the aggregation biases produced by transit excess time (TEXS) and travel time (TRNT), as they are represented by all possible combinations of pair classifications under the independence assumption (i.e., representing each by its marginal class histogram as if it were independent of the other variable). This figure indicates the pattern of diminishing marginal error with better representation. For every given representation of TRNT, the curve monotonically decreases as the accuracy of the representation of TEXS improves, but the rate of decrease is slower and levels off as its representation approaches the marginal distribution. These same results occur for all GALs, for all variables, and for all DRMs examined.

As we shift our attention to a pair representation of the variables, each with the same number of classes, the same behavior of diminishing marginal error continues to hold. There is a large decrease in error

Figure 5. Aggregation bias for two transportation alternatives.




(about 5.5 percent) from the weighted averages method to representing TEXS and TRNT with a two-class histogram for each; the next step (three-class histogram for each) improves the bias only by about 1.5 percent, and so on.

We have assumed that TEXS and TRNT are uncorrelated. Consequently, they were represented each by an approximated marginal distribution. The results, as expressed by Figure 4, do not contradict this assumption. The figure illustrates that, with every class added to represent any of the variables, (a) the overall error is reduced and (b) this reduction is made in quite a smooth manner, where the rate of reduction for each variable is unaffected by the DRM of the other variable.

As can be seen from Figure 4, the largest portion of the total bias that can potentially be corrected by representing a single variable with its marginal distribution is already achieved by a two-class representation (in our example, about two-thirds of the total error). In the case of TEXS, the two-class representation reduces the error by about 2 percent compared to 3 percent by marginal distribution, while the corresponding figures for TRNT are 3.5 and 5.2 percent. These results are consistent with those of Koppelman (6), who observed the same pattern for his model. As our attention shifts to pair representation, an analogous phenomenon occurs.

## Aggregation Bias and Transportation Alternatives

The aggregate prediction bias is a function of the transportation alternative under examination. As Figures 2 and 5 illustrate, the difference between the prediction bias of two transportation alternatives is not necessarily small, although the same model is used for both, and only one variable, TRNT (which expresses the change in policy), is changed.

In our example, the difference between the prediction biases increases with the level of aggregation for all DRMs, especially for the weighted averages method, and other DRMs that do not represent more accurately the variable indicating the change in policy.

SUMMARY

The specific conclusions summarizing the findings about the behavior of the aggregation bias are

1. The application of a nonlinear disaggregate demand model to aggregate data may result in substantially biased predictions;
2. A large aggregation bias is likely to be introduced even in very low GALs;
3. As aggregation increases, the bias increases monotonically but with a diminishing marginal rate;
4. A representation of an explanatory variable (which is uncorrelated with others), even if only with a small number of classes, reduces its error contribution substantially;
5. The bias reduction is generally greater if a number of variables are represented with few classes (two or three), than if only one or two of them are represented very accurately;
6. The aggregation bias is a function of the transportation alternative; and
7. If only some variables are represented more accurately, while other variables are represented by their means, then a residual aggregation bias of a significant magnitude may occur.

The practical implications of these findings are

1. Aggregate predictions with disaggregate models are not meaningful unless the aggregation biases are explicitly considered;
2. Since the aggregation bias is systematic and significant, as in our case study, it is important to correct it;
3. Classification methods applied to correcting the bias are not always efficient (in view of the possible high residual value of this bias of 13-34 percent, sophisticated bias correction methods should be developed and applied); and
4. If it is desired to test a variety of transport options and if the variable reflecting them is changed over a wide range, then it is important to represent more accurately the distribution of this variable.

Indeed, the empirical analysis has focused on examining a specific model structure in a specific transportation system. However, the characteristics of the applied PMM demand model appear to be representative of many other nonlinear demand models in terms of the functional form (i.e., s-shaped curve), the number and nature of explanatory variables, and the value of the model's parameters (or elasticities). Similarly, the Washington, D.C., data seem to be representative of many urban areas in terms of their variety of socioeconomic characteristics, trip patterns, available transportation alternatives, and so on.

It is, of course, possible that the demand function will be of unusual nature (e.g., step function, many discontinuities), or that the analysis situation will have extreme characteristics (e.g., very high- and very low-income people living in the same blocks). But such cases are not very common. Assuming that the characteristics of our case study are representative of many analysis situations, then in view of preliminary results (6) and in the absence of any evidence to the contrary, the empirical results obtained here appear to indicate the general behavior of the aggregation bias.

ACKNOWLEDGMENTS

This study was part of a project at the Massachusetts Institute of Technology, sponsored by the U.S. Department of Transportation for development of an aggregate model of urban travel behavior. The data were provided by the Metropolitan Washington Council of Governments.

REFERENCES

1. Charles River Associates. A Disaggregate Behavioral Model of Urban Travel Demand. Rept. prepared for the Federal Highway Administration, U.S. Department of Transportation, Mar. 1972.
2. M. L. Manheim. Fundamental Properties of Systems of Demand Models. Department of Civil Engineering, Massachusetts Institute of Technology, Cambridge, Discussion Paper T70-1, 1970.
3. T. J. Atherton and M. Ben-Akiva. Transferability and Updating of Disaggregate Travel Demand Models. Paper presented at the 55th Annual Meeting, TRB, Jan. 1976.
4. M. E. Ben-Akiva. Structure of Passenger Travel Demand Models. Department of Civil Engineering, Massachusetts Institute of Technology, Cambridge, PhD thesis, 1973.
5. F. Koppelman. Prediction With Disaggregate Models: The Aggregation Issue. TRB, Transportation Research Record 527, 1974, pp. 73-80.
6. F. Koppelman. Travel Prediction With Models of Individual Choice Behavior. Department of Civil Engineering, Massachusetts Institute of Technology, Cambridge, PhD thesis, June 1975.
7. F. Koppelman. Methodology for Analyzing Errors in Prediction With Disaggregate Choice Models. TRB, Transportation Research Record 592, 1976, pp. 17-23.
8. U. Landau. Sketch Planning Models in Transportation Systems Analysis. Department of Civil Engineering, Massachusetts Institute of Technology, Cambridge, PhD thesis, Feb. 1976.
9. D. McFadden. The Measurement of Urban Travel Demand. Journal of Public Economics, Vol. 3, No. 4, Nov. 1974.
10. D. McFadden and F. Reid. Aggregate Travel Demand Forecasting From Disaggregated Behavioral Models. TRB, Transportation Research Record 534, 1975, pp. 24-37.
11. A. Talvitie. Aggregate Travel Demand Analysis With Disaggregate or Aggregate Travel Demand Models. Transportation Research Forum, Vol. 14, No. 1, 1974.
12. C. Difiglio and M. F. Reed. Transit Sketch Planning Procedures. Paper presented at the 54th Annual Meeting, TRB, Jan. 1975.
13. F. Dunbar. Evaluation of the Effectiveness of Pollution Control Strategies on Travel: An Application of Disaggregated Behavioural Models. Paper presented at the 16th Annual Meeting of the Transportation Research Forum, Toronto, Nov. 1975.
14. E. Ruiter and M. Ben-Akiva. A System of Disaggregate Travel Demand Models: Structure Component Models and Application Procedures. Rept. prepared for Cambridge Systematics Inc., Cambridge, MA, Jan. 1977.
15. Peat, Marwick, Mitchell and Company. Implementation of the N-Dimensional Logit Model: Final Report. Paper prepared for Comprehensive Planning Organization, San Diego County, CA, May 1972.

*Publication of this paper sponsored by Committee on Passenger Travel Demand Forecasting and Committee on Traveler Behavior and Values.*

# Planning Model for Transportation Corridors

Antti Talvitie, Department of Civil Engineering, State University of New York at Buffalo

**The paper describes a planning model for transportation corridors and outlines its application to one such study. The components of the model—demand, level-of-service equilibration, and computation of travel impacts—all have new features that make the model system easy and inexpensive to operate. Demand is predicted on the basis of a representative sample of households from the study area by using a disaggregate logit model; level of service is expressed by means of equations, which avoids the pitfalls and costs of lengthy network coding procedures; equilibration is accomplished by solving, in essence, the demand and level-of-service equations. Finally, the consequences of policies can be computed by any user-defined market segment. This is made possible by use of a representative household sample and the disaggregate demand and level-of-service models.**

In this paper a brief description of a planning model for transportation corridors is given, and its application to one corridor study—the I-580 corridor in the San Francisco Bay Area—is outlined. Because the model was developed for this particular study, the model and the study environment are interwoven. Nevertheless, the model system is transferable in its present form to other corridors characterized by a strong radial movement. A complete description of the model system, its computational efficiency, and the results of the pilot application are given elsewhere (1).

The components of the model system are (a) prediction of demand, (b) prediction of transportation system level of service, (c) computation of the equilibrium of demand and transportation level of service, (d) assessment of the costs and revenues of transportation policy alternatives, and (e) processing of information for evaluation of the results. Before describing these components separately, I shall briefly relate what is new and what is omitted from this policy-planning tool.

The present model is different from the traditional transportation analysis models in five ways. First, travel forecasting is done by using a representative sample of households in the study area. This facilitates the use of disaggregate forecasting techniques to calculate the total travel demand as the sum of individual travel choices. Second, levels of service (LOS) for both access and linehaul are expressed by means of equations. These service models, which translate policies and plans into LOS, are utilized in a manner that provides disaggregate values for all LOS components except peak-period linehaul time. The peak-period linehaul travel time is equilibrated to be equal for all travelers having the same origin-destination (O-D) zones. The third new feature of the model system is the method of equilibration. Equilibration is accomplished, in essence, by solving the demand and service equations simultaneously. Fourth, owing to the use of disaggregate forecasting, sample consequences of plans can be computed for any user-defined market segment. Fifth and finally, the model system is inexpensive and easy to operate and apply.

The model system is still incomplete in important ways. The cost side of the transportation system is not yet an integral part of the models, even though costs were part of the study. Nonwork and auto-ownership decisions are not included in the present travel demand model. The inclusion of such trips and decisions in the model is not conceptually difficult. Nonwork trips can be easily added on at small cost. For the particular case study for which the model system was developed, the inclusion of nonwork trips was not considered necessary. The incorporation of auto-ownership decisions is also possible, although more costly, particularly if equilibration of such decisions is attempted. Again, for a short run, for which the model is now being applied, auto ownership is unlikely to be discernibly affected by the proposed transportation policies. Thus, the treatment of auto ownership as an exogenous variable is justified. Finally, and most importantly, the feedback relationship among transportation plans and policies and location of residences and jobs is assumed not to exist in spite of evidence to the contrary.

Because of these omitted relationships, the assessment of results of the technical analyses on consequences of various actions must be done with sensitivity to the omissions.

The components of the model system—demand, service, equilibrium, assessment of costs and revenues, and processing information for evaluation—are described next. These are followed by a brief description of the application of the model system.

## PREDICTION OF DEMAND

Prediction of peak-period travel demand is done in two steps. First, on the basis of base yearly information about the distribution of households and jobs and of socioeconomic attributes of individuals, a representative sample of households is projected for the desired future year(s). In planning for the short run this sample can be the base-year sample itself. For the present application a special method was developed for projecting a representative sample for 1976 from 1970 census information (2).

Second, given the O-D work trip demands of households and the socioeconomic attributes of households, the mode choices are calculated using the familiar logit model. This work-trip mode-choice model is discussed in detail next.

The structure of the chosen mode-choice model is best shown by means of the tree diagram in Figure 1, where the numbers in parentheses are the numbers by which modes are identified in this study.

There are eight primary modes. For many of them there are also subchoices regarding access or egress modes. For the first three modes (driving alone, ride sharing, and local bus) access and egress are always assumed to be by walking, and there is no subchoice involved.

For express bus (mode 4), the access and egress modes, labeled a and b, are determined independently of each other by using an access mode-choice model estimated from a San Francisco Bay Area Rapid Transit (BART) ridership sample. The access and egress modes having the highest probability of choice are assumed chosen. This determines how the access and egress attributes are computed for the door-to-door trip.

For BART (modes 5-8), the access mode is determined jointly with the linehaul mode, while the egress

Figure 1. Tree diagram of the mode-choice model.

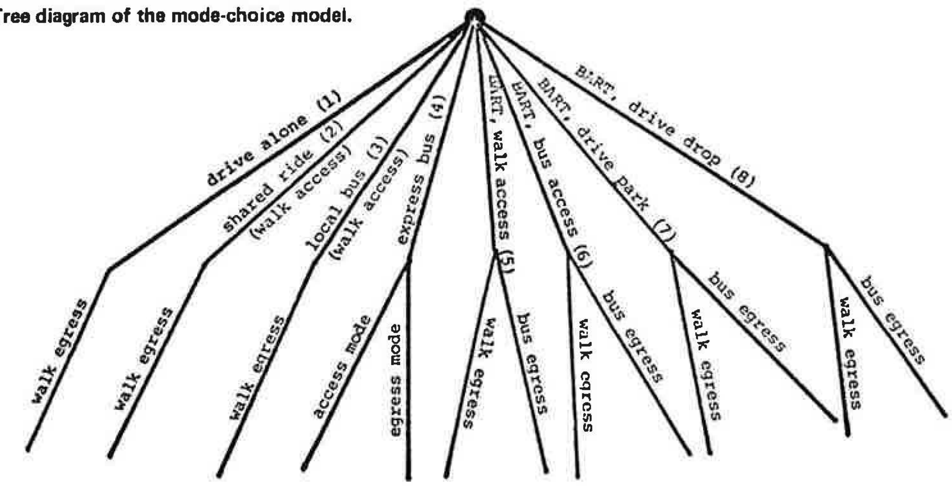



Table 1. Access mode-choice and primary mode-choice models.

| Variable | Access Mode-Choice Model | | Primary Mode-Choice Model | |
|---|---|---|---|---|
| | Alternative[a] | Coefficient (t-value) | Alternative[b] | Coefficient (t-value) |
| Cost divided by wage (min) | 2, 3 | -0.120 0 (4.2) | 1-8 | -0.038 35 (4.3) |
| In-vehicle time (min) | 2, 3, 4 | -0.065 2 (2.3) | 1-8 | -0.048 48 (2.6) |
| Walking time (min) | 1, 2, 3, 4 | -0.065 1 (3.6) | 1-8 | -0.047 10 (3.2) |
| First headway (min) | 2 | -0.032 2 (1.4) | 3, 4-8 | -0.023 38 (0.9) |
| Transfer time (min) | NA[c] | - | 3, 4-8 | -0.063 52 (2.9) |
| Household income ($) | 3, 4 | 0.000 059 (1.7) | 1.2 | -0.000 054 7 (1.2) |
| Autos per driver | 3 | 3.142 (3.2) | 1, 7, 4 (if drove) | 1.717 (2.6) |
| Autos per driver | 4 | -1.014 (1.1) | 2, 8, 4 (if driven) | 0.680 6 (1.1) |
| Drivers | 3 | 0.680 (2.3) | 1, 7, 4 (if drove) | 0.660 2 (2.5) |
| Drivers | 4 | 0.680 (2.3) | 2, 8, 4 (if driven) | 0.728 0 (2.8) |
| Distance to station ≤0.8 km | 1 | 4.166 (3.5) | NA | - |
| Employment density | NA | - | 1 | -0.001 923 (4.1) |
| Alternative-specific dummy | 1 | 3.473 (2.9) | 1 | -1.167 (1.2) |
| Alternative-specific dummy | 2 | 4.470 (4.0) | 2 | -2.099 (2.3) |
| Alternative-specific dummy | 3 | -0.648 (0.8) | 4 | -1.120 (1.4) |
| Alternative-specific dummy | NA | | 5 | -1.422 (2.9) |
| | NA | | 6 | -0.832 4 (1.9) |
| | NA | | 7 | -3.581 (4.3) |
| | NA | | 8 | -4.065 (5.1) |

Note: 1 km = 0.62 mile.

[a] Alternative numbers are 1 = walking, 2 = bus, 3 = drive, and 4 = driven.
[b] Alternative numbers are given in Figure 1.
[c] Variable not used in the model.

mode from station to work is decided by using the access mode-choice model. Again, the maximum egress mode is chosen. In addition to access and egress modes, access and egress stations also need to be determined for express bus and BART modes. This can be done by using the access mode-choice model referred to earlier; here a shortcut was employed whereby the station nearest to the individual's home or work was assumed chosen.

A few words are in order to support such a "mongrel" mode-choice model, particularly as it could not be derived from such well-known principles of choice behavior as utility maximization because they were not available at the time the decision regarding the model structure was made. Thus, there was no theoretical pressure to choose a sequentially estimated model with the access and egress modes characterized by an inclusive price variable that is consistent with the logit model (3).

Second, the choice of egress mode, normally ignored in both demand analysis and forecasting, appears both necessary and desirable. However, inclusion of egress mode choice as a joint decision with access mode and linehaul mode would have doubled the number of BART modes and would have left so few observed choices in each alternative that estimating alternative-specific constants would have been inaccurate. A likely violation of the "independence from irrelevant alternatives" (IIA) assumption would also have occurred. The same reasons can be cited for keeping the express bus as one mode. The number of observations did not permit the estimation of access mode choice jointly with express bus linehaul choice. And, if the prediction of the use of express bus access modes was done with the joint structure, the IIA property would have been especially aggravating, because express bus on an exclusive lane is probably very similar to BART in its performance.

In summary, then, the adopted model structure is a compromise between conflicting interests. It is believed that the chosen structure does not result in the overprediction of transit use by the IIA property, that the separation of transit demand between BART and express bus modes is accomplished satisfactorily, and that the equilibration of important system variables is possible in a straightforward manner. In short, the mode-choice model is tailored to suit the needs of the policy analysis not vice versa. It may also be noted that the installation of a (recursive) logit model would not be an involved task.

The specifications of the access mode-choice model and of the primary mode-choice model are given in Table 1.

## PREDICTION OF DEMAND: MARKETS AND METHOD OF AGGREGATION

In order to predict the modal demand volumes, the study corridor is divided into segments (nine in the present application). These segments are mutually exclusive and cover the entire length of the corridor and serve as the markets on which equilibration is done. Thus, the travel demands in each segment are not obtained by link and bus line but by mode and access mode. If necessary, auto demand can be divided into arterial and freeway volumes; station use by access mode can also be obtained.

The demand of travel on mode m over segment j is computed as the summation of individual modal demands expanded by the sampling rate $\theta$. The disaggregate forecasting method expressed mathematically is

$$D_{jm} = \sum_{k=1}^{K} (1/\theta)\, \delta_j^k P_m^k \qquad m = 1 - M, j = 1 - J \tag{1}$$

where

$\theta$ = sampling rate,
$P_m^k$ = probability of individual k choosing mode m computed using the mode-choice model in Table 1,
K = sample size,
$\delta_j^k$ = fraction of segment j used by individual k on the trip ($0 \le \delta_j^k \le 1$), and
$D_{jm}$ = demand (the number of trips) made within or through segment j on mode m.

## TRANSPORTATION LOS MODELS

Each transportation policy alternative plan is described by a set of options. The options considered in the present model system are shown below.

| Policy Option | Application |
|---|---|
| Bus line spacings | Access |
| Bus stop spacings | Access, linehaul |
| Bus/BART headways | Access, linehaul |
| Bus/BART fares | Access, linehaul |
| Location of BART stations | Access |
| Residential and employment density distributions | Access |
| Auto operating cost | Access, linehaul (per kilometer cost) |
| Availability of transit service | Linehaul |
| Tolls | Linehaul (segment j) |
| Parking cost | Destination zone |
| Highway and street capacities (no. of lanes) | Linehaul (segment j) |
| Highway or street capacity reserved for priority use | Linehaul (segment j) |
| Signal green time allotted to major flow direction | Linehaul (segment j) |

The values of these options define a policy or plan and yield the values of the LOS attributes for access and linehaul. These LOS values are calculated from the relationships between the attributes and policy actions.

### Access Service Models

The specific relationships developed for the access models are given in Tables 2 and 3. Their detailed development is described elsewhere (4). Only few words are said here, mostly on their application.

Application of the access travel time models in policy analyses is done in the following way. It is assumed that the access and egress travel times on various modes have a log-normal distribution whose mean and standard deviations are given by the equations. The log-normal distribution was chosen mostly on empirical grounds.

Given the O-D zones of workers and the transportation system attributes of these zones defined in terms of the policy options, random draws are made from the log-normal distributions of access and egress time components for each individual in the sample. In this way disaggregate values for access and egress times are obtained for each individual. In order to facilitate application, it is assumed that the access times do not require equilibration. This assumption is tenable for other access variables except walking time in a BART parking lot. Equilibration of this variable poses no conceptual problems.

### Linehaul Service Models

The linehaul travel times, which require equilibration, were derived by using an extension of the point bottleneck method (5, 6). In order to utilize the method, the bottleneck must be identified for each segment. The freeway and arterial capacities for a bottleneck are

$C_j^N = 1770(FW_j - PL_j)$ freeway capacity on nonpriority lanes,
$C_j^P = 1770 PL_j$ freeway capacity on priority lanes, and
$C_j^A = 1200 \times AW_j \times G_j$ arterial capacity,

where 1770 VPH (vehicles per hour) and 1200 VPHG (vehicles per hour green time) are the lane capacities for freeways and arterials, respectively, and $FW_j$, $AW_j$, and $PL_j$ are the number of lanes in the segment bottleneck for freeways, arterials, and priority lanes. $G_j$ is the green time split to the major flow direction.

It is then assumed that travelers distribute themselves so as to minimize their travel times (Wardrop's first principle); the combined service level of parallel capacities $C_j^A$ and $C_j^F$ is obtained by summing them horizontally.

Table 2. Definitions of explanatory variables.

| Variable | Mnemonic | Designation | Definition |
|---|---|---|---|
| CO | Coordinates | $m_x + m_y$ | Sum of the stations or major bus line coordinates (measured from the center of the zone), in kilometers[a] |
| ST | Stops | $b_x + b_y$ | Sum of bus stop intervals in each direction, in kilometers |
| SP | Spacing | $s_x + s_y$ | Sum of bus route spacings in each direction, in kilometers |
| L | Side | $\ell$ | Side length of zone, in kilometers |
| HD | Headways | $h_x + h_y$ | Sum of minor bus line headways, in min |
| HX | Headway | $H_x$ | Major bus line headway to x-direction ($H_y = 0$), in min |
| BMI | Bus kilometers | $60 \times \ell^2 [(1/s_x h_x) + (1/s_y h_y)]$ | Bus kilometers of service within zone (over and above offered by major line $H_y$) |
| SD | Standard deviations (of development density) | $v_x + v_y$ | Sum of the development density standard deviations |
| HO | Hole | $(r_x/v_x) + (r_y/v_y)$ | Note, if station is outside the zone, then $r_x = [m_x - (\ell/2)]/v_x$ and $r_y = [m_x - (\ell/2)]/v_y$ |
| DU | Dummy | 0 | Station inside the zone |
| | | 1 | Station outside the zone |
| DV | Dependent variable | | Appropriate travel time or its standard deviation |

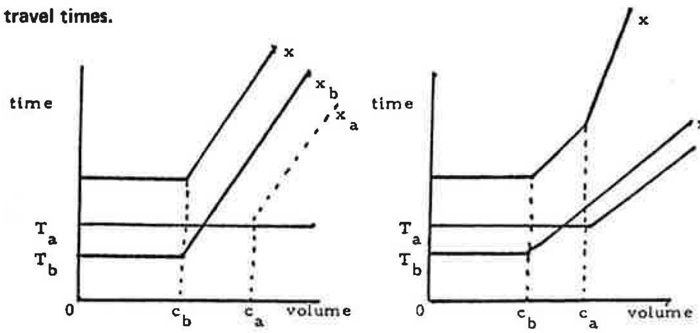
[a] Note that the summation of the variables simply means that the coefficients of the summed variables are constrained to be equal.

**Table 3. Estimated models for the access travel times on various modes.**

| | Walking or Driving Time to Station and Speed = f(variables, $a_i$) | | | | Walking Time to Bus Stop = f(variables, $a_i$) | | | | Bus Ride Time to Station Times Speed = f(variables, $a_i$) | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Mean | | Standard Error | | Mean | | Standard Error | | Mean | | Standard Error | |
| Variable[a] | Coefficient | t-Value | Coefficient | t-Value | Coefficient | t-Value | Coefficient | t-Value | Coefficient | t-Value | Coefficient | t-Value |
| Constant $a_0$ | -0.271 | 4.6 | -0.0271 | 1.1 | 1.004 | 1.4 | 0.830 | 1.6 | -0.282 | 2.9 | -0.099 1 | 2.2 |
| CO $a_1$ | 0.272 | 32.1 | 0.175 | 48.4 | - | - | - | - | 0.275 | 7.4 | 0.147 | 17.6 |
| ST $a_2$ | - | - | - | - | 5.472 | 6.0 | 0.922 | 1.5 | - | - | - | - |
| SP $a_3$ | - | - | - | - | 1.158 | 7.7 | 0.656 | 6.3 | - | - | - | - |
| L $a_4$ | 0.272 | 32.1 | 0.175 | 48.4 | - | - | - | - | 0.281 | 10.0 | 0.188 | 17.6 |
| HD $a_5$ | - | - | - | - | 0.093 1 | 9.8 | 0.047 7 | 7.2 | -0.006 20 | 4.7 | -0.000 554 | 0.9 |
| HX $a_6$ | - | - | - | - | 0.045 5 | 2.7 | 0.019 0 | 1.6 | -0.002 17 | 0.8 | 0.001 64 | 1.4 |
| BMI $a_7$ | - | - | - | - | 0.001 37 | 1.9 | 0.003 23 | 6.3 | - | - | - | - |
| SD $a_8$ | 0.138 | 11.2 | 0.0159 | 3.0 | - | - | - | - | 0.150 | 8.9 | 0.040 1 | 5.2 |
| HO $a_9$ | 0.293 | 4.7 | -0.176 | 6.6 | - | - | - | - | - | - | - | - |
| DU $a_{10}$ | 0.563 | 12.6 | -0.0949 | 5.0 | 0.562 | 1.6 | 0.846 | 3.4 | 0.942 | 8.2 | 0.0 | - |
| F-value | 593.8 | | 720.3 | | 41.9 | | 34.0 | | 293.0 | | 257.4 | |
| $R^2$ | 0.89 | | 0.91 | | 0.49 | | 0.44 | | 0.87 | | 0.83 | |
| CV | 0.16 | | 0.15 | | 0.22 | | 0.30 | | 0.23 | | 0.20 | |
| Standard error of estimate | 0.24 | | 0.10 | | 2.20 | | 1.53 | | 0.34 | | 0.15 | |
| Mean value of DV | 1.53 | | 0.67 | | 10.1 | | 5.1 | | 1.49 | | 0.79 | |

[a]See Table 2 for mnemonics.

**Figure 2. Single bottleneck versus network travel times.**




The travel times for driving alone on highway modes over segment j of length $L_j$ for any given volume $B_j$ are given by

$$x_{j1} = \{ [T_j^F + \min(T_j^A - T_j^F); \max(0; B_j^N/C_j^N - 1) P/2] + \max[0; (B_j^N + B_j^A)/C_j - 1] P/2 \} \times 60 \quad (2)$$

where $B_j^N$, $B_j^P$, and $B_j^A$ are peak-hour auto-equivalent demands (from the demand models) on nonpriority and priority freeway lanes and arterials, $T_j^F$ and $T_j^A$ are their free speed travel times in hours, and P is the length of the peak period in hours.

Thus for a nonpriority carpool

$$x_{j2} = x_{j1} + \text{constant penalty (assumed 5.0 min)} \quad (3)$$

while for a priority carpool it is

$$x_{j2} = \{ T_i^F + \min [(T_j^A - T_i^F); \max(0; B_j^P/C_j^P - 1) P/2] \} \times 60 + \text{penalty (5.0 min)} \quad (4)$$

and for local bus

$$x_{j3} = x_{j1}^A + \text{delays due to acceleration-deceleration} + \text{delays due to boarding and alighting passengers} \quad (5)$$

and for express bus

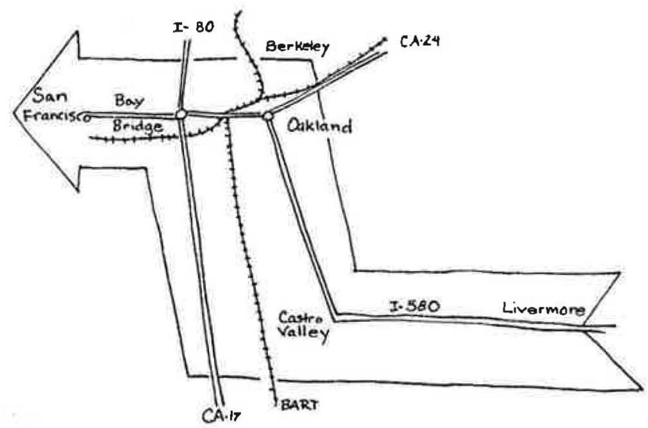
$$x_{j4} = x_{j2} + \text{delays due to acceleration-deceleration} + \text{delays due to boarding and alighting passengers} \quad (6)$$

It is instructive to contrast the above point bottleneck formulation of segment travel time to the normal network approach.

Consider a segment that consists of two roads in sequence with travel time curves $x_a$ and $x_b$. Now the single bottleneck approach identifies the restraining capacity, which is shown to be on road b (Figure 2) and assumes that there is no capacity restraint on road a. Hence, road a is characterized by a constant free speed travel time, $T_a$, and the segment travel time curve is derived by a vertical addition of $T_a$ and $x_b$. This can be compared to vertically adding $x_a$ and $x_b$, which is done by the network algorithms. It can be seen that in the range of travel volumes from zero to $C_a$, the single and multiple bottleneck formulations give identical travel times. Beyond $C_a$ the travel times are different, the multiple bottleneck travel time being larger than the single bottleneck travel time. It is not clear which is correct. If the two consecutive road links are independent, as in the case of a low-volume arterial street governed with unsynchronized traffic signals, then the multiple bottleneck formulation (every signal is a bottleneck) is approximately correct. However, if the consecutive links are not independent, as in the case of freeways and most arterials, then the single bottleneck version is approximately correct.

The single bottleneck formulation is very attractive because of its simplicity and small data requirements. Its success will depend in large measure on whether the delays in a segment of a corridor are due to congestion and whether the corridor segments are independent. Because the segments here are several kilometers long and

Figure 3. I-580 corridor.




the delays during the peak are mostly due to congestion, the assumption leading to a point bottleneck model are likely to be satisfied.

May and Keller (5) have used the point bottleneck model on a single freeway segment several kilometers long with apparent success. The prevalence of signals and stop signs and other disturbances on arterials suggests, however, that the travel time and volume function ought to be estimated statistically using collected data.

In summary, once the plan or policy options have been specified, the equations in Table 3 will yield specific values for access attributes of various modes for each household (individual) in the sample and will define the parameters of the linehaul time equations. The simultaneous solution to the demand equations and to the linehaul travel time equations is the sought equilibrium. The method employed in equilibrating these equations is discussed elsewhere (6). Suffice it to note here that the equilibration procedure [utilizing a fixed-point algorithm developed by Scarf (7)] was found to be both efficient and instructive.

## EVALUATION MEASURES

The evaluation measures pertaining to the technical analyses provide support for subjective political decision making and allow the consideration of different consequences that alternative transport policies have on different groups in the transportation corridor. Accordingly, the model system permits any grouping of people (market segments) in the representative sample. A segmentation that divides the population into nine groups was chosen in the present study. These nine market segments constitute three income groups—low, medium, and high—and three work trip groups—suburb to suburb, suburb to central business district (CBD), and urban to urban. Various summary cross-comparison tables between market segments and alternative plans are also possible.

The evaluation measures calculated for the market segments include: the values of the trip time and cost components and their standard deviations, mode and access-mode use, trip distances, agency costs by mode (not yet integrated in the model), revenues by mode and market segment, and passenger- and vehicle-kilometers of travel.

## SAMPLE APPLICATION IN THE I-580 CORRIDOR

The model system is being applied to the I-580 corridor study in the San Francisco Bay Area. The study is still in progress, and only an indication of the type of results that can be obtained can be given. In order to gain an appreciation of the great complexity of the study, a brief description of the area is given.

The geographic scope of the study includes areas for transbay and corridor transit ridership and automobile use along I-580 from Vasco Road in Dublin over the San Francisco-Oakland Bay Bridge to the San Francisco peninsula. The corridor is 72 km (45 miles) long, and about two million households occupy the immediate neighborhood of the transportation corridor (see Figure 3).

The direct cause of the present study was the approval of a major widening of I-580 in Alameda County in December 1974. This approval was conditioned by the Metropolitan Transportation Commission (MTC) to include facilities for preferential bus and carpool treatment in the form of reserved lanes and to study the feasibility of extending exclusive bus and carpool operations throughout the length of I-580 from Vasco Road to the San Francisco-Oakland Bay Bridge.

A sample of results is given in Figure 4. By examining Figure 4, it is seen first that, proportionally, BART riders are well-to-do people. For all plans, over 10 percent of high-income travelers use BART, while only 5 percent or less of low-income people use BART. Across the income groups, only 8-10 percent of the BART users are low-income travelers (<$10 000/year); over 55-60 percent of the users have a high income (>$20 000/year); the remainder are middle-income travelers. Carpooling appears to be the most popular mode among low-income people; 40 percent of them share rides. The corresponding figure for high-income people is 20 percent or slightly less.

Figure 4 also contains a surprise: modal shares are not affected much by traffic-engineering measures or transit improvements. Increases in fares (plans I-K) seem to have the greatest impact in aggregate shares. This surprise becomes understandable when it is noted that the plots of Figure 4 pertain to all travelers affected by the corridor transportation policies, including the Bay Bridge, not only those living or working within the corridor area.

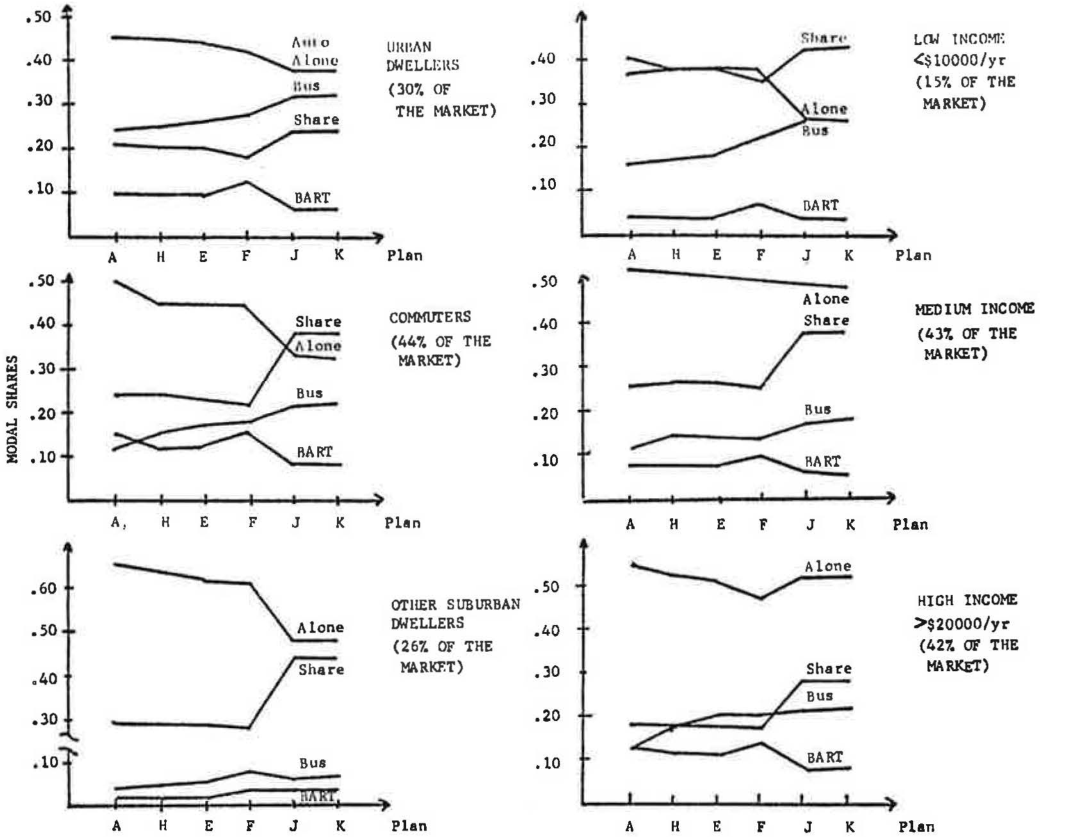
It may be noted that the provision of a multiple occupancy vehicle priority lane on I-580 (plan L) reduces the driving alone volume and BART ridership by 2 and 5 percent respectively. Shared-ride volume has increased by 2.5 percent and bus volume by 7 percent. The introduction of express bus service in extended priority lane (plan B) further reduces driving alone and BART volumes, with a slight increase (1 percent) also taking place in the shared-ride mode. There is a substantial increase (27 percent) in bus volumes.

The initiation of a full BART service with 3-min headways and increased feeder bus service, in addition to the express bus service in a priority lane (plan F) results in a 10 percent drop in driving alone and a 5 percent decrease in ride sharing. Buses have increased their ridership, and BART experienced a 15 percent increase in ridership.

In general it appears that the introduction of expanded bus service with current fare structure reduces BART ridership unless BART headways are substantially reduced and feeder service is improved.

Results pertaining to the full cost plan (plans J and K) are quite interesting. The full cost includes only the monetary costs of supplying the service (but no user time costs) and the replacement costs of facilities and vehicles. An exception was made in the case of BART, where the historical costs of construction and land acquisition were used in place of the replacement costs. This procedure is incorrect, but the replacement costs

Figure 4. Modal shares by residence and by income group.




of BART facilities were overwhelmingly large. The feeling was that sunk costs are sunk, but that the bounds and the costs of continued service must be paid for. A flat fare was used for local bus by using average travel distance; all other fares were based on distance traveled.

Not surprisingly, the imposition of full cost causes a substantial reduction in driving alone (28 percent) and BART use (39 percent). Shared-ride and bus modes, operating on exclusive lanes whenever possible, experienced substantial increases (46 and 56 percent respectively). As a consequence of full cost pricing, the average trip cost for low-income families would go up 85 percent (from $0.73 to $1.35) and 105 percent for high-income families (from $1.25 to $2.57).

Expressed as a function of income, the yearly work trip currently takes 5.6 percent of the low income and 2.0 percent of the high income. After imposing the full cost pricing, the corresponding percentages are 10.4 and 4.0. So, even though the full costs hit the higher income people harder, their proportional take from the low income is still substantially more than that from the high income. These figures do ignore the fact that, with exceptions, low-income people drive less expensive cars, keep the cars longer, and often do their own maintenance and servicing. Thus the car cost for low-income people is likely to be an overestimate.

The picture changes somewhat when only the transit fares are considered. Currently, the average transit fare is $0.50 for the low-income and $0.80 for the high-income users. After imposing the full costs, the fare for low-income users increases to $0.55 (10 percent), and the fare for the high-income users goes to $1.60 (100 percent). It can be inferred from these numbers that the current fare structure makes the low-income users pay more nearly for the actual costs of transportation than the high-income users do. A generalization of this is that the flat fare systems or distance-rebated fares are regressive [for other cross-subsidies see Hoachlander (8)].

The impacts on other LOS attributes besides user cost were also interesting. In brief, the provision of a priority lane did worsen the service levels on the segment from Livermore to Castro Valley. The provision of a priority lane from Castro Valley to the Bay Bridge increased the speeds of the high-occupancy vehicles, while the speeds of the nonpriority vehicles either remained unchanged or increased. Interestingly, the full cost plans J and K provide the highest speeds, with plan K being marginally better. In general, the priority-lane plans conferred benefits to higher-income groups and to the suburbanites; the urban dwellers and the low-income people (who are not the same) did not benefit from the provision of the priority lane.

The beneficiaries of the transit improvements studied here were also the suburbanites, especially the high-income commuters. Urban dwellers and the low-income people were largely left with their present levels of service. An exception to this is plan F, which does improve the access components across all the income groups; even with the drastic changes of plan F the urban dwellers gain only in reduced wait times. The full cost plans J and K also improve both the linehaul and wait time components rather evenly across the market segment, but the walking times to transit remain unchanged.

In sum, the currently popular policies seem to confer benefits mostly to the well-to-do suburbanites. These policies, coupled with current fare structures, also provide heavy subsidies to the well-to-do suburbanites. It is likely that if the strategies are pursued without remedial actions in fares, they will further encourage urban sprawl and, unfortunately, increase energy consumption.

This brief examination of the results may be concluded by noting one interesting item not evident from the tables presented here. This most interesting, intuitively obvious result is the stability of the Bay Bridge travel volume. With the exception of plans J and K (full BART service and full costs), the work-trip volume crossing the bridge is only 1300 less than now. Interestingly, the 1300 reduction in bridge volume can be expected as a result of offering full BART service with 3-min headways (current headways are 12 min) and doubling the feeder bus service to BART, in addition to express bus service in an exclusive lane.

The reason for the insensitivity of the Bay Bridge volume is that, while travelers within the corridor do shift to transit in response to protransit policies, those not originating from the corridor but traveling over the Bay Bridge are switching to cars. Thus, if Bay Bridge congestion is a social ill that needs to be eliminated, one effective way to do it is by charging more for travel. In particular, an eminently sensible charge of full costs (which does not include time costs) appears to come close to attaining the noncongestion objective.

## CONCLUSIONS

Experience with the corridor model system indicates that it is flexible, fast, and inexpensive to operate. The ten or so plans were run in one working week at a cost of $35 per run. These attributes of the model system owe much to the way in which policy statements are translated into LOS variables (equations rather than networks) and the new way of equilibrating service and demand. The disaggregate forecasting method employed also proved to be quite useful; it enabled the analysis of equity issues in transportation service and costs.

In spite of the many innovations and strengths of the model system, there are also some weak points. The first element in need of improvement is the work-trip destination choice model. There is some evidence, too lengthy to detail here, that the linear probability model utilized in that task is not a very good predictor. However, it may not be any worse a predictor than any current work-trip distribution model; research work on that topic is recommended.

The second element requiring improvement consists of the automobile travel time equations. The single bottleneck concept together with the Wardrop aggregation of parallel links is too insensitive to changes in highway capacity. Many specific items require research in order to improve this state of affairs. For example, more research is needed in route-choice behavior between arterials and expressways. The same holds true for developing equations for travel time that have more explanatory variables than just volume and capacity.

Improvements in the two above-mentioned items require only minor changes in the workings of the model system itself, but the development of specific models could require several years.

The third element that needs attention is the establishment of relationships among location of residence, places of work, and transportation LOS. This is a very large and complex task, where empirical knowledge, to say nothing of theories and models, is sorely needed. Nonetheless, it must rank high in its importance and deserves careful investigation by the planning agencies and others involved in planning tasks.

## ACKNOWLEDGMENTS

I have benefited from discussions with Daniel McFadden during the formulation of the study approach. Melvin Cheslow of the Urban Institute provided a good critical review of an early draft of this paper that materially improved the final product. Jerome Berkman did the expert programming of the model system. This research was supported in part by a National Science Foundation grant for Research Applied to National Needs to the University of California, Berkeley.

## REFERENCES

1. A. Talvitie and others. A Corridor Study of Interstate I-580 From Livermore to San Francisco. Urban Travel Demand Forecasting Project, Institute of Transportation Studies, Univ. of California, Berkeley, Final Rept. No. 10, 1978.
2. D. McFadden and others. Demographic Data for Policy Analysis. Urban Travel Demand Forecasting Project, Institute of Transportation Studies, Univ. of California, Berkeley, Vol. 8, 1977.
3. D. McFadden. Quantitative Methods for Analyzing the Travel Behavior of Individuals: Some Recent Developments. Urban Travel Demand Forecasting Project, Institute of Transportation Studies, Univ. of California, Berkeley, 1977.
4. A. Talvitie and Y. Dehghani. Models for Transportation System Performance. Paper presented at the International Symposium on Travel Supply Models, Montreal, 1977.
5. A. D. May and H. E. M. Keller. A Deterministic Queuing Model. Transportation Research, Vol. 1, No. 2, 1976, pp. 117-128.
6. I. Hasan and A. Talvitie. An Equilibrium Mode-Split Model of Work Trips Along a Transportation Corridor. Paper presented at the 2nd World Conference on Transportation Research, Rotterdam, Netherlands, 1977.
7. H. Scarf. The Approximation of Fixed-Points of a Continuous Mapping. SIAM Journal of Applied Mathematics, Vol. 15, 1967, pp. 1328-1343.
8. E. Hoachlander. Bay Area Transit—Who Pays and Who Benefits. Institute of Urban and Regional Development, Univ. of California, Berkeley, Working Paper 267, 1976.

*Publication of this paper sponsored by Committee on Passenger Travel Demand Forecasting and Committee on Traveler Behavior and Values.*

# Hypernetworks and Supply-Demand Equilibrium Obtained With Disaggregate Demand Models

Yosef Sheffi, Department of Civil Engineering, Massachusetts Institute of Technology, Cambridge
Carlos F. Daganzo, Department of Civil Engineering, University of California, Berkeley

**This paper presents a framework for discussing many transportation demand and supply-demand equilibrium problems. It regards the sequence of choices an individual faces when he or she is about to make a travel (or not-to-travel) decision as a case choice of a route on an abstract network (hypernetwork). Hypernetworks are intimately related to the multinomial probit (MNP) model of travel choice. For instance, the multivariate normal distribution underlying this model enables one to represent processes of travel choice as route choices on networks and to use the networks as visual aids in conceptualizing the specification of covariance matrices for MNP choice models. Hypernetworks enable us to carry out supply-demand equilibrium analyses with disaggregate demand models on a mathematically consistent basis (heuristic equilibration techniques based on feedback loops do not necessarily converge, as shown with a simple counterexample). This greatly enhances the potential of probabilistic discrete-choice disaggregate demand models, since it is now possible to avoid their mispredictions when applied to congested transportation systems.**

This paper departs from the main line of thought in previous transportation planning research. It suggests that most transportation forecasting problems can be regarded as network problems and that by doing so it is possible to address several outstanding and seemingly unrelated problems in a unified way. Although the urban passenger transportation planning process is used throughout this paper as an example, it should be noted that the concepts introduced here are applicable to small-scale problems and sketch-planning issues as well. In fact, if one wanted to adapt our approach to the very large models that are sometimes used in urban transportation planning, further research would be needed. A discussion and literature review of existing problems with the above-mentioned transportation planning process follows.

The most noticeable and widely used approach to transportation equilibrium analysis is the Urban Mass Transportation Administration's (UMTA's) Urban Transportation Planning System (UTPS) (1, 2, 3). The UTPS is a battery of computer programs designed to perform the above analyses, which include trip generation, trip distribution, modal split, and traffic assignment. Each phase in the process has a methodology of its own that has been extensively discussed in the literature (4, 5, 6, 7, 8, 9, 10, 11).

There are other computer packages that attempt to perform equilibrium analysis [such as Dodotrans (12), which, unlike the early versions of the UTPS, is an explicit equilibration package] and are based on similar ideas. A review of many of these packages can be found in a report by Peat, Marwick, Mitchell (13).

Although these transportation planning tools are commonly accepted among transportation planners, they have received severe criticism in the literature in the last several years.

Some of the criticism is general and points out the deficiencies of all large-scale models (14, 15, 16, 17, 18). Some of it is directed at specific models used in the planning process (19, 20, 21, 22, 23, 24). Yet others have based their criticism on a more fundamental issue that applies to small-scale problems as well—the behavioral assumptions that underlie this process (25, 26). The latter line of criticism led to the so-called disaggregate behavioral demand models (27, 28), which, by using individuals or households as the study units, attempted to capture choice-makers' behavior. Some of these models can be interpreted according to the economic principle of utility maximization, giving them a flavor of causality and behavioral realism.

Since the use of disaggregate data is more efficient—disaggregate models need less data than aggregate models to get a specific confidence level on the estimated coefficients (29)—and because the estimated coefficients are potentially independent of the distribution of the explanatory variables, those models have gained popularity among planners and are used increasingly in practice. Examples of applications and further developments of disaggregate demand models can be found elsewhere (30, 31, 32, 33, 34, 35).

These models, however, raised a new set of unresolved issues. The first of those difficulties is the aggregation problem (36, 37, 38, 39), i.e., how to use disaggregate demand models to predict the behavior of the population of choice makers. The second difficulty is in incorporating these models into a supply-demand equilibrium framework. The third difficulty is that most of the disaggregate modeling and estimation effort has been with models such as logit (40, 41) that involve assumptions that are sometimes unrealistic and often fail to capture reasonable user behavior. Obvious examples are the blue bus—red bus problem (42), the case of nested alternatives (43), and route-choice problems (44).

In addition, some of the issues that arise from the heuristic nature of the transportation planning process still remain, as happens for instance with its failure to naturally represent the intimate interdependencies among the various phases that comprise it, including the performance characteristics of the system (the "supply" side).

This is a problem with the traditional (sequential) planning process because, if, for example, modal split is performed before trip distribution, transit trips might be allocated to zone pairs not used by transit. If modal split is performed after trip distribution, too many trips might be allocated from, say, an origin with low car ownership to a destination connected to it by highway only. Although sequential models can avoid this problem by the use of accessibility measures (45, 46, 47, 48), the numerical values of these accessibility measures are not usually known a priori, which creates some problems. For instance, in congested systems, travel time (an explanatory variable appearing in trip generation, trip distribution, and modal split models) can only be determined after the traffic assignment step.

In order to circumvent such problems it is suggested that the model system be iterated several times in order to achieve a state of equilibrium, which, for the case of probabilistic and discrete choice demand models, has not been formally defined. However, due to the high computation costs involved, this is seldom done in practice.

Last, there are some aggregation problems that remain even if aggregate models are used. For instance, although market segmentation might enhance predictions (49, 50) and is commonly used in practice (trip generation and auto ownership studies are typical), no firm guidelines have been given in the literature as to the necessary extent of the segmentation. Similarly, in the traffic assignment step, no definite criteria exist on how to represent the network, i.e., how to locate the zone centroids, or on how to decide on the number and select the characteristics of dummy centroid connectors.

In summary, the following issues can be identified as those

1. Concerning disaggregate choice models in general: (a) the aggregation problem (including market segmentation), (b) incorporation in equilibrium analysis, and (c) alternatives to logit;
2. Concerning the traditional process in general: (a) equilibrium formulation and equilibration procedure, (b) consistency throughout the steps, and (c) network representation.

The objective of this research is to provide a framework within which some of these issues can be resolved. Of course, some of the problems have already been solved; numerical approaches to probit (51, 52) and generalized logit (53)—both reasonable alternatives to logit—have already been developed. Some of them, such as the aggregation problem (39), are partially solved, and yet some of them, such as efficient equilibration approaches with probabilistic choice models, remain unsolved.

This paper uses some of these results and some new ideas to formulate a solution (at least partial) to the above-mentioned problems. Mathematical formulations and algorithmic steps are not within the scope of this paper, which concentrates on the concepts and the application. A comprehensive treatment of our approach is included in Sheffi (54).

Several equilibrium models have been recently developed. The first ones dealt (rigorously) with route choice and network equilibrium only, by casting the problem as a mathematical program (55, 56). Ruiter and Ben-Akiva (57) developed a complete equilibrium-forecasting system incorporating an integrated set of production-oriented disaggregate demand models, and a conceptually similar model system is described by Jacobson (58). Neither of the last two methods, however, is guaranteed to produce the desired results in terms of convergence to a defined equilibrium. A formal solution to the equilibration problem over a transportation corridor, using disaggregate demand models, was obtained by Talvitie and Hasan (59). Their approach consists of formulating the equilibration as a fixed-point problem and solving it with the algorithm proposed by Scarf (60).

The approach taken in this research is to view and formulate all choice processes as route-choice processes on abstract networks, which we call hypernetworks, and to use an efficient procedure to analyze stochastic hypernetwork equilibrium problems. Although a numerical example is provided at the end, the emphasis of this paper is on presenting a concept rather than a new technique.

The following section, which discusses the idea of hypernetworks and relates it to existing approaches to forecasting, is followed by a section that shows how hypernetworks are related to MNP models and how they can be used to alleviate some of the above problems. Next the concept of equilibrium on a hypernetwork is defined and the possible failure of heuristic equilibrium approaches to converge is illustrated. The results obtained with a mathematical equilibration procedure that has been recently developed by the authors are also presented.

## HYPERNETWORKS

We assume that the various alternatives open to travelers in choice situations (mode, route, destination, etc.) can be viewed as paths in a hypothetical network (a hypernetwork) made up of links characterized by disutilities. We also assume that, as in route-choice problems (44), people select the shortest route, that is, the alternative with the lowest disutility. This is consistent with the principle of utility maximization of choice theory.

Assume for instance that we are concerned with a modal split-route choice problem for one single origin-destination pair, and to further simplify matters assume that there are one transit mode and two automobile routes. Figure 1 represents a possible configuration of the hypernetwork for such a problem.

In the figure there are three hyperpaths corresponding to the three alternatives. Links OA and OB represent the inherent disutility of the two modes, fare and comfort, and links AD and BD represent the travel time characteristics of the three alternatives. Choice of a car implies that the shortest route in the hypernetwork consists of the car link and a route through the street network.

In the most general case, link disutilities may be flow dependent (e.g., travel time under congested conditions), fixed (e.g., fare on a transit line), and/or multiattributed if it is so desired. It should be noted at this point, however, that the algorithm described in the sequel requires the modeling of links that exhibit flow-dependent utility as single attributed (i.e., as multiattributed with fixed weights on the attributes). Disutilities are also assumed to be additive so that the disutility of an alternative (hyperpath) equals the sum of the disutilities of the links that make it up. Both of these assumptions are discussed by Sheffi (54).

By modifying the structure of a hypernetwork, one can affect the probabilistic structure of the corresponding choice problem. This will be seen in the next section as we show how the probability of choice is affected by network topology. For instance, Figure 2 displays an alternative representation of the problem in Figure 1 that, as will be seen later, would have approximately the independence of the irrelevant alternative (IIA) property.

Figure 3 demonstrates a more complicated choice problem that can be represented by a hypernetwork. It displays a hypernetwork for a combined modal split, route, and destination choice problem, where a fraction of the population does not have access to the car mode. The links of this network are of two types. The ones belonging to the street network are real links and are associated with travel time and travel cost. All other links are dummy links representing different dimensions of the problem. For instance, the links leading from the destinations to D represent the unattractiveness of the specific destinations and the links labeled car and transit represent the unattractiveness of the respective modes. Note that $O_2$ does not have access to the street network in order to represent market segments that do not own an automobile. The number of hyperpaths in this network is larger than in the preceding example; as a matter of fact, in real problems this number can be so large as to preclude enumeration of all possible hyperpaths.

These examples were intended to illustrate that it is possible to construct a hypernetwork for many choice problems and that different market segments can be

Figure 1. Simple hypernetwork example for mode and route choice.

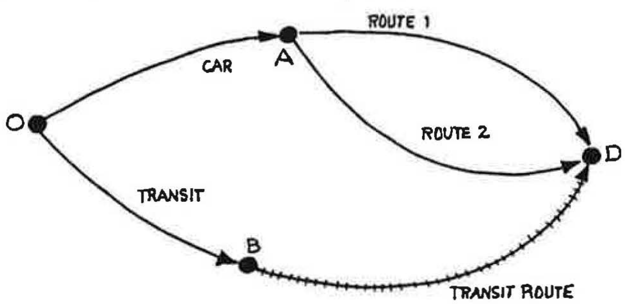



Figure 2. Independent utility representation of the hypernetwork of Figure 1.

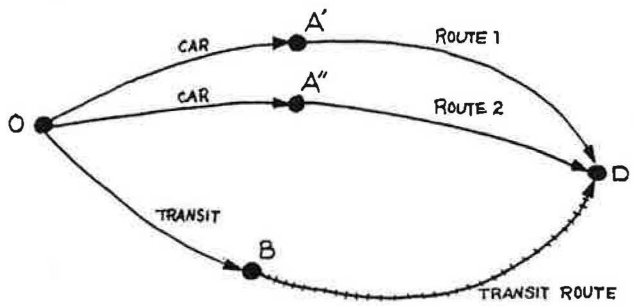
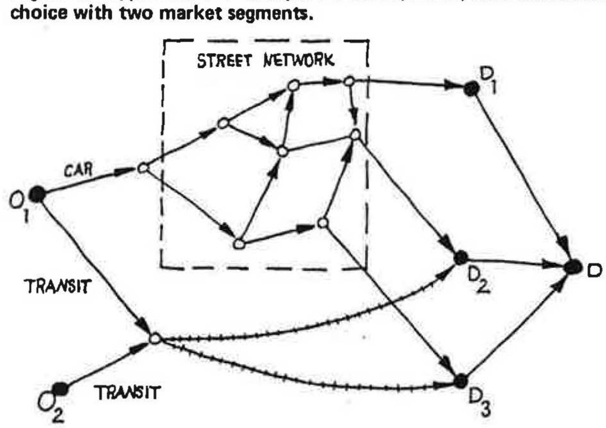



Figure 3. Hypernetwork example for mode, route, and destination choice with two market segments.




adequately handled by appropriate representation.

Before launching a more in-depth exploration of hypernetwork properties and applications, it is worthwhile to note that the idea of hypernetworks has been latent in the literature for some time. As early as 1972 at the Williamsburg Conference, Wilson (61) noted that

> It is tempting as computer capacity expands, to think of assigning on multimodal networks, in effect, possibly directly to routes on an abstract modal basis. . . . This is another example of a class of mathematical aggregation problems.

Manheim (62) formulated the transportation planning process as a network, using Dial's assignment method (63) to predict flow on the hypernetwork. However, since Dial's assignment algorithm is based on a logit formulation, it suffers from the IIA property, which, in route-choice problems, can be shown to produce unacceptable results (44, 64, 65, 66).

Dafermos suggested an integrated equilibrium flow model for transportation planning (67), based again on visualizing the whole transportation planning process as a solution to a network assignment problem. In her words,

> We adopt the natural behavioral assumption that each user chooses his origin, his destination, as well as his path as to minimize his "travel cost". Of course, "travel cost" should be interpreted in a very liberal fashion. In reality additional factors such as "attractiveness" of the origins (residential areas) and destinations (places of work) have to be taken into account but this can be incorporated into the model as "travel cost" by a straight-forward modification of the network. . . . Interestingly, we establish a mathematical equivalency which reduces integrated transportation problems for a network into assignment problems for a modified network.

Dafermos' model, although very similar to the hypernetwork concept, is not quite as general, because she was working exclusively with deterministic travel costs over the modified network. This explicitly excludes many demand models from the realm of application of her model, since, as it is assumed with deterministic equilibrium traffic assignment methods, users are identical (this excludes disaggregate demand models), fully informed (which excludes logit, probit, and other stochastic models) individuals making consistently perfect decisions.

The well-known elastic demand traffic assignment problem formulated by Beckmann, McGuire, and Winsten (68) can be solved with existing fixed demand traffic assignment problems on an expanded network (69). Such an expanded network can be viewed as a hypernetwork since it has dummy links going from each origin to each destination in order to represent the no-travel alternative.

It is also worth noting that traditional trip distribution models such as the entropy model can be cast as hypernetwork problems; Golob and Beckmann (70) showed the equivalence of entropy maximization and utility maximization over a hypernetwork similar to the one in Figure 3. Others (71, 72, 73, 74) contain mathematical programming formulations of the combined distribution-assignment problem that can also be regarded as hypernetwork problems.

The concept of a hypernetwork thus seems reasonable and flexible enough to be applied to many transportation forecasting problems. We shall now proceed to an analysis of the hyperpath choice process.

## HYPERPATH CHOICE AND MULTINOMIAL PROBIT

In this section we consider the choice process to be hyperpath when the attributes characterizing the disutilities of each link are given. We thus concentrate on the choice process and leave the more complex network equilibration issues to the next section.

The behavioral basis of the approach presented in this section is the economic concept of random utility (75, 76, 77, 78, 79). Early applications of the concept of random utility were suggested by McFadden (80) and Kohn, Manshi, and Mundel (81).

It has long been recognized from empirical studies that highway route choice is not a deterministic proposition, as seen, for instance, in the early traffic diversion studies (82, 83, 84). This gave rise to traffic assignment methods that do not allocate all the traffic to the shortest route (63, 85, 86, 87). These methods, however, are deficient (44, 64, 65, 66, 88), and only recently has a general theory of route choice been developed (44).

Others (44) have presented an analytically consistent method that accounts successfully for the topology of the networks. It is based on a probit model of route choice and it admits a straightforward generalization for hypernetworks with multiattributed link disutilities. This is done below.

Assume a general hypernetwork composed of links representing various dimensions of travel choice. Every link, i, in the network is associated with some disutility, $U_i'$ (we use primes for variables associated with links):

$$U_i' = V_i' + \xi_i' \quad (1)$$

where $V_i'$ is the measured (known) disutility, and $\xi_i'$ is a random error term distributed across the population. The vector of error terms, $\xi'$, is assumed to be multivariate normally (MVN) distributed with zero mean and a known variance-covariance matrix $\Sigma'$. Thus, using vector notation, $U' \sim \text{MVN}(V', \Sigma')$.

As an aside, note that the structure of the variance-covariance matrix $\Sigma'$ can only be decided on reasonableness grounds, as is the case with any specification decision. For instance, corr $(\xi_i'; \xi_j')$ can be set equal to zero if one can reasonably assume that links i and j are sufficiently unrelated as to be perceived independently by the decision maker. As argued by Daganzo and Sheffi (44), disutilities of links belonging to the street network can be considered independent. If this is not the case, there may be a representation of the problem that will admit independent link error terms. If the error terms on links OA' and OA" in Figure 2 are considered independent, the IIA property arises; however, the representation in Figure 1 overcomes this. Finding such a representation is equivalent to finding a reasonable specification of the problem.

The disutility minimization approach yields a multinomial probit (MNP) model of hyperpath choice, since the disutilities of hyperpaths are also MVN distributed. This can be seen by considering the hypernetwork link-route incidence matrix, $\Delta = \delta_{ik}$, where $\delta_{ik} = 1$, if link i belongs to hyperpath k, or 0, if otherwise. Letting U denote the vector of perceived hyperpath disutilities (the disutility vector of our choice model), we see that

$$U = U' \times \Delta \quad (2)$$

and $U \sim \text{MVN}(V, \Sigma)$ where $V = V' \times \Delta$ and $\Sigma = \Delta^T \times \Sigma' \times \Delta$ (T denotes the transportation operation).

Equation 2 implies that the covariance of two alternatives is heavily dictated by the amount of overlap of their corresponding hyperpaths. If $\Sigma'$ is diagonal, the covariance of two hyperpaths is given by the sum of the variances of the common links. Consequently, the topology of the hypernetworks bears directly on the probabilistic structure of the corresponding choice model.

Heterogeneous populations are well handled by hypernetworks, for, if we assume that the vector of attributes entering the disutility functions is MVN distributed across the population and that $\Sigma'$ is fixed, the choice process also follows an MNP law at the aggregated level (39). Non-normally distributed attributes can be handled by market segmentation (hypernetwork representation), as was done in Figure 3.

Of course, MNP models can be applied to choice problems without hypernetworks, but the graphical aid provided by visualizing the degree to which hyperpaths overlap helps to conceptualize reasonable parameters of the matrix $\Sigma$, especially for problems involving more than one stage of the traditional transportation planning process. Note, for instance, that the logit model arises from a random utility model with independent error terms and that it therefore corresponds approximately to hypernetworks with nonoverlapping routes. It should also be noted that, although many choice problems can be cast as hypernetwork problems, the use of MNP codes is basic to the analysis of general hypernetworks.

Hypernetworks thus present the same advantages and disadvantages of MNP models. Namely, MNP models and hypernetworks solve, or at least alleviate, the market segmentation problem, since an MNP model of individual choice is also MNP for the aggregate predictions (39). The step sequence issue (e.g., should mode choice be predicted before destination choice, after it, or simultaneously with it), posed by Ben-Akiva (89), who demonstrated the feasibility of a simultaneous approach, has already been addressed in this paper. Our approach is equivalent to a simultaneous MNP choice model whose covariance matrix can be studied visually.

Although hypernetworks enable us to visualize choice problems in a unified way and can help select appropriate probabilistic structures for choice models, their main advantage is that they will enable us to perform supply-demand equilibrium analysis on a mathematically consistent basis with disaggregate demand models. This subject is covered in the next section.

## EQUILIBRIUM

Although most of the work related to transportation equilibrium has been done in the context of traffic assignment, equilibrium analyses should not be restricted to route-choice situations, especially since in problems dealing with congested transportation facilities the supply considerations can be a more important determinant of use than the demand function itself. For instance, what good does it do to have a sophisticated demand model that predicts a transit park-ride ridership larger than what the access parking lots can accommodate?

Although the equilibrium problem has been addressed in the literature in connection with the traditional planning process (and as was mentioned earlier there is a rich literature dealing with heuristic iterative schemes involving feedback loops and accessibility measures), there is no general definition of equilibrium in the transportation market when the demand side is modeled as a probabilistic proposition. Because of this, the following definition is proposed. The equilibrium criterion is the condition such that "at equilibrium no user perceives that he or she can decrease his or her disutility by unilaterally changing alternatives."

This is a generalization of the stochastic user equilibration principle (44) of traffic assignment, which, in turn, is a generalization of Wardrop's rule (90).

The equilibrium solution is obtained by the solution of two systems of equations representing the demand and supply relationships. For a hypernetwork with only one origin, demand is

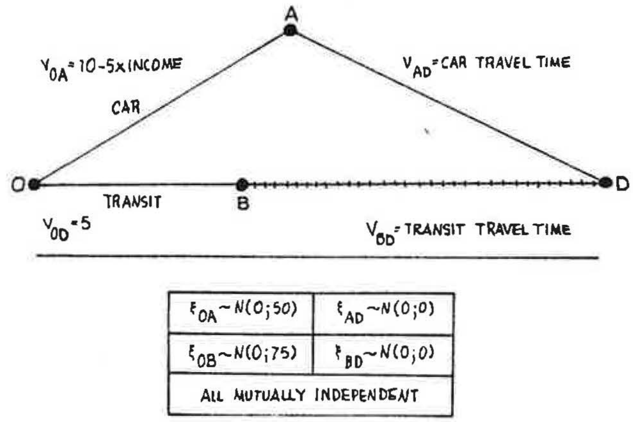
$$\Pr\{U_k < U_j; \forall j\} = X_k \Big/ \sum_{\forall j} X_j \leftrightarrow ; \forall k \quad (3)$$

where

- $X_k$ = number of users selecting alternative k,
- $\Sigma X_j$ = known fixed quantity (the population size), and
- $U_j$ = disutility of travel alternative j as perceived by an individual chosen at random from the population.

This equation is, of course, merely a statement of the weak law of large numbers. It states that the (predicted) market share of each alternative equals the choice probability for a randomly sampled choice maker, which is determined by the distribution function of the measured

Figure 4. Binary hypernetwork example.




disutility across a population.

If the latter distribution is approximately multivariate normal [Bouthelier and Daganzo (39) have identified some conditions for this], the distribution function is totally characterized by a vector of means V and a covariance matrix $\Sigma$, and in the general case one has the supply equations

$$V = V(X);\ \Sigma = \Sigma(X) \qquad (4)$$

This equation states that the vector of mean disutilities and the corresponding covariance matrix are functions of the use on each one of the alternatives; i.e., the total link flows on the hypernetwork. In instances where $\Sigma$ can be considered independent of X, supply modeling techniques can be used to determine Equation 4. Discussion of this point is, however, beyond the scope of this paper (54).

We now use a simple example with two alternatives and a known equilibrium solution to demonstrate how heuristic iterative techniques that alternatively solve Equations 3 and 4 fail to converge. The example is also used to demonstrate a new algorithm that is mathematically proved to converge.

Figure 4 displays a hypernetwork corresponding to an idealized modal split problem. The street network is represented by one link (AD); the inherent characteristics of the car mode are represented by a link leading into the street network (OA); and the transit alternative is similarly represented (links OB and BD). The measured link disutilities are written down by each link. The distribution of the link error terms is also specified in the figure.

The supply equations are defined for each link and expressed in travel time units: link OA = 10-5 income (independent of flow); link OD = 5 (independent of flow). Thus

$$\text{Car travel time} = 10/(1 - X_{AD}) \qquad (5A)$$

where $X_{AD}$ is the flow on link AD.

$$\text{Transit travel time} = 15 \quad \text{(independent of flow)} \qquad (5B)$$

Using Equation 2, the choice model corresponding to the hypernetwork turns out to be

$$U_{CAR} = 10 - 5 \text{ income} + \text{car travel time} + \xi_{CAR} \qquad (6A)$$

$$U_{TRANSIT} = 5 + \text{transit travel time} + \xi_{TRANSIT} \qquad (6B)$$

$$\text{with } \Sigma = \begin{Bmatrix} 50 & 0 \\ 0 & 75 \end{Bmatrix}$$

Equations 6A and 6B, however, are not ready for use because they do not represent the disutilities of a user sampled at random from the population. Thus, carrying out an aggregation procedure with, say, income being a normally distributed attribute with mean and variance equal to four yields, as the reader can check,

$$U_{CAR} = -10 + \text{car time} + \xi_{CAR} \qquad (7A)$$

$$U_{TRANSIT} = 5 + \text{transit time} + \xi_{TRANSIT} \qquad (7B)$$

$$\text{with } \Sigma = \begin{Bmatrix} 150 & 0 \\ 0 & 75 \end{Bmatrix}$$

If instead of a normal attribute, such as income, we had a zero-one variable, or market segmentation needed (39), one would have introduced additional origins representing the different market segments and would have connected them to A and B, as was done in Figure 3. In any case, Equations 7A and 7B yield for the probability of car choice:

$$P_{CAR} = \Pr\{U_{CAR} < U_{TRANSIT}\} = \Pr\{U_{CAR} - U_{TRANSIT} < 0\}$$
$$= \Phi[(15 + \text{transit time} - \text{car time})/(225)^{1/2}] \qquad (8)$$

where $\Phi$ is the standard normal cumulative distribution curve. Equation 8 of our example corresponds to Equation 3, and Equations 5A and 5B correspond to Equation 4.

Assuming that we are studying one unit of population, ($\sum_{\forall k} X_k = 1$), $P_{CAR} = X_{AD}$, and Equation 8 reduces to

$$X_{AD} = \Phi\{2 - [\tfrac{2}{3}/(1 - X_{AD})]\} \qquad (9)$$

Equation 9 can be solved graphically, as shown in Figure 5A, and yields $X_{AD} = 0.61$. Since in most instances it is impossible to reduce the original problem to a manageable set of equations (e.g., in multicommodity networks with just a few links), efficient numerical techniques must be sought, especially since heuristic iterative algorithms do not necessarily converge. This is shown below.

A typical heuristic procedure involving feedback loops consists of solving Equations 5A, 5B, and 8 alternatively, until a convergence criterion is met. In our case and to better illustrate the point, we select an initial value, $X_{AD} = 0.62$, very close to the equilibrium solution.

The table below displays the results obtained with the heuristic approach, which the reader can verify. The same iterative scheme could have been carried out graphically on Figure 5 to yield a familiar pattern known to economists as the cobweb model (Figure 6).

| Iteration | $X_{AD}$ | Iteration | $X_{AD}$ |
|---|---|---|---|
| 0 | 0.620 | 6 | 0.736 |
| 1 | 0.599 | 7 | 0.301 |
| 2 | 0.630 | 8 | 0.852 |
| 3 | 0.579 | 9 | 0.006 |
| 4 | 0.661 | 10 | 0.908 |
| 5 | 0.512 | 11 | 0 |

The approach developed to solve for the equilibrium in the transportation market has been formally described (54). It is computationally very efficient, can handle many origins on the hypernetwork, does not require that all the alternative routes be enumerated, and can take explicitly into account finite hypernetwork link capacities. This last feature proves to be invaluable for analysis of problems involving parking and other finite capacity transportation facilities.

Technically, the algorithm solves an associated disutility minimization program for the hypernetwork. The

**Figure 5. Numerical solution of the hypernetwork example.**

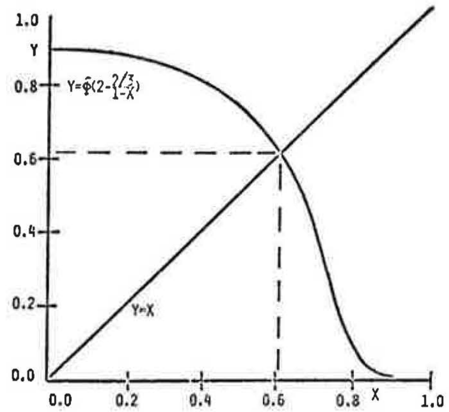
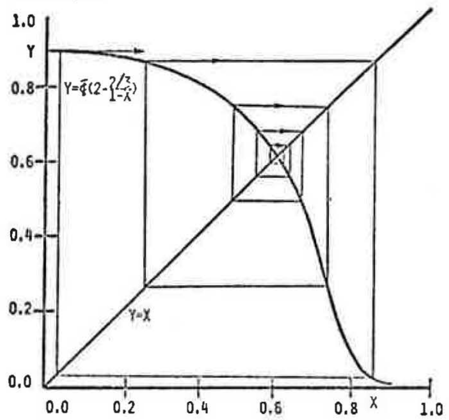



**Figure 6. Divergence pattern of a naive iterative technique.**




disutility additivity assumption mentioned in the preceding section is used to separate between the flow-dependent and flow-independent link disutilities using the optimality principle of dynamic programming. Then the assignment over the links associated with the flow-independent disutility is carried out by using analytical expressions for the flow allocation and the total disutility for all the populations of each origin zone. These quantities are computed at every iteration of the flow-dependent links equilibration to ensure a global minimum for the associated minimization program. A detailed description of the algorithm can be found in Sheffi (54) and a special application of it to the spatial aggregation problem of traffic assignment in Daganzo (91).

The major assumptions of the methodology include the above additivity of the flow-dependent component in all the disutility functions. For example, if car travel time over the street network is modeled as flow dependent, it has to enter the disutility functions in a generic linear-additive form. Furthermore, each hypernetwork link can be modeled as stochastic, as multiattributed, or as flow dependent (54). Another important limitation on our methodology is that the covariance matrix of the multivariate distribution underlying the MNP models involved has to be independent of the associated vector of means, otherwise Equation 3 does not follow an MNP model (39, 54).

For our particular example, with a relatively poor initial point $X_{AD} = 0.5$, the algorithm converges in six iterations. The table below displays the results.

| Iteration | $X_{AD}$ | Iteration | $X_{AD}$ |
|---|---|---|---|
| 0 | 0.5 | 4 | 0.637 |
| 1 | 0.748 | 5 | 0.571 |
| 2 | 0.664 | 6 | 0.610 |
| 3 | 0.555 | 7 | 0.610 |

## DISCUSSION AND SUMMARY

This paper has presented a framework for carrying out supply-demand equilibration of a transportation market. It is argued that choice problems can be regarded as route-choice problems on abstract hypernetworks. This idea has been latent in the literature, and, as shown, it is intimately related to multinomial probit models. Hypernetworks may help us conceptualize useful parameters of MNP models, especially for the covariance matrix, with not too many additional parameters.

We addressed the important issue of equilibrium analysis by providing a formal definition of equilibrium. It is also shown, by means of a counterexample, that heuristic iterative algorithms do not necessarily converge to the equilibrium solution and that a newly developed, efficient mathematical algorithm does indeed converge.

The example and the paper do not include a specific discussion of the spatial aggregation problem, but, as shown elsewhere (39, 91), the intrazonal level-of-service attributes can be well approximated with MVN distributions, and therefore MNP models can be used to predict the loading of the network. This obviates the need for centroids and dummy links.

Although most of the emphasis of our discussion was on the urban transportation planning process, microscopic problems can also be represented as hypernetworks and handled with the above-mentioned equilibrium algorithm. Typical examples of applications would be finding equilibrium on a dial-a-ride market. Supply equations for these systems have recently been developed with queueing theory (92). Other examples are selection of parking lots in a downtown area as a function of their capacities and locations, airport selection as a function of an intercity origin-destination table, and the location and service characteristics of the individual airports.

In summary, hypernetwork theory seems to be a flexible and powerful methodology that enables us to directly address some problems that had not been previously tractable.

## REFERENCES

1. Urban Transportation Planning—General Information, Federal Highway Administration, U.S. Department of Transportation. Rept., March 197?.
2. UMTA Transportation Planning System (UTPS) Reference Manual. Urban Mass Transportation Administration Planning Methodology and Technical Support Division, U.S. Department of Transportation, Aug. 1974.
3. D. Brand. Travel Demand Forecasting: Some Foundations and a Review. HRB, Special Rept. 143, Dec. 1973, pp. 239-282.
4. A. A. Douglas and R. J. Lewis. Trip Generation Techniques. Traffic Engineering and Control, Vol. 12, No. 10, Nov. 1970.
5. Guidelines for Trip Generation Analysis, U.S. Department of Transportation, Office of Planning, Bureau of Public Roads, June 1967.
6. A. G. Wilson. A Statistical Theory of Spatial Distribution Models. Transportation Research, Vol. 1, No. 3, 1967.

7. P. Loubal and R. Potts. A Mathematical Model for Trip Distribution. Univ. of California, Berkeley, 1970.
8. J. M. McLynn and others. Analysis of a Market Split Model. Journal of Research of National Bureau of Standards, Vol. 72B, No. 1, 1968.
9. D. A. Quarmby. Choice of Travel Mode for the Journey to Work: Some Findings. Journal of Transportation Economics and Policy, Vol. 1, No. 3, 1967.
10. B. V. Martin and M. L. Manheim. A Research Program for Comparison of Traffic Assignment Techniques. HRB, Highway Research Record 88, 1965, pp. 69-84.
11. M. J. Huber, H. B. Boutwell, and D. K. Witheford. Comparative Analysis of Traffic Assignment Techniques with Actual Highway Use. NCHRP Rept. 58, 1968.
12. M. L. Manheim and E. M. Ruiter. DODOTRANS 1: A Decision-Oriented Computer Language for Analysis of Multi-Mode Transportation Systems. HRB, Highway Research Record 314, 1970, pp. 135-163.
13. Peat, Marwick, Mitchell. A Review of Operational Urban Transportation Models, Final Report. Cambridge, MA, Rept. No. DOT-TSC-496, 1973.
14. R. Bolsan. New Rules for Judging Analytical Techniques in Urban Planning. In Analytical Techniques, American Society of Planning Officials, Chicago, 1970, pp. 75-84.
15. P. O. Roberts. Model System for Urban Transportation Planning: Where Do We Go From Here? HRB, Highway Research Record 309, 1970, pp. 34-44.
16. W. Alonzo. Predicting Best With Imperfect Data. Journal of the American Institute of Planners, Vol. 34, 1968, pp. 248-255.
17. D. B. Lee. The Future of Models in Transportation. Paper presented at the 1971 Annual Meeting of the American Institute of Planners, San Francisco, Oct. 1971.
18. D. B. Lee. Requiem for Large-Scale Models. Journal of the American Institute of Planners, Vol. 39, No. 3, May 1973, pp. 163-178.
19. H. Cassoff and H. Deutschman. Trip Generation: A Critical Appraisal. HRB, Highway Research Record 297, 1969, pp. 15-30.
20. R. J. Lewis. Comments on Regression Techniques in Trip Generation Analysis. Urban Traffic Model Research Symposium, Planning and Transport Research and Computation (International), June 1970.
21. D. Hartgen and G. Tanner. Behavioral Models of Mode Choice. New York State Department of Transportation, Albany, March 1970.
22. Charles River Associates. A Model of Urban Passenger Travel Demand in the San Francisco Metropolitan Area. Rept. prepared for the California Division of Bay Toll Crossings, San Francisco, 1967.
23. D. Brand. Theory and Methods in Land Use and Travel Demand Forecasting. Paper presented at the 51st Annual Meeting, HRB, Jan. 1972.
24. M. Manheim. Fundamental Properties of Systems of Demand Models. MIT, Cambridge, MA, MIT Discussion Paper T70-1 (initial draft), Nov. 1970.
25. R. G. McGillivray. Some Problems With Urban Transportation Planning Models. Traffic Quarterly, Vol. 26, No. 4, Oct. 1972, pp. 547-559.
26. C. A. Lave. A Behavioral Approach to Modal Split Forecasting. Transportation Research, Vol. 3, No. 4, Dec. 1969, pp. 463-480.
27. S. Reichman and P. Stopher. Disaggregate Stochastic Models of Travel Mode Choice. HRB, Highway Research Record 369, 1971, pp. 91-103.
28. M. G. Richards. Disaggregate, Simultaneous Urban Travel Demand Models: A Brief Introduction. Transportation, Vol. 3, No. 4, Dec. 1974, pp. 335-342.
29. Charles River Associates. A Disaggregate Behavioral Model of Urban Travel Demand. Federal Highway Administration, U.S. Department of Transportation, 1972.
30. P. R. Stopher and T. E. Lisco. Modelling Travel Demand: A Disaggregate Behavioral Approach, Issues and Applications. In Transportation Research Forum, Richard B. Cross, Oxford, IN, 1970, pp. 195-214.
31. D. McFadden. The Measurement of Urban Travel Demand. Journal of Public Economics, Vol. 3, 1974, pp. 303-328.
32. T. J. Adler and M. Ben-Akiva. Joint-Choice Model for Frequency, Destination, and Travel Mode for Shopping Trips. TRB, Transportation Research Research Record 569, 1975, pp. 136-150.
33. M. Ben-Akiva and S. Lerman. A Disaggregate Behavioral Model of Automobile Ownership. Paper presented at the 54th Annual Meeting, TRB, Jan. 1975.
34. S. R. Lerman. Location Housing, Auto Ownership and Mode to Work: A Joint Choice Model. MIT, Cambridge, PhD thesis, Aug. 1975.
35. P. Watson and others. Factors Influencing Shipping Mode Choice of Intercity Freight: A Disaggregate Approach. Proc., 16th Annual Meeting of the Transportation Research Forum, 1975.
36. F. S. Koppelman. Guidelines for Aggregate Travel Prediction Using Disaggregate Choice Models. Paper presented at the 55th Annual Meeting, TRB, Jan. 1976.
37. R. B. Westin. Prediction From Binary Choice Models. Journal of Econometrics, Vol. 2, 1974, pp. 1-16.
38. D. McFadden and F. Reid. Aggregate Travel Demand Forecasting From Disaggregated Behavioral Models. TRB, Transportation Research Record 534, 1975, pp. 24-37.
39. F. Bouthlier and C. F. Daganzo. Aggregation With Multinomial Probit and Estimation of Disaggregate Models With Aggregate Data: A New Methodological Approach. Paper presented at the 57th Annual Meeting, TRB, Jan. 1978.
40. D. McFadden. Conditional Logit Analysis of Qualitative Choice Behavior. In Frontiers in Econometrics (P. Zarembka, ed.), Academic Press, New York, 1973.
41. T. Domencich and D. McFadden. Urban Travel Demand: A Behavioral Analysis. American Elsevier, New York, 1975.
42. J. Mayberry. Structural Requirements for Abstract-Mode Models of Passenger Transportation. In The Demand for Travel, Theory and Measurement (R. Quandt, ed.), Heath, Lexington, MA, 1970.
43. Y. Sheffi. Estimating Choice Probabilities Among Nested Alternatives. Transportation Research, in preparation.
44. C. F. Daganzo and Y. Sheffi. On Stochastic Models of Traffic Assignment. Transportation Science, Vol. 11, Aug. 1977, pp. 253-274.
45. S. H. Putman. Further Results From the Integrated Transportation and Land Use Package (ITLUP). Transportation Planning and Technology, Vol. 3, 1976, pp. 165-173.
46. M. Ben-Akiva and S. R. Lerman. Disaggregate Travel and Mobility Choice Models and Measures

of Accessibility. Paper presented at 3rd International Conference on Behavioral Travel Modelling, Australia, April 1977.
47. M. Q. Dalvi. The Measurement of Accessibility: Some Preliminary Results. Transportation, Vol. 5, No. 1, March 1976.
48. L. Sherman, B. Barber, and W. Kondo. Method for Evaluating Metropolitan Accessibility. TRB, Transportation Research Record 499, 1974, pp. 70-82.
49. G. C. Nicolaidis, M. Wachs, and T. F. Golob. Evaluation of Alternative Market Segmentations for Transportation Planning. Paper presented at 56th Annual Meeting, TRB, Jan. 1977.
50. H. Lovelock. A Market Segmentation Approach to Transit Planning. Proc., 16th Annual Meeting of Transportation Research Forum, 1975, pp. 247-258.
51. C. F. Daganzo, F. Bouthelier, and Y. Sheffi. An Efficient Approach to Estimate and Predict With Multinomial Probit Models. Paper presented at 56th Annual Meeting, TRB, Jan. 1977.
52. C. F. Daganzo, F. Bouthelier, and Y. Sheffi. Multinomial Probit and Qualitative Choice: A Computationally Efficient Algorithm. Transportation Science, Vol. 11, No. 4, Nov. 1977, pp. 338-358.
53. D. McFadden. A Closed-Form Multinomial Choice Model Without the Independence From Irrelevant Alternatives Restrictions. Urban Travel Demand Forecasting Project, Institute for Transportation Studies, Univ. of California, Berkeley, Working Paper No. 7703, 1977.
54. Y. Sheffi. Transportation Networks Equilibration With Discrete Choice Models. Civil Engineering Dept., MIT, Cambridge, PhD thesis, May 1978.
55. S. Nguyen. An Algorithm for the Traffic Assignment Problem. Transportation Science, Vol. 8, 1977, pp. 203-216.
56. L. J. LeBlanc, E. Morlok, and W. Pierskalla. An Efficient Approach to Solving the Road Network Equilibrium Traffic Assignment Problem. Transportation Science, Vol. 9, 1975, pp. 309-318.
57. E. Ruiter and M. Ben-Akiva. A System of Disaggregate Travel Demand Models, Structure, Component Models, and Application Procedures. Paper presented at 57th Annual Meeting, TRB, Jan. 1978.
58. J. Jacobson. Case Study Comparison of Alternative Urban Travel Forecasting Methodologies. Civil Engineering Dept., MIT, Cambridge, MS thesis, 1977.
59. A. Talvitie and I. Hasan. An Equilibrium Model System for Transportation Corridor and Its Application. Paper presented at the International Symposium of Travel Supply Models, Montreal, Nov. 1977.
60. H. Scarf. The Computation of Economic Equilibrium. Cowels Foundation Monograph No. 24, 1973.
61. A. G. Wilson. Travel Demand Forecasting: Achievements and Problems. HRB, Special Rept. 143, 1973.
62. M. Manheim. Proposed Direct Equilibrium Approach Using Exponential Functions. Memorandum to R. Dial, Sept. 1973.
63. R. B. Dial. A Probabilistic Multipath Traffic Assignment Model Which Obviates Path Enumerations. Transportation Research, Vol. 5, No. 2, 1971, pp. 83-111.
64. M. Schneider. Probability Maximization in Networks. Proc., International Conference on Transportation Research, Bruges, Belgium, 1973, pp. 748-755.
65. M. Florian and B. Fox. On the Probabilistic Origin of Dial's Multipath Traffic Assignment Model. Transportation Research, Vol. 10, No. 6, 1976, pp. 339-341.
66. J. Burrell. Multipath Route Assignment: A Comparison of Two Methods. In Traffic Equilibrium Methods: Lecture Notes in Economics and Mathematical Systems, Springer Verlag, New York, Vol. 118, 1976.
67. S. Dafermos. Integrated Equilibrium Flow Models for Transportation Planning. In Traffic Equilibrium Methods: Lecture Notes in Economics and Mathematical Systems, Springer Verlag, New York, 1976.
68. M. Beckmann, C. B. McGuire, and C. Winsten. Studies in the Economics of Transportation. Yale Univ. Press, New Haven, 1956.
69. G. B. Dantzig, S. F. Maier, and Z. F. Lansdowne. The Application of Decomposition to Transportation Network Analysis. U.S. Department of Transportation, Rept. DOT-TSC-OST-76-26, Oct. 1976.
70. T. F. Golob and M. J. Beckmann. On the Metaphysical Foundations of Traffic Theory: Entropy Revisited. In Traffic Flow and Transportation (G. F. Newell, ed.), Elsevier, New York, 1972, pp. 109-118.
71. M. Bruynooghe. Un Modele Integre de Distribution et d'Affectation de Trafic sur un Reseau. Rept. from Institut de Recherche des Transports, Departement de Recherche Operationnelle et Informatique, Arcueil, France, 1969.
72. J. A. Tomlin. A Mathematical Programming Model for the Combined Distribution—Assignment of Traffic. Transportation Science, Vol. 5, 1971, pp. 122-140.
73. S. P. Evans. Some Models for Combining the Trip Distribution and Traffic Assignment Stages in the Transport Planning Process. In Traffic Equilibrium Methods (M. J. Beckmann and H. P. Kunzi, eds.), Springer Verlag, New York, 1976.
74. M. Florian, S. Nguyen, and J. Ferland. On the Combined Distribution—Assignment of Traffic. Centre de Recherche sur les Transports, Univ. of Montreal, Publ. No. 8, 1974.
75. J. Marschak. Binary Choice Constraints and Random Utility Indicators. In Mathematical Models in the Social Sciences (K. J. Arrow and others, eds.), Stanford Univ. Press, Mathematical Studies in the Social Sciences, No. 4, 1960.
76. R. D. Luce and P. Suppes. Preference, Utility and Subjective Probability. In Handbook of Mathematical Psychology (R. D. Luce, ed.), Wiley, New York, Vol. 3, 1965, pp. 249-410.
77. R. Bock and L. Jones. The Measurement and Prediction of Judgement and Choice. Holden-Day, New York, 1968.
78. A. Tverski. Choice by Elimination. Journal of Mathematical Psychology, Vol. 9, 1972, pp. 341-367.
79. C. F. Manski. The Analysis of Qualitative Choice. Department of Economics, MIT, Cambridge, PhD dissertation, June 1973.
80. D. McFadden. The Revealed Preference of a Government Bureaucracy. Institute for International Studies, Univ. of California, Berkeley, Technical Paper No. 17, Nov. 1968.
81. M. Kohn, C. Manski, and D. Mundel. An Empirical Investigation of Factors Influencing College-Going Behavior. Annals of Economic and Social Measurement, Vol. 5, No. 4, 1976, pp. 391-419.
82. A. D. May and H. L. Michael. Allocation of Traffic to Bypasses. HRB, Bulletin No. 61, 1955.
83. E. W. Campbell and R. S. McCorgar. Objective

and Subjective Correlates of Expressway Use. HRB, Bulletin No. 119, 1956.
84. E. Moskowitz. California Method for Assigning Diverted Traffic to Proposed Freeways. HRB, Bulletin No. 130, 1956.
85. J. E. Burrell. Multipath Route Assignment and Its Applications to Capacity Restraint. Proc., 4th International Symposium on the Theory of Traffic Flow, Karlsruhe, Germany, 1968.
86. R. L. Tobin. An Extension of Dial's Algorithm Utilizing a Model of Tripmakers' Perception. Transportation Research, Vol. 11, 1977, pp. 337-342.
87. H. Von Falkenhausen. Traffic Assignment by a Stochastic Model. Proc., 4th International Conference on Operational Science, 1966, pp. 415-421.
88. Y. Sheffi. A Note on the Turn and Arrival Likelihood Algorithms of Traffic Assignment. Transportation Research, in preparation.
89. M. Ben-Akiva. Structure of Passenger Travel Demand Models. TRB, Transportation Research Record 526, 1974, pp. 26-42.
90. J. G. Wardrop. Some Theoretical Aspects of Road Traffic Research. Proc., Institution of Civil Engineers, London, Part II: Research and Theory, Vol. 1, 1952, pp. 325-378.
91. C. F. Daganzo. Network Representation, Continuum Approximations and a Solution to the Spatial Aggregation Problem of Traffic Assignment. Stochastic Traffic Assignment Project, Institute for Transportation Studies, Univ. of California, Berkeley, Working Paper No. 7702, 1977.
92. C. F. Daganzo, C. Hendrickson, and N. Wilson. An Analytic Model of Many-to-One Demand Response Transportation Systems. Proc., 7th International Symposium on the Theory of Traffic Flow and Transportation, Kyoto, 1977.

*Publication of this paper sponsored by Committee on Passenger Travel Demand Forecasting and Committee on Traveler Behavior and Values.*

# Disaggregate Travel Demand Models for the San Francisco Bay Area

## System Structure, Component Models, and Application Procedures

Earl R. Ruiter, Cambridge Systematics, Inc., Cambridge, Massachusetts
Moshe E. Ben-Akiva, Massachusetts Institute of Technology and Cambridge Systematics, Inc., Cambridge

**Significant advances have recently been made in developing and applying disaggregate behavioral travel demand models to many aspects of urban travel decisions. What has not previously been developed is a full set of urban models integrated into a complete forecasting system for use by a metropolitan planning organization. The purpose of this paper is to describe the first such system, which was developed for the Metropolitan Transportation Commission, the designated metropolitan planning organization for the San Francisco area. First, the background of the current modeling project is briefly set out, followed by a description of the structure of the model system. The model development process—estimation, prediction testing, and validation—is described, and two computerized model application procedures—a regional network analysis system compatible with available urban transportation planning packages and a generalized policy analysis system based on random sample forecasting—are presented. Conclusions concerning the advantages and disadvantages of the system of disaggregate models are presented.**

Significant advances have recently been made in developing and applying disaggregate behavioral travel demand models to many aspects of urban travel decisions (1, 2, 3, 4). What has not previously been done is the development of a full set of urban models and their integration into a complete forecasting system for use by a metropolitan planning organization. The purpose of this paper is to describe the first such system, which was developed for the Metropolitan Transportation Commission (MTC), the designated metropolitan planning organization for the San Francisco area. The models have been developed by the Travel Model Development Project, carried out by a consultant team consisting of the COMSIS Corporation, Cambridge Systematics, Inc., and Barton-Aschman Associates. Because of the number of models included in the system, this paper must be an overview of the system as a whole, rather than a detailed description of each individual model. Complete documentation of the project is available in a three-volume final report (5).

### BACKGROUND

The MTC is the successor to the Bay Area Transportation Study Commission (BATSC), the original region-wide multimodal transportation planning agency in the Bay Area. Although BATSC carried out the traditional first steps of metropolitan transportation planning—collecting and analyzing land-use and travel data—neither it nor MTC was previously successful in developing an accepted complete land-use and travel modeling system that could be used to forecast future travel patterns. In cooperation with the Association of Bay Area Governments (ABAG), the projective land use model (PLUM) was developed to forecast future land use, employment location, residential location, and socioeconomic characteristics (6).

Although at least two generations of trip-making models that use these forecasts to predict future travel have been developed, both MTC and other Bay Area agencies have been reluctant to use them because of deficiencies in their ability to represent existing travel or to provide reasonable future estimates. The lack of such a travel model system led MTC to fund the Travel Model Development Project and to select an approach to develop a model system that would be, as much as possible, based on disaggregate travel demand modeling techniques.

The domain of the models to be developed was clear. They should begin where PLUM ends, using PLUM's land-use and other travel-related forecasts as input; they should deal with all aspects of urban passenger travel, including the assignment of transit person trips and highway vehicle trips to the appropriate facilities; they should provide the ability to conduct areawide planning studies.

The data base for model development also became clear; although many special purpose surveys collected since 1965 cover specific subareas of the region, the 1965 BATSC travel surveys remain the most recent complete travel data set. In addition, all necessary related data—highway and transit networks, zone level-of-service data (access times and distances and parking costs, for example), and land-use data (obtained largely from "backcasts" to 1965, using PLUM, of data collected in 1970)—were available for 1965. All such data were also available with a common zoning system of 290 zones and 30 districts.

## MODEL SYSTEM REQUIREMENTS

Against this background, the following major requirements for the MTC travel demand model system were identified.

1. Validity: The system must accurately represent travel behavior, which occurs as a result of an interconnected set of household and individual decisions;
2. Comprehensiveness: The system must represent the full range of urban travel decisions;
3. Policy relevance: The system must be sensitive to all relevant policy options; and
4. Flexibility: The system must be usable for analysis at varying levels of detail and spatial and time scales.

Because disaggregate travel demand modeling techniques provide significantly improved capabilities to meet each of these requirements, they have been used to develop the MTC model system. Disaggregate models estimated at the household, person, or trip level are used for aggregate forecasting at the zone level. The theoretical and statistical advantages of this approach over conventional aggregate modeling techniques have been extensively discussed in the literature (7). The major characteristics of the resulting model system meet the above requirements in a system that is (a) an operational areawide transportation planning tool, (b) based entirely on disaggregated travel demand models, and (c) estimated by using conventional urban transportation study data.

The remainder of this paper describes the key features of the model system and emphasizes unique characteristics and major improvements over a typical conventional system. The discussion focuses on two major areas: first, the model system structure—the interrelations among travel choices and the structure of individual models—and, second, the model development process—the process of empirical model estimation and testing, and the techniques used for aggregate forecasting. The paper concludes with a summary of the key advantages and disadvantages of the disaggregate approach to travel modeling.

## MODEL SYSTEM STRUCTURE

### Travel Choices Represented

In general, a travel demand model is concerned with those household and individual decisions that result in trips being made. However, some other choices are so strongly interrelated with actual trip-making choices that it is impossible to separate them from such decisions. For example, the choice of residential location is not in itself a trip-making decision. However, the combination of a worker's employment location choice and his or her household's location decision has as its consequence a trip choice, i.e., daily work trips.

For this reason, the general framework from which the components of the MTC model system are derived begins with a partition of all possible household and household member decisions into two sets: those that are relevant to transportation analysis and those that, for practical purposes, can be ignored. This partition produces the following set of travel-related household decision choices:

1. Employment location (for all workers),
2. Residential location,
3. Housing type,
4. Automobile ownership,
5. Mode to work (for all workers),
6. Frequency (for nonwork trips of each purpose),
7. Destination (for nonwork trips of each purpose),
8. Time of day (for nonwork trips of each purpose),
9. Mode (for nonwork trips of each purpose), and
10. Route (for all trips).

For most situations, this set of decisions must be represented in a complete model system. In theory, each decision may be dependent on the rest. For example, where one chooses to live is obviously linked to the housing type and the level of automobile ownership one selects. Similarly, shopping trip destination and mode are likely to be closely linked.

This perspective, if carried through completely, would produce a model of unmanageable dimensions, since all possible combinations of choices would, for practical purposes, be limitless; useful models would then be impossible to develop. Fortunately, there are some interrelationships among components of this set that are of a fundamentally different character than others. Some of the decisions, such as residential location choice, have high transaction costs and are consequently stable over fairly long intervals; other choices, such as the frequency of social and recreational trips, are altered on a daily basis. Some decisions are more logically represented as being made collectively by the household, while others can be approximated as individual choices. Thus, it is possible to formulate explicit behavioral hypotheses and to establish a structure of the total set of choices as a logical working hypothesis. Such an explicit structure greatly simplifies model development. This structure is termed a hierarchy of choice (8).

Figure 1 illustrates the three-stage choice hierarchy represented by the MTC model system. At the highest level are urban development decisions, which are long run in nature: employers decide where to provide jobs and developers decide where to provide housing of various types. Next come household mobility decisions made more frequently; these include where to live and work, how many household members will have jobs, how often they each will go to work, how many autos to

Figure 1. A three-stage choice hierarchy.

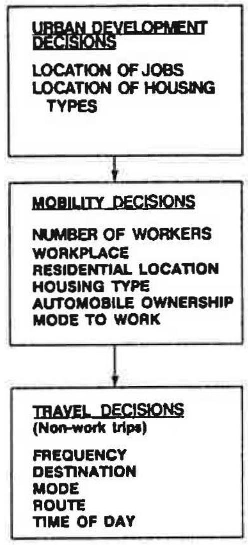
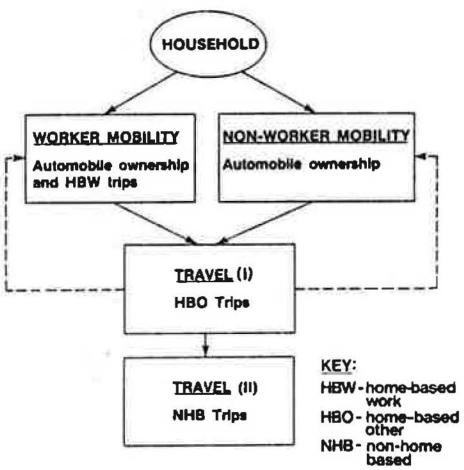



Figure 2. Overall structure of the MTC model system.




own, and which modes will be used to make work trips. Finally come short-run travel decisions made almost daily: frequency, destination, and mode for nonwork trips, and time of day and route for all trips.

The PLUM land-use model used by MTC predicts each of the development decisions shown in Figure 1, plus the residential location decision. The travel demand models described in this paper predict each of the remaining household decisions.

Figure 2 shows in greater detail how the mobility and travel choice levels of this choice hierarchy are represented in the MTC model system; the mobility decisions of households with and without workers are handled separately, and travel decisions are divided into two groups. Home-based other travel (nonwork) is predicted according to all home-based travel decisions. The dotted lines indicate that there is also an indirect provision for nonwork travel (HBO) to affect mobility decisions.

Although the structure shown in Figure 2 is strongly related to the conceptual system of disaggregate travel demand models shown in Figure 1, it incorporates a number of features and approximations necessary for producing a practical regional forecasting system. These include the following.

1. The residential location and housing type choices are external to the present model development effort because they are predicted by the previously developed PLUM system, which also predicts income distributions and work force by zone.
2. The mobility choice block for households with workers distinguishes between primary and secondary workers in a household. Each household with workers has only one primary worker, or breadwinner. All additional workers are termed secondary.
3. The modeling system deals separately with home-based and non-home-based trips. This simplifies the representation of trip chains (a trip from home, followed by one or more non-home-based trips, followed finally by a trip to home), an area in which basic conceptual development is continuing (9). Also, it allows the model system to deal with one-way trips, in accordance with traditional practice, rather than with the roundtrips more commonly considered in disaggregate modeling.
4. For a number of closely related travel decisions for which joint models have been previously predicted, a series of sequential rather than joint models has been developed. Examples are auto ownership and mode choice for primary workers, and nonwork trip frequency, destination, and mode choice. However, due to the structures of these sequential models, joint effects are not ignored.
5. There are exceptions to the hierarchy indicated by the solid-line arrows connecting the choice blocks. These are shown by the dotted-line arrows. Each of these represents an accessibility-like variable in the higher level model (auto ownership for households without workers, for example) that is obtained from a lower level model (home-based other destination and mode choice). Each of these variables is based on the full set of variables of the lower level model. These variables allow consistent representation of level-of-service effects in spite of the sequential structure of the model system.
6. The time-of-day decision is modeled by using historical peaking characteristics rather than a choice model based on the relationship between peak and off-peak transportation system characteristics.
7. The vehicle occupancy choice decision for nonwork travel is made by using historically observed rates rather than disaggregate choice models.
8. The route-choice decision is modeled by using conventional capacity restraint assignment techniques.

Before describing the models in each of the choice blocks illustrated in Figure 2, the definitions of the various trip purposes and modes will be given, the nature of the linkages implied by the dashed lines in the figure will be made explicit, and the types of independent variables used in the system will be described.

### Trips Represented

The trip purposes used in model development are: home-based work trips (HBW), or all trips between home and work; home-based other trips (HBO), or home-based nonwork travel represented by two sets of models, one for generalized shopping trips (including also medical-dental, business-related, and serving passenger purposes—HBSH), and one for social-recreational trips (including eating, visiting, and recreation purposes—HBSR); and non-home based (NHB), or all trips that do not begin or end at home. These three purpose groups include all surveyed trips except school trips.

The modes considered in the models include auto and pickup drivers and passengers, as well as all bus, streetcar, railroad, and jitney trips. Trips by trucks and taxi drivers and passengers, and by walkers, are not represented in the models.

## Linkages Between the Models

The components of the MTC model system are linked in two ways: first, "lower level" models are conditional on the predicted choices by "higher level" models, as indicated by the solid arrows in Figure 2; second, feedback in the form of composite or accessibility variables are calculated by using lower level models and are included in higher level models, as indicated by the dotted arrows.

The first type of linkage is determined by the assumed choice hierarchy and the resulting sequence of models. Variables resulting from higher level choices are predetermined for lower level choices and are attributes of the household or the individual that do not vary among alternative lower level choices. For example, auto ownership is treated as a household characteristic in the HBO models.

The composite variables represent expectations of the outcomes of lower level choices that could be different among alternatives of higher level choices. For example, level of service by transit for shopping trips affects auto ownership. However, this variable depends on the household choice of shopping trips, a decision made only conditional on the household auto ownership. Thus, the specific shopping level of service is indeterminate in the choice of auto ownership. However, composite variables representing overall shopping level of service for alternative auto ownership levels can be determined. The attributes that vary among lower level choices are aggregated and included as composite variables in the models of higher level choices.

All systems of travel demand models include, to some extent, such composite variables. Examples would be weighted by "inclusive prices" (3) and by the transit accessibility variable used in a trip generation model by the Metropolitan Washington Council of Governments (10). However, the formulation of these composite variables is often arbitrary and results in counterintuitive predictions. The composite variables used in the MTC model system are derived in a way that is consistent with the underlying assumption of the models.

If a lower level choice is modeled by using the logit model, the composite variable defined over these choice alternatives is constructed as the expected maximum utility from this choice process. If the outcome of this choice were known, then the composite variable would be taken as the expected utility of the chosen alternative. However, since it is assumed, in developing a choice model such as logit, that the alternative with the highest utility is selected, then we can calculate the expected value of the maximum utility. For the logit model, this is equal to the natural logarithm of the denominator of the logit probability model (11).

Because the denominator of the logit function must be computed in forecasting anyway, the use of this expression as a variable in related models requires little or no additional computation. In the MTC model system such variables were used in several models as the expected maximum utility from the entire set of alternatives or as the expected maximum utility from a subset of the alternatives. For example, the shopping trip frequency model includes as an independent variable the denominator of the shopping destination and mode-choice model. The auto ownership model includes the ratios of expected maximum utilities from shopping travel by auto and transit for different auto ownership levels, which are calculated by using appropriate partial sums of the denominator. Larger values of this ratio indicate a greater need for a car for shopping travel, which will therefore result in increased auto ownership.

## Variables Included

Disaggregate model estimation techniques provide greater efficiency in the use of survey data than aggregate techniques do. As a result, more variables can be included in the model system, thereby increasing the sensitivity of travel forecasts to changes in the urban environment and in government policies.

Figure 3 presents a summary of the independent variables in the MTC model system in terms of the submodels of a conventional system. The independent variables are classified into four groups:

1. Highway level of service (auto travel time and out-of-pocket cost),
2. Transit level of service (fare and wait time),
3. Land use (retail employment), and
4. Socioeconomic attributes of potential travelers (annual income).

These variables affect all the travel-related choices that were described in the previous section. Figure 3 indicates the classes of variables typically omitted from various submodels of conventional systems and the improvements in sensitivity to level of service (LOS) and socioeconomic characteristics in the disaggregate model system. Examples of improved model sensitivity include the impacts of level of service by auto and transit on HBO trip generation and NHB travel and the effect of socioeconomic characteristics on trip distribution. The added sensitivity to LOS characteristics permits a more credible forecasting of the consequences of pricing policies, auto restraint measures, and other service changes that affect not only modal split but generation and distribution as well.

# COMPONENT MODELS

In the light of the general material presented in the previous sections, each of the four travel choice blocks shown in Figure 2 can be discussed by presenting the component models and showing their interrelationships.

## Worker Mobility Models

Figure 4 displays the details of the worker mobility models, which include a full set of HBW models and a model of auto ownership for households with workers. Workers are differentiated into primary (one per working household, chosen on the basis of income per worker) and secondary (all other workers) groups. For each, sequential models of trip frequency, work place choice (distribution), and mode choice are provided. The work place choice models use accessibility terms for each destination by all modes obtained from the mode-choice models. Household auto ownership is affected by the work place choice of the primary worker, by means of an expected utility variable that measures the relative ease of traveling to this destination by auto and transit. In addition, the relative ease of traveling to all shopping destinations by auto and transit is allowed to influence household auto ownership levels.

## Nonworker Mobility Models

Because they make no work trips, only auto ownership is predicted for households without workers. The model uses information on the relative accessibility to shopping destinations by auto and transit, as well as per capita household income and residential density measures to predict probabilities of owning a given number of autos.

## Travel Models I: Home-Based Other

As shown in Figure 5, for each home-based other trip purpose—shopping and social-recreational—two models exist. These models predict trip frequency and the joint choice of destination and mode. Trip frequency is dependent on the auto ownership predicted in the mobility blocks, the autos remaining after all work travel is predicted, and the expected utility of travel to all available destinations by either auto or transit. This structure allows the amount of nonwork travel to vary as auto use for work travel varies, and as the level of service in the transportation system varies.

## Travel Models II: Non-Home Based

Figure 6 presents the structure of the NHB models. Home-based trip attractions by mode are put directly into a joint model of NHB frequency and destination choice. The frequency decision is binary; i.e., either the traveler goes home (frequency = 0) or he or she makes an NHB trip (frequency = 1) to one of the destination alternatives in the choice set. The details of these NHB models are in another paper in this Record by Ben-Akiva, Sherman, and Kullman. Explicit modeling of mode choice can be omitted if the observed frequency of tours involving mode switching is negligible. In this case, the mode choice for NHB trips can be assumed to be determined by the home-based trip mode choices; i.e., separate NHB models could be estimated for each mode, and the trip table inputs to the NHB models are for a specific mode.

# MODEL DEVELOPMENT PROCESS

An aggregate forecasting system based on disaggregate models offers more opportunities for refinement and validation than does a system based on purely aggregate models. This is because, with disaggregate model systems, the individual models can be tested in both their aggregate (e.g., at the zone level) and disaggregate forms.

## Estimation and Validation

Figure 7 presents a schematic outline of the overall model development process starting with estimation. Disaggregate estimation requires a sample of observed

Figure 3. A comparison of variables included in the MTC and conventional model systems.

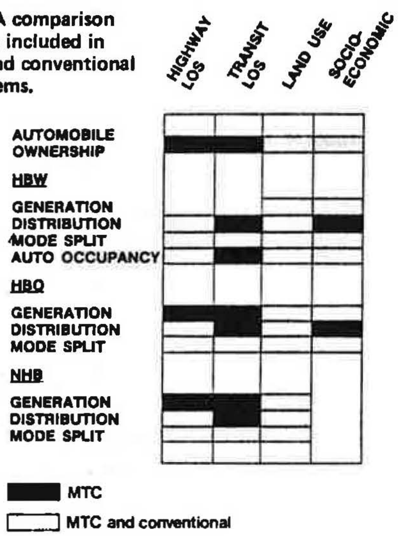



Figure 4. Worker mobility models.

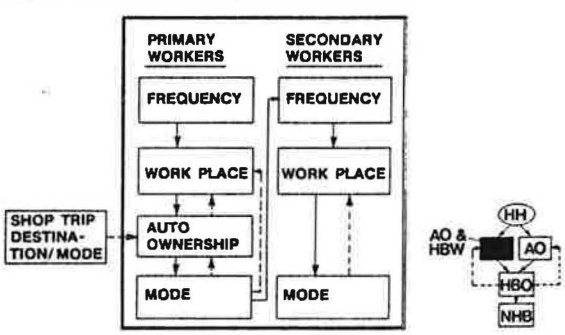



Figure 5. Travel models I: home-based other trips.

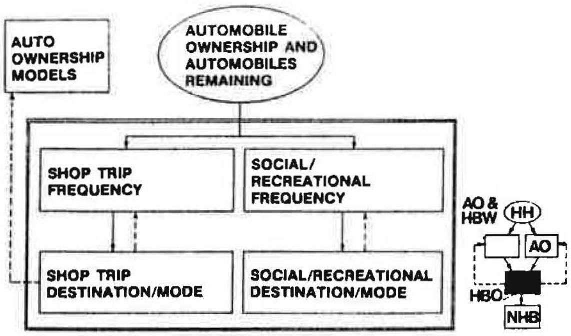



Figure 6. Travel models II: non-home-based trips.

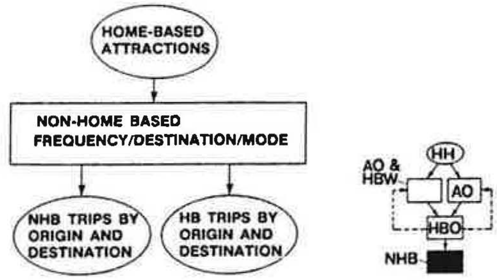
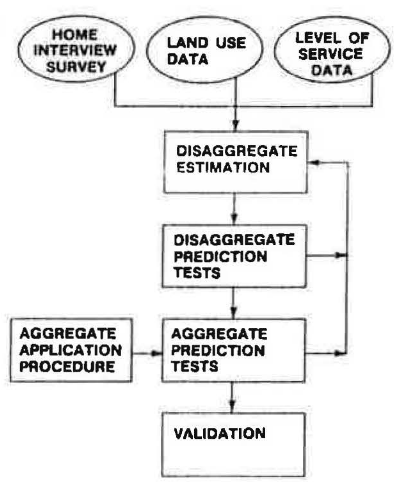



Figure 7. Disaggregate model development process.



travel decisions and socioeconomic characteristics from a home interview or other survey, land-use data to describe the attractiveness of destinations, and level-of-service data to describe model alternatives. In terms of types of data required, the aggregate and disaggregate approaches are identical. The key difference for disaggregate systems arises from the significant reduction in sample size required to yield statistically significant parameters.

As shown in Figure 7, following model estimation, the next step in the overall model development process is a series of disaggregate prediction tests, a step unique to the disaggregate approach. The estimated component models can be tested one at a time, passing each observed trip in the estimation data set through the disaggregate model system to produce tables of predicted and observed choices. Weaknesses in model specification may show up as systematic mispredictions by market segment (such as income) or by explanatory variable (such as travel time). The feedback loop from disaggregate prediction to disaggregate estimation in Figure 7 represents the decision to return to the estimation step based on the failure of a given model specification. Only after each model has passed the disaggregate tests does the process proceed to aggregate prediction and validation.

The base-year aggregate validation procedure for a disaggregate model system is essentially the same as for conventional aggregate systems. With disaggregate models it may be necessary to modify the forecasts by making transformations of the utilities. This requirement arises from the use of average households per zone to represent all households in a zone. The difference between the average behavior of all households and the behavior of an average household is referred to as aggregation bias (12, 13).

We found it necessary in the MTC model system to add distance-related factors to the utilities of the destination choice alternatives to match the observed trip length distribution. Also, in attempting to match zone-to-zone or district-to-district interchanges, we had to add trip interchange adjustment factors (in some cases) to obtain the equivalence between observed and predicted data. In this context, however, one important difference exists between traditional aggregate and disaggregate model systems. A complete disaggregate model system can be estimated with data from as few as 1000 households. The zone interchanges obtained from a sample that small are too sparse to use as the basis of zone interchange adjustments. Either one must rely on the trip length adjustments as the sole means of validating the distribution models with the survey data or one must augment the survey sample used for model estimation with additional observations. In the case of the MTC models, this sample was used in its entirety to ensure that the aggregate versions of the models matched observed travel patterns.

If the disaggregate testing procedure in Figure 7 is followed carefully, there should be no need to return from the aggregate tests to the disaggregate estimation, since the only changes made in the models involve adjustment of the utilities.

Disaggregate estimation and disaggregate and aggregate prediction testing were carried out iteratively as a part of the MTC model development process. As a result, the model system matches observed data in all of the following respects, at the district and district-to-district levels (there are 30 districts in the MTC analysis area): (a) person trips produced and attracted, (b) average trip length for person trips by production district, (c) mode choice by production district, and (d) trip interchanges by mode.

The final validation of the model requires external data sources and preferably data from before and after a change in the transportation system that can directly be compared with the model predictions. This step is being carried out by MTC in their continuing use of the model system.

## Model Application Procedures

As part of the travel model development project, two computerized procedures have been developed to apply the demand models described in this paper. The first is oriented toward the application of an aggregate version of the model system for detailed regional network analysis in either the short- or long-range time frame. The second is oriented to more rapid generalized policy analysis in the short- to medium-range time frame. Each of these procedures is described in this section.

### MTC Regional Network Analysis

MTC regional network analysis (MTCFCAST) is a package of programs based on all of the MTC travel demand models that predicts regional travel patterns and volumes from regional socioeconomic information and the level of service data for existing and proposed modes of transportation. This computer system represents an alternative to the demand estimation portions of the traditional urban transportation planning (UTPS) packages (14, 15) and is integrated with them in its external data structure and its use of their data-processing, report-formatting, and traffic-assigning programs. This compatibility with the UTPS package greatly enhances the effectiveness of MTCFCAST. This system provides the aggregate application procedure used for aggregate prediction tests, as shown in Figure 7.

The primary product of MTCFCAST is a set of eleven 24-h person-trip tables on

1. Driving alone home-based work trips,
2. Shared ride home-based work trips,
3. Transit home-based work trips,
4. Auto home-based shopping trips,
5. Transit home-based shopping trips,
6. Auto home-based social-recreational trips,
7. Transit home-based social-recreational trips,
8. Auto non-home-based trips,
9. Transit non-home-based trips,
10. Auto home-based trips, and
11. Transit home-based trips.

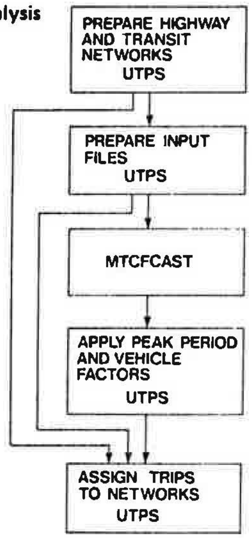
These trip tables may be put directly into the UTPS matrix manipulation and network assignment routines to produce 24-h or peak-hour highway vehicle and transit person-trip assignments. The entire composite MTCFCAST-UTPS analysis framework is shown in Figure 8.

The MTCFCAST forecasting system is similar to but more sophisticated than the conventional trip generation-trip distribution-modal split methodology. It incorporates or provides the necessary inputs to UTPS routines that incorporate each of the models described in the previous section. Its structure is essentially that shown in Figure 2. Aggregation is performed by market segmentation by using average socioeconomic values for each of three income groups initially, followed by segmentation based on auto ownership level after the prediction of auto ownership in the mobility blocks.

### Short-Range Generalized Policy Analysis

Short-range generalized policy analysis (SRGP) is a

Figure 8. The MTCFCAST analysis framework.




computerized procedure that applies a subset of the MTC travel demand models for analysis. The procedure is designed to produce rapid turnaround estimates of the consequences of broadly defined transportation policy options. SRGP processing and outputs are based on an input sample of home interview survey households. The program estimates the travel behavior of the individual households subject to user-controlled facilities for expanding the results in whatever manner is appropriate to the problem universe. This approach takes full advantage of the disaggregate nature of the demand models. Aggregation does not take place until the expansion, after all estimation is complete, and can be straightforward and without bias.

It is the use of sample households that also lends SRGP its short-range applicability. The procedure has no facilities for modeling the long-range dynamics of attributes of households, such as location and life cycle progression or the number of workers and their choice of occupations and job locations. Activity distributions (the zone extent and intensity of employment, shopping facilities, social and recreational opportunities, etc.) are also provided exogenously to the procedure. It was previously applied to other urban areas using alternative disaggregate models (16, 17).

Because of its orientation to short-range analysis, only the following models, which represent short-range choices, of the full MTC model system are included: auto ownership for worker households, auto ownership for nonworker households, HBW mode choice, HBSH trip production, HBSR trip production, HBSH destination-mode choice, and HBSR destination-mode choice.

As these models are applied to each household in turn, summary impacts are accumulated and reported for household income class groups or other segmentation. SRGP also has the capability to retrieve the results of a previous run to produce comparison tables.

## CONCLUSIONS

As the first production-oriented system developed for use by a metropolitan planning organization that is based on a consistent theory of traveler behavior and disaggregate model estimation, the MTC model system has achieved the following major accomplishments.

Disaggregate models have been integrated into a working aggregate model system that resembles the conventional systems with which planners are familiar.

The model system is a reasonable compromise between validly representing the theories of individual travel behavior and developing a practical, large-scale planning model.

The model system has improved behavioral properties compared to its previous aggregate counterpart. It is relevant to the low capital cost alternatives that must be evaluated now and is sensitive to a wider range of variables than aggregate models.

The model system provides an additional option over the full detail and sketch planning approaches supported by conventional aggregate models. Random sample enumeration is a valuable, quick, low-cost method of analyzing a wide range of policies and projects.

Cost reductions in the development of disaggregate model systems arise from the vastly reduced amount of data needed. Employing random sample enumeration techniques also reduces costs of using the models in certain situations.

The only disadvantage that exists in the use of this approach relative to conventional aggregate models is the increased complexity of the system. Due to improved behavioral representation, more models are estimated; they are closely interconnected; and they are not yet well understood by practitioners. The cost of a full zone aggregate application of a disaggregate model system is marginally higher than its aggregate counterpart, but this cost differential is not a significant issue since the major costs of both types of systems are network skimming and assigning.

Work undertaken to date has shown that this modeling approach is feasible, that careful estimation and testing are necessary during the model development phase, that extensive training is necessary to familiarize planners with the new approach, and, of course, that the resulting model systems are sufficiently improved over the conventional system they replace to warrant the investment in training and model development.

## ACKNOWLEDGMENTS

We are indebted to many people for their contributions to the MTC Travel Model Development Project and to this paper. We would specifically like to thank William Davidson, Hanna Kollo, and Gordon Shunk of MTC, as well as the following Cambridge Systematics, Inc., staff members and consultants who contributed to the project: Richard Albright, Jeffrey MacMann, Marvin Manheim, Patrick O'Keefe, Len Sherman, Antti Talvitie, and David Welland. The project was funded directly by the San Francisco area's Metropolitan Transportation Commission and indirectly by the Urban Mass Transportation Administration, U.S. Department of Transportation.

## REFERENCES

1. M. E. Ben-Akiva. Structure of Passenger Travel Demand Models. Department of Civil Engineering, Massachusetts Institute of Technology, Cambridge, PhD thesis, 1973.
2. M. E. Ben-Akiva and S. R. Lerman. Some Estimation Results of a Simultaneous Model of Auto Ownership and Mode Choice to Work. Transportation, Vol. 3, No. 4, 1974.
3. Charles River Associates. A Disaggregate Behavioral Model of Urban Travel Demand. NTIS, Springfield, VA, 1972.
4. D. L. McFadden. The Measurement of Urban Travel Demand. Journal of Public Economics, 1974, pp. 303-328.
5. Cambridge Systematics. Metropolitan Transportation Commission Travel Model Development

Project: Final Report. Cambridge, MA, Vols. 1, 2, and 3, 1978.
6. W. Goldner. Projective Land Use Model (PLUM), BATSC Technical Report 219. Bay Area Transportation Study Commission, Berkeley, CA, 1968.
7. M. G. Richards and M. E. Ben-Akiva. A Disaggregate Travel Demand Model. Saxon House/Lexington Books, Lexington, MA, 1975.
8. M. E. Ben-Akiva, S. R. Lerman, and M. L. Manheim. Disaggregate Models: An Overview of Some Recent Research Results and Practical Applications. Proc., Planning and Transport Research and Computations Summer Meeting, London, 1976.
9. T. J. Adler. Modelling Non-Work Travel Patterns. Department of Civil Engineering, Massachusetts Institute of Technology, Cambridge, PhD thesis, 1976.
10. Analysis of Zonal Level Trip Generation Relationships. Metropolitan Washington Council of Governments, Technical Rept. 10, 1974.
11. M. E. Ben-Akiva and S. R. Lerman. Disaggregate Travel and Mobility Choice Models and Measures of Accessibility. Paper presented at the 3rd International Conference on Behavioral Modelling, Australia, Apr. 1977.
12. F. S. Koppelman. Guidelines for Aggregate Travel Predictions Using Disaggregate Choice Models. TRB, Transportation Research Record 610, 1976, pp. 19-24.
13. F. S. Koppelman and M. E. Ben-Akiva. Aggregate Forecasting with Disaggregate Travel Demand Models Using Normally Available Data. Proc., World Conference on Transportation Research, Rotterdam, 1977.
14. FHWA Computer Programs for Urban Transportation Planning. Federal Highway Administration, 1974.
15. Urban Transportation Planning System: Reference Manual. Urban Mass Transportation Administration, 1977.
16. T. J. Atherton, J. H. Suhrbier, and W. A. Jessiman. Use of Disaggregate Travel Demand Models to Analyze Carpooling Policy Incentives. TRB, Transportation Research Record 599, 1976, pp. 35-40.
17. M. E. Ben-Akiva and T. J. Atherton. Choice-Model Prediction of Car-Pool Demand: Method and Results. TRB, Transportation Research Record 637, 1978, pp. 13-17.

# Non-Home-Based Models

Moshe E. Ben-Akiva, Cambridge Systematics, Inc., and Department of Civil Engineering, Massachusetts Institute of Technology, Cambridge
Len Sherman and Brian Kullman, Cambridge Systematics, Inc., Cambridge, Massachusetts

**This paper describes a practical model of non-home-based travel that can be incorporated in existing urban transportation model systems. The model is estimated by using a disaggregate sample of trips drawn from the 1965 home interview survey of the San Francisco Bay Area for the Metropolitan Transportation Commission. The model predicts trip generation, distribution, and mode split with full sensitivity to travel times, costs, and zone characteristics. The paper describes the overall model structure and estimation results. Comparisons with other research on non-home-based travel are drawn and recommendations for future research presented.**

There is a clear need to better understand non-home-based (NHB) travel behavior. In large urban areas, NHB travel may represent over 20 percent of total vehicle trips, and an even larger proportion when pedestrian travel is counted. Moreover, emerging transport policies focusing on downtown people-mover systems, free or reduced-fare transit circulation systems within central business districts (CBDs), auto restricted zones, and activity center connectors have drawn increased attention to NHB travel patterns.

In modeling urban travel patterns, it has been traditional to classify trips into three categories: home-based work (HBW) trips, home-based other (HBO) trips, and non-home-based (NHB) trips. Models of non-home-based travel have received the least attention. From a behavioral standpoint, non-home-based travel should be sensitive to the same factors that are conventionally associated with HBW and HBO travel patterns. For example, accessibility and the spatial distribution of opportunities should clearly influence the generation and distribution of NHB trips, and yet most existing model systems treat NHB travel as a fixed proportion of home-based trips or use simple generation rates that take no account of transport accessibility. Also lacking from most conventional approaches is a behaviorally plausible representation of trip chaining.

Some recent research has attempted to develop explicit models of trip chaining. Adler, for example (18, 19), has estimated disaggregate choice models in which the choice set consists of alternative trip chains with two or more links. Each alternative's utility was characterized by a total tour level of service and spatial opportunities at the destinations included in the tour. Another approach was taken by Horowitz (20), who used regression analysis to model the frequency and number of stops made on household trip chains as a function of level of service. Because of their complexity or structure, neither of these approaches is suitable for inclusion in an urban area travel demand forecasting system, but they do establish the importance of characterizing the interdependence of non-home-based and home-based travel.

The models reported here explicitly represent this interdependence as a Markovian process where the decision of an individual to continue a tour with a non-home-based trip depends only on conditions at a specific trip end, not on the sequence of trips that may have led to the trip end. This structure was dictated by practical considerations, since in aggregate forecasting applications it is not computationally feasible to treat specific tour sequences as explicit choice alternatives.

While the theory, estimation, and preliminary validation of the NHB models reported here were developed at the disaggregate level, the models were designed to be incorporated into an aggregate travel demand forecasting system to be used by the Metropolitan Transportation Commission (MTC) of the San Francisco Bay Area (21, and the paper by Ruiter and Ben-Akiva in this Record). Thus, from the beginning, model development was concerned with balancing theoretical and behavioral plausibility with model practicality and ease of use. The key advantages of the resulting models are their sensitivity to transportation level of service and their explicit representation of the interdependence of home-based and non-home-based travel.

## DEFINITIONS AND TERMINOLOGY

Non-home-based travel consists of all trips made between two non-home locations. Examples of NHB trips are a journey from work to a place to eat lunch or the journey from a shopping location to work. While it is possible to further differentiate NHB travel by trip purpose, in the empirical work of this study all travel purposes were grouped together.

In developing a theory of NHB travel behavior, it is convenient to adopt a notation where trips are defined by purpose at the origin and destination. Trip end purposes can then be classified into three groups: home (H), work (W), and other (O). In these terms, NHB travel occurs whenever a trip has W or O purposes at each end. Thus, the number of NHB trips between zones e and k can be written as the sum of the following sets of trips:

$$\text{NHB trips from e to k} = W_eO_k + O_eW_k + O_eO_k + W_eW_k \qquad (1)$$

where

$W_eO_k$ = trips with purpose W at origin e to destination k with purpose O,
$O_eW_k$ = trips with purpose O at origin e to destination k with purpose W,
$O_eO_k$ = trips with purpose O at origin e to destination k with purpose O, and
$W_eW_k$ = trips with purpose W at origin e to destination k with purpose W.

Similarly, home-based (HB) trips may be defined as follows (ignoring home-to-work trips):

$$\text{HB trips from e to k} = H_eW_k + H_eO_k + W_eH_k + O_eH_k \qquad (2)$$

where

$H_eW_k$ = trips with purpose H at origin e to destination k with purpose W,
$H_eO_k$ = trips with purpose H at origin e to destination k with purpose O,
$W_eH_k$ = trips with purpose W at origin e to destination k with purpose H, and
$O_eH_k$ = trips with purpose O at origin e to destination k with purpose H.

It was assumed in this study that there is no need to differentiate between work-based and non-work-based, non-home-based trips. Accordingly the above notation can be greatly simplified by designating A to stand for any non-home-based purpose, i.e., A = W + O. This definition leads, for example, to the following relations:

$$H_eA_k = H_eW_k + H_eO_k \qquad (3)$$

$$A_eA_k = W_eO_k + O_eW_k + O_eO_k + W_eW_k \qquad (4)$$

The symbols A and H can be used to represent both the type and the direction of trips as indicated in Figure 9. The terminology and definitions in Figure 9 are essential for the development of the model described below.

## THEORETICAL MODEL DEVELOPMENT

Each trip maker at a non-home location is always faced with the choice between returning home or traveling to a non-home location. If the traveler returns home, then one may say that a NHB trip was not made. If a traveler does continue on to another non-home location, then the traveler must choose between alternative destinations and modes. This sequence is repeated until the traveler eventually returns home. This view of the NHB choice process makes it clear that the traveler's decision process can be associated with the trip end. That is, each home-based trip (and succeeding NHB trips) constitutes a potential NHB trip.

Given this general framework, two assumptions were adopted to simplify the empirical analysis and to facilitate the use of the models in aggregate forecasting. First, it was assumed that each traveler's decision is independent of previous decisions, i.e., that the decision maker has no memory of past decisions. This assumption obviates the need to represent several alternative trip chains (e.g., two-leg versus three-leg versus four- or five-leg tours) as explicit choice alternatives. Second, it was assumed that travelers making NHB vehicle trips would continue on the same mode as their outbound home-based trip; i.e., mode switching is not explicitly modeled. Analysis of San Francisco travel data indicated an extremely low incidence of vehicle mode switching between HB and NHB trip legs. The models, in any event, do not ignore trips that switch mode. They simply predict these trips as if they continued on the same mode as on the HB trip. As a result of this assumption, separate NHB models could be estimated conditional on auto or transit choice for the home-based trip.

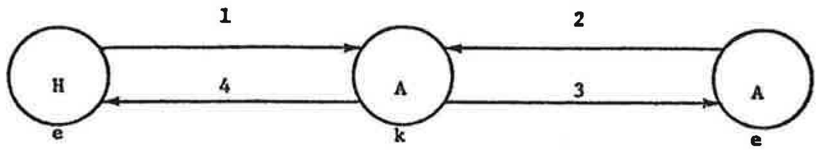
It is apparent from the above discussion that the analysis assumes that a traveler at a non-home location has choices over two dimensions: either to return home (NHB frequency = 0) or to make an NHB trip (NHB frequency = 1) to a specific alternative trip end. Using the approach of random utility choice models employed in disaggregate travel demand models, the preferences of an individual for these choice alternatives can be represented by the following probabilities:

1. The first is $p(f = 1, e|k)$ = probability of making an NHB trip (f = 1) to trip end location e, which could be either an origin or a destination depending on the form of the model—a point that will be discussed below—given that the opposite trip end is in zone k; and
2. The second is $p(f = 0|k)$ = probability that an NHB trip is not made given a trip end, either origin or destination, in zone k.

There will be one such probability for each mode. For simplicity, we will not carry the mode subscript in our notation.

It is possible to model NHB trips from either a destination or an origin perspective. The former represents a traveler choosing from alternative non-home destinations (NHB f = 1) given a non-home trip origin,

**Figure 9. Non-home-based trip types.**




$H_eA_k$ = type 1 trips
$A_kH_e$ = type 3 trips } HB

$A_eA_k$ = type 2 trips
$A_kA_e$ = type 4 trips } NHB

while the latter represents alternative non-home origins from which a traveler could have reached a non-home destination. Although the former approach appears more reasonable, both types of models were estimated in this study in order to calculate directional splits by trip purpose.

The disaggregate model of NHB frequency and trip end was specified as a joint choice logit model:

$$p(f = 1, e\,|\,k) = \exp V_{ek} \Big/ \left(\exp V_{ok} + \sum_{e' \epsilon A_k} \exp V_{e'k}\right) \quad (5)$$

$$p(f = 0\,|\,k) = \exp V_{ok} \Big/ \left(\exp V_{ok} + \sum_{e' \epsilon A_k} \exp V_{e'k}\right) \quad (6)$$

where

$A_k$ = set of feasible trip end choices from trip end k (note that this set may depend on the mode), and

$V_{ok}$, $V_{ek}$ = linear parameter utility functions for the alternatives f = 0 and f = 1 and trip end e, respectively.

The utility associated with choosing trip end (zone) e was represented by the cost and travel time between zones k and e and zone e's attractiveness in terms of the density and level of its population and employment. The zero frequency function contained proxy measures for factors that would decrease the probability of NHB travel.

## MODEL ESTIMATION

### Definition of Choice Set

The San Francisco MTC region was divided into 30 districts, each of which represented a possible trip end choice. In addition, in NHB destination (origin) choice models, all zones in the origin (destination) district were considered as possible choices. This procedure was designed to reduce the variance associated with the level-of-service measures. Analysis of our data showed that over 85 percent of NHB trips were intradistrict. Hence, for the great majority of the chosen trips, level of service was represented at the zone interchange level, rather than district to district. In the NHB transit models a potential destination was considered unavailable if transit service was unavailable. The maximum size of the choice set was 29 districts (not including the trip end district k) plus 26 zones (maximum; many districts have fewer) plus 1 zero frequency alternative which equals 56 alternatives maximum.

### Estimation Data Set

The estimation data set for the NHB models consisted of a sample of 11 249 trip records from 1347 households randomly drawn from MTC's 1965 home interview survey tape. Screening of all the trips for invalid modes (taxi, truck, walk) and purpose (school) reduced the set of valid observations to 8216 records, of which 3214 were NHB f = 1 and 5092 NHB f = 0. These resulting data were subdivided into four files corresponding to auto, transit, origin choice, and destination choice model forms. The resulting numbers of trip records are shown below.

| Model Type | Population | Freq = 1 | Freq = 0 |
|---|---|---|---|
| Auto, destination choice | 4730 | 1789 | 2941 |
| Transit, destination choice | 308 | 57 | 251 |
| Auto, origin choice | 4759 | 1789 | 2970 |
| Transit, origin choice | 265 | 57 | 208 |

The small differences in the population of trips from which NHB trips could be made as represented in the origin and destination choice models are due to incomplete trip chains contained in the survey data.

### Specification of Independent Variables

The variables that were used in the models are listed below.

| Variable | Description |
|---|---|
| ZEROFCON | 1 in zero frequency utility function, 0 otherwise |
| CBDOCON | 1 in zero frequency utility function if origin zone (destination zone if NHB origin choice model) is in CBD, 0 otherwise |
| LN(TT) | Natural log of total travel time in f = 1 utility functions, in minutes |
| COST | Travel cost in f = 1 utility functions, in cents |
| EMPDENS | Employment density: total employment divided by total area at the destination (origin if NHB origin choice model) alternative in f = 1 utility functions, in employees per square kilometer |
| LN(P/E) | Natural log of the zone fraction of regional population divided by zone fraction of regional total employment at the destination (origin if NHB origin choice model) alternative in f = 1 utility functions |
| LN(EMP) | Natural log of the zone fraction of regional total employment at the destination (origin if NHB origin choice model) alternative in f = 1 utility functions |

Table 1. Structure of the NHB models.

| Model | ZEROFCON | CBDOCON | LN(TT) | COST | EMPDENS | LN(P/E) | LN(EMP) |
|---|---|---|---|---|---|---|---|
| Auto destination choice | | | | | | | |
| f = 0 | $\alpha_1$ | $\alpha_2$ | | | | | |
| f = 1 | | | $\beta_1$ | $\beta_2$ | $\beta_3$ | $\beta_4$ | 1.0 |
| Transit destination choice | | | | | | | |
| f = 0 | $\gamma_1$ | $\gamma_2$ | | | | | |
| f = 1 | | | $\delta_1$ | $\delta_2$ | $\delta_3$ | $\delta_4$ | 1.0 |
| Auto origin choice | | | | | | | |
| f = 0 | $\theta_1$ | $\theta_2$ | | | | | |
| f = 1 | | | $\phi_1$ | $\phi_2$ | $\phi_3$ | $\phi_4$ | 1.0 |
| Transit origin choice | | | | | | | |
| f = 0 | $\pi_1$ | $\pi_2$ | | | | | |
| f = 1 | | | $\epsilon_1$ | $\epsilon_2$ | $\epsilon_3$ | $\epsilon_4$ | 1.0 |

Table 2. Summary of model estimation results.

| | Destination Choice | | | | | | Origin Choice | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Auto | | | Transit | | | Auto | | | Transit | | |
| Variable | EC[a] | SE[b] | t-Statistic | EC[a] | SE[b] | t-Statistic | EC[a] | SE[b] | t-Statistic | EC[a] | SE[b] | t-Statistic |
| CBDOCON | 0.313 5 | 0.068 2 | 4.59 | - | - | - | 0.368 4 | 0.068 3 | 5.39 | 1.880 | 0.458 0 | 4.11 |
| ZEROFCON | -11.22 | 14.39 | -0.78 | -7.57 | 1.47 | 5.14 | -10.75 | 7.90 | -1.36 | -4.53 | 1.37 | 3.32 |
| LN(TT) | -2.294 | 0.073 6 | 31.2 | -1.033 | 0.306 1 | -3.38 | -2.418 | 0.074 0 | -32.7 | -0.898 2 | 0.239 8 | -3.75 |
| COST | -0.015 18 | 0.001 51 | -10.1 | -0.016 08 | 0.006 91 | -2.33 | -0.013 38 | 0.001 47 | -9.09 | -0.017 75 | 0.006 09 | -2.65 |
| LN(P/E) | 0.470 0 | 0.030 6 | 15.38 | 0.608 8 | 0.209 9 | 2.90 | 0.398 1 | 0.030 0 | 13.26 | - | - | - |
| EMPDD | 0.002 242 | 0.000 6 | 3.87 | 0.003 566 | 0.002 0 | 1.77 | 0.001 887 | 0.000 5 | 3.72 | - | - | - |
| LN(EMP) | | 1.0 | | | 1.0 | | | 1.0 | | | 1.0 | |
| Percent correct predictions | | 11.9 | | | 55.9 | | | 11.8 | | | 50.0 | |
| $\rho^{-2}$ | | 0.546 4 | | | 0.724 5 | | | 0.552 7 | | | 0.716 3 | |
| Likelihood ratio statistic | | 19 341 | | | 1554 | | | 19 681 | | | 1322 | |

[a]EC = estimated coefficient. [b]SE = standard error.

The structure of the model showing how each variable enters in the f = 0 and f = 1 utility function is given in Table 1, which emphasizes the fact that no assumptions were made that the coefficients of identical variables in the origin and destination choice models would be the same. The variable LN (EMP) is a measure of the size of a destination alternative and its coefficient is constrained to take the value of 1.0. This constraint is necessary if the model is independent of the zone system used for estimation. In fact, in aggregate forecasts made with the NHB models described here, a different zone system has been used.

### Estimation Results

Estimation results for the four model types are shown in Table 2. For each variable the cell entries indicate estimated coefficient, standard error, and t-statistic. In preliminary estimations not reported here, gross employment density was used as a variable to explain walk trip propensity (i.e., in the f = 0 alternative), but its estimated coefficient was insignificant. Also, in preliminary runs, it was determined that in- and out-of-vehicle travel time could not be statistically distinguished, so only total travel time was ultimately used. The coefficient of the size variable LN(EMP) is constrained to 1.0 in all cases.

The relatively high value for the $\rho^{-2}$-statistic derives from the fact that the f = 0 alternative is chosen with high frequency, while it is presumed in the "equally likely" case that it has only an equal chance of being chosen along with each f = 1 destination.

An indication of the reasonableness of the model results is the implied value of time (VOT). For the recommended models, the values of time for a range of travel times between 20 and 60 min were calculated. These results, based on the estimated models in Table 2, are shown below.

| Model | Total Travel Time (min) 20 | 40 | 60 |
|---|---|---|---|
| NHB destination choice, auto | 4.53 | 2.27 | 1.51 |
| NHB origin choice, auto | 5.42 | 2.71 | 1.81 |
| NHB destination choice, transit | 1.92 | 0.96 | 0.64 |
| NHB origin choice, transit | 1.52 | 0.76 | 0.51 |

Auto VOT is higher than transit VOT for every travel time and model type (origin or destination choice). For both modes, travelers show a decreasing sensitivity to marginal time savings as total travel time rises. These results are similar to those obtained from the HBO destination and mode-choice models estimated from the same data set (21).

## FORECASTING WITH THE NHB MODELS

Forecasting with the NHB models is complicated by two factors. First, the models predict the non-home-based travel decision of a traveler at the trip end of either a home-based or a non-home-based trip; and, second, the directional split of the home-based trip is unknown.

Consider, for example, using a destination choice approach where the directional split of home-based trips is known. In this case, non-home-based trips could be determined by predicting trips in sequence: first HB and then NHB. However, since the models will predict a non-zero probability to every possible destination zone, this approach is computationally infeasible for application to large networks that typify major urban areas. Lerman and others (22) used a variant of this procedure by simulating each trip leg of a tour conditional on the previous trip end, but their approach is far too complex to be practical for aggregate forecasting applications.

Assuming that there are N possible zones, the total number of NHB flows is $N^2$, and this prediction problem could be viewed as a solution of a system of $N^2$ simultaneous equations. Each equation expresses the expected number of non-home-based trips from a given origin to a given destination as follows:

$$A_k A_e = P(f = 1, e|k)\left(HA_k + \sum_{e'} A_{e'} A_k\right) \tag{7}$$

where $HA_k$ is $\sum_{e'} H_{e'} A_k$. The unknown directional split of the home-based trips adds 2N more equations and unknowns. The first N of these equations defines the directional split of the home-based attractions:

$$HA_k + A_k H = HBA_k \tag{8}$$

where $HBA_k$ is home-based attractions at k.

The second set of N equations defines the conservation of flows at the non-home ends:

$$HA_k + \sum_e A_e A_k = A_k H + \sum_e A_k A_e \tag{9}$$

These systems of equations (with known or unknown HB direction split) are too complex to be solved directly. Therefore, a simplified method based on the use of both origin and destination choice models was devised.

To distinguish between the two NHB model types, we introduce the subscript i = 1, 2 to refer to the choice of NHB destination (i = 1) or NHB origin (i = 2):

$$P_i(f = 1, e|k) = \exp V^i_{ek} \Big/ \left(\exp V^i_{ok} + \sum_{e'} \exp V^i_{e'k}\right) \tag{10}$$

where

i = 1 implies e = destination, k = origin, and
i = 2 implies e = origin, k = destination.

The equations predicting NHB trips are

$$A_k A_e = P_1(f = 1, e|k)\left(HA_k + \sum_{e'} A_{e'} A_k\right) \tag{11}$$

$$A_e A_k = P_2(f = 1, e|k)\left(A_k H + \sum_{e'} A_k A_{e'}\right) \tag{12}$$

Summing both sides of these equations over trip ends and denoting the summation over trip ends e by the absence of the subscript e, we may write

$$A_k A = F_{1k} \times (HA_k + AA_k) \tag{13}$$

$$AA_k = F_{2k} \times (A_k H + A_k A) \tag{14}$$

where $F_{1k}$ and $F_{2k}$ are the fraction of potential NHB trips choosing zone k as the origin or destination of an NHB trip. Specifically these fractions may be written as

$$F_{1k} = \sum_e P_1 (f = 1, e|k) \tag{15}$$

$$F_{2k} = \sum_e P_2 (f = 1, e|k) \tag{16}$$

Equations 13 and 14 can now be solved for $A_k A$ and $AA_k$:

$$A_k A = (F_{1k} HA_k + F_{1k} F_{2k} A_k H)/(1 - F_{1k} F_{2k}) \tag{17}$$

$$AA_k = (F_{2k} A_k H + F_{1k} HA_k)/(1 - F_{2k} F_{1k}) \tag{18}$$

Substituting Equations 17 and 18 into 11 and 12 yields

$$A_k A_e = P_1 (f = 1, e|k) [(HA_k + F_{2k} A_k H)/(1 - F_{1k} F_{2k})] \tag{19}$$

$$A_e A_k = P_2 (f = 1, e|k) [(A_k H + F_{1k} HA_k)/(1 - F_{1k} F_{2k})] \tag{20}$$

Either Equation 19 or Equation 20 could be used to forecast trips, but not both, since these equations predict opposite ends of the same NHB trips. It is more natural to use Equation 19, which predicts NHB trips from origin zone k to destination zone e based on both the home-to-any and the any-to-home trip ends in zone k. Although only one of the equations is needed to forecast, note that the influence of both origin and destination models enters in Equations 19 and 20 through the terms $F_{1k}$ and $F_{2k}$.

Starting from Equations 13 and 14, the complete set of relations used for NHB aggregate forecasting can now be derived by adding the two equations for the directional split of home-based attractions at k and flow conservation at k. The entire set of equations can then be solved simultaneously to obtain expressions for $HA_k$, $A_k H$, $A_k A$, and $AA_k$, which depend only on $F_{1k}$, $F_{2k}$, and known $HBA_k$:

$$HA_k = [(1 - F_{2k})/(2 - F_{1k} - F_{2k})]\ HBA_k \tag{21}$$

$$A_k H = [(1 - F_{1k})/(2 - F_{1k} - F_{2k})]\ HBA_k \tag{22}$$

$$AA_k = [F_{2k}/(2 - F_{1k} - F_{2k})]\ HBA_k \tag{23}$$

$$A_k A = [F_{1k}/(2 - F_{1k} - F_{2k})]\ HBA_k \tag{24}$$

The equations serve two purposes: they directionally split the forecasts of $HBA_k$ (home-based trips; use Equations 21 and 22), and they directly predict NHB generations and distribution (use Equation 23 or Equation 24).

## CONCLUSIONS AND DIRECTIONS FOR FURTHER RESEARCH

The models presented in this paper were developed as part of a complete system of disaggregate travel demand models. However, they can be used in a conventional urban transportation model system to replace existing NHB trip models.

The simplifying assumptions of the model structure presented were dictated by practical considerations of model development resources and computer cost for model application.

Given the overall model structure developed in this paper, the specific models could be improved by further disaggregation of purposes, inclusion of mode choice, and addition of other variables to better represent attractions of travel opportunities and transport services.

## ACKNOWLEDGMENTS

This research was conducted as part of a comprehensive travel demand development project sponsored by the Metropolitan Transportation Commission (MTC) of San Francisco. The assistance and cooperation of Gordon Shunk, Hanna Kollo, Bill Davidson, and Earl Ruiter are gratefully acknowledged. Comments on an early draft from Joel Horowitz were particularly useful. The findings and conclusions in this paper are strictly ours.

REFERENCES

18. T. J. Adler. Modelling Non-Work Travel Patterns. Transportation Systems Division, Department of Civil Engineering, Massachusetts Institute of Technology, Cambridge, PhD thesis, 1976.
19. T. J. Adler and M. E. Ben-Akiva. A Theoretical and Empirical Model of Trip Chaining Behavior. Transportation Research, 1978, in press.
20. J. Horowitz. Effects of Travel Time and Cost on the Frequency and Structure of Automobile Travel. TRB, Transportation Research Record 592, 1976, pp. 1-5.
21. Metropolitan Transportation Commission Travel Model Development Project: Final Report. Cambridge Systematics, Cambridge, MA, Vols. 1, 2, and 3, 1978.
22. S. R. Lerman and others. A Method for Estimating Patronage of Demand Responsive Transportation Systems. U.S. Department of Transportation, Transportation Systems Center, Final Rept., DOT-TSC-977, Dec. 1976.

# Discussions

Frederick C. Dunbar, Charles River Associates, Cambridge, Massachusetts

Urban travel demand research in the 1960s and early 1970s gave rise to the hope that travel forecasting models could be developed that would satisfy the following objectives. They would predict the changes in travel demand from adding new modes or other new transportation alternatives, evaluate the consequences of alternative policy options applied to the existing transportation system, be easily transferred from one urban area to others, and make full use of disaggregate data to properly specify travel behavior.

In a major step toward these objectives, McFadden and Domencich (23) showed that probability choice models could be extended using multinomial logit (MNL) to estimate equations representing a wide range of travel decisions. These models were designed to predict the short-run responses of trip makers to transportation system changes. In addition to mode choice, which had traditionally been analyzed with probability choice models, MNL was applied to trip time of day, destination, and frequency decisions. These separate models were linked with accessibility measures to form an integrated travel forecasting structure.

The San Francisco MTC system is an initial attempt to utilize this basic MNL model structure in a large-scale urban transportation planning package. It also incorporates some important refinements in travel demand model estimation with MNL. The first of these is a rigorous definition of the choice structure of mobility decisions based on the work of Ben-Akiva and Lerman (1, 24). The second refinement is an appropriate specification for sequential models with inclusive prices (25). Finally, there has been some attempt to adjust the attraction variables in the destination and frequency choice models to make the equations transferable to alternative zone systems.

Given the scope of the MTC system, it is appropriate to review it in terms of the four objectives of travel demand models presented above. The major advances in large-scale travel forecasting procedures represented by the MTC system appear to be that

1. Policy-sensitive accessibility measures have been incorporated in the separate components of travel demand,
2. New alternatives to the traveler choice set, such as new modes, can be unambiguously forecast by using the independence from irrelevant alternatives (IIA) property of multinomial logit [see Charles River Associates (26) for a discussion of the strengths and limitations of this capability], and
3. The destination choice models make better use of trip record data than does the traditional gravity model calibration approach, thereby gaining, potentially, greater accuracy in forecasting the trip distribution effects of transportation system and land-use changes.

There are several major limitations on the MTC system that are discussed in more detail below. These include, first, problems with model estimation and model specification that combine to prevent isolation of short-run travel response from other effects; these problems should lead to forecasts that are overly responsive to system changes and to biased policy evaluations. Second, the destination and frequency choice model is not generally transferable, and, third, potential advantages of disaggregate data analysis have gone unrealized, perhaps because of the dependence upon the traditional large-scale home interview survey.

## POLICY SENSITIVITY

The MTC system should give biased travel demand forecasts and policy evaluations. This results, in part, from using relatively simple model structures estimated on cross-section data. This statistical design is an unfortunate constraint on most travel demand analysis. In addition, there appear to be model specification errors that compound the drawbacks of using cross-section data.

## CROSS-SECTION BIAS

Not enough attention has been given to cross-section bias in travel demand model estimation. As an example of this type of problem, consider either of the MTC system's disaggregate models of nonwork trip frequency estimated from a household travel survey. These households will have chosen locations for themselves based on, among other things, their preferences for accessibility to alternative activities. Households with low preferences for accessibility will live in areas with high impedance measures for nonwork trips. Households with high preferences for accessibility will live where there are low travel times and costs to a large number of trip-generating activities.

The frequency choice model will use the correlation between trip frequency and accessibility and incorrectly assign an increased frequency response to a policy that increases accessibility.

In fact, households will have revealed their preference for trip frequency by their location decisions. Rather than forecasting trip frequency, the model is partially distributing households to locations as a result of accessibility, given their demands for a certain number of trips. As a result, the model should overpredict the travel demand response to increased trip frequency.

Practitioners should be cognizant of potential cross-section bias, not only in frequency models but also in destination and mode-choice estimates and forecasts. A lack of appreciation of this problem can lead to model specifications that fit the data better than alternatives but lead to worse forecasts of traveler responses to transportation systems. Two examples of model specification problems are discussed below.

## MARGINAL VALUE OF TIME IN THE NON-HOME-BASED DESTINATION/ FREQUENCY MODEL

Specification of the utility function in the travel models will in some ways exacerbate cross-section bias. The MTC system's non-home-based destination and frequency choice model, for example, has the logarithm of travel time as an argument in the utility function. Thus, the marginal value of time is inversely proportional to the amount of travel time between an origin and a destination (O-D); that is, the marginal value of time declines with respect to time spent in travel.

This formulation is inconsistent with the behavior of travelers who maximize utility under time constraints.

The policy implications of this model specification are that for two zone pairs with equal flows, the time benefits of improving travel time by 1 min for an O-D link of 10 min is equivalent to improving travel time by 6 min for an O-D link of 60 min. Project evaluation will be biased toward improving links associated with short origins and destinations inasmuch as cost per kilometer of an improvement is constant or increasing. Thus, the model is policy sensitive, but it may give the wrong priorities to policy alternatives because of its specification.

One problem with estimating the marginal value of time from cross-section data is that people with a low value of time will take longer journeys, which creates a negative statistical correlation between marginal value of time and length of the trip. However, this correlation does not tell us that any given individual has a decreasing marginal value of time when choosing among alternative destinations. This latter concept of value of time is what is relevant in analyzing traveler responses to system changes and in evaluating projects.

## WORK-TRIP DISTRIBUTION

The purpose of the MTC model is to forecast travel demand with exogenously determined land-use patterns. Presumably, the land-use model (PLUM) forecasts zone aggregates of employment and residences that are input to the MTC travel demand system. Thus, long-run residence and employment choice decisions are modeled in PLUM. A transportation policy that redistributes work trips would do so primarily by redistributing zone household locations and employment opportunities. However, the MTC travel demand model keeps these elements constant and reallocates workers among existing places of employment.

For short-run travel demand analysis, it is preferable to keep work-trip distribution constant (as is done in the short-range generalized policy analysis package). For long-run travel demand analysis, work-trip redistribution should be included in the land-use forecasting system with a methodologically sound linkage to the travel demand models.

## TRANSFERABILITY

Strong conditions are required on the use of zone activity variables in the MNL model in order to ensure that arbitrary repartitioning of the zones does not violate conservation of travel flows (27). Even more stringent conditions are required either if the nontrip alternative is included in the model or if the inclusive price (the log of the denominator) from the destination choice model is to be in the utility function of a higher level frequency choice model.

The non-home-based travel model in the MTC system can be used as an example of the problems in transferring destination and frequency choice models. The developers of the model use total zone employment as an attractor variable; it is specified in log form with the coefficient constrained to unity. If this were the only activity variable in the utility function, then the model could be transferred without arbitrarily changing travel demand forecasts. Unfortunately, including other activity variables such as zone attractors in the model leads to a violation of conditions that would ensure transferability. To see this, we rewrite the MTC non-home-based model as follows, ignoring the employment density variable and the zero frequency alternative:

$$P_i = e^{V_i} E_i^{1-a} Y_i^a \Big/ \sum_{j=1}^{n} e^{V_j} E_j^{1-a} Y_j^a \tag{25}$$

where

$P_i$ = probability of the ith alternative;
$V_i$ = level of service component of the utility function;
$E_i$ = employment in the ith destination zone;
$Y_i$ = population in the ith destination zone; and
$a$ = estimated coefficient.

If a repartitioning of the zones occurs such that $\Delta E$ and $\Delta Y$ are taken from destination zone 1 and added to destination zone 2, then the conservation of trips would require the following equality:

$$\begin{aligned}(e^{V_1} E_1^{1-a} Y_1^a + e^{V_2} E_2^{1-a} Y_2^a) \Big/ \sum_{j=1}^{n} e^{V_j} E_j^{1-a} Y_j^a = \\ [e^{V_1}(E_1 - \Delta E)^{1-a}(Y_1 - \Delta Y)^a + e^{V_2}(E_2 + \Delta E)^{1-a}(Y_2 + \Delta Y)]^a \\ \div [e^{V_1}(E_1 - \Delta E)^{1-a}(Y_1 - \Delta Y)^a + e^{V_2}(E_2 + \Delta E)^{1-a}(Y_2 + \Delta Y)^a \\ + \sum_{j=3}^{n} e^{V_j} E_j^{1-a} Y_j^a]\end{aligned} \tag{26}$$

With population and employment acting as attractors, it is necessary for a to be within the unit interval so that the coefficients on zone attraction are positive. However, if trips are to be conserved in cases where the number of choices is three or greater, it will in general be necessary for a to be either zero or one. The values for a estimated in the MTC model are within the unit interval but tend to be around 0.5, which would lead to forecasting error if the model were transferred to an alternative zone system.

Another problem in transferability arises because changing the number of destination alternatives will, in and of itself, affect the frequency of travel. In a joint frequency and destination choice model, such as the non-home-based model in the MTC system, this effect is quite direct.

Suppose, for example, that the denominator in Equation 25 doubles with a twofold increase in destination possibilities. Then, using numbers from the destination sample for the MTC model, the probability of taking a trip would increase from, say, 0.38 (the proportion of nonzero frequency records in the destination sample) to 0.55. This implies a much higher elasticity of trip frequency than is intuitively plausible. Note that, if frequency choice is modeled separately from destination choice, the coefficient on accessibility in the frequency choice model could be less than unity, which would moderate this effect.

DISAGGREGATE DATA ANALYSIS

As a closing and hopefully constructive comment, disaggregate data analysis should focus more on distinguishing variations among households' travel behavior. Different trip makers in a zone have different choice constraints and different sensitivities to transportation LOS variables. These distinctions are now blurred when one model structure is applied to a zone containing a heterogeneous population. In this regard, large-scale planning models have not been significantly improved with the application of disaggregate travel demand models. An important topic for future research is how zone forecasting can be made more precise through the use of disaggregate data analysis that would segment trip makers demographically into groups with distinct travel behavior.

REFERENCES

23. T. Domencich and D. McFadden. Urban Travel Demand: A Behavioral Analysis. North Holland, Amsterdam, 1975.
24. S. R. Lerman. A Disaggregate Behavioral Model of Urban Mobility Decisions. Center for Transportation Studies, Massachusetts Institute of Technology, Cambridge, Rept. 75-5, 1975.
25. D. McFadden. Properties of the Multinomial Logit (MNL) Model. Institute of Transportation Studies, Univ. of California, Berkeley, Urban Travel Demand Forecasting Project Working Paper 7617, 1976.
26. Disaggregate Demand Models. Charles River Associates, Cambridge, MA, Project 8-13, Phase I Rept., Sept. 1975.
27. Disaggregate Demand Models. Charles River Associates, Cambridge, MA, Project 8-13, Phase II Draft Rept., Dec. 1977.

Gordon A. Shunk and Hanna P. H. Kollo, Metropolitan Transportation Commission, Berkeley, California

This discussion is a comment from the user's perspective rather than a critique of the content of the papers. We feel that the papers adequately represent the work done for MTC by the authors. Further discussion of theory and technique does not seem as important as a few comments about the experience we have had in using the models. This is important to anyone considering development or use of such models.

A little background is appropriate first. MTC contracted with the consultant team (COMSIS, Cambridge Systematics, and Barton-Aschman Associates) in 1975 to develop a complete set of travel forecasting models. In phase one of that project the consultants reviewed the needs of MTC for travel forecasting models and reviewed the MTC data base. Then they recommended how we should proceed and told us what it would cost. We accepted their recommendations, and phase two began late in 1975. The model system was delivered about a year later. Some additional validation work was necessary and was completed by mid-1977. The entire effort took about two years and cost nearly $250 000.

MTC gained knowledge and experience with the models as we monitored the consultants' work. MTC conducted the aggregate validation for 1975 for the HBW models and has been using only those models, preparing travel forecasts in one study. MTC also had some brief experience with the entire model system early in 1977, before the models were finally validated for 1965. Our experience has been primarily with the regional network analysis system. We have done some further development work on the policy analysis system, but that is not reported here.

Our position is one of concern that the models may not do the job we need or may not do it within reasonable time and resource constraints. The models may be excellent theoretically and quite accurate—or as accurate as possible given the data base. We shall deal with the behavioral nature purported for the models, some changes necessary to validate and use the models, and their cost effectiveness.

A claim made for these models is that they represent traveler decision-making behavior. Experience with our models indicates that such claims may be somewhat inflated. Our opinion is based on the kinds of adjustments needed to make the models match various measures of existing trip patterns. This judgment is partly the result of our feeling that models should be reasonably consistent within a given data base. We have not found this to be true with our models and data base, but either might be suspect.

What concerns us most are the dramatic changes to the constants of the estimated models to eliminate large prediction errors. The constants represent the unknown or nonquantifiable factors and therefore are subject to considerable estimation error. Even when they are changed, there is no assurance that the constants will retain their values in the future or, if they do not, how they might change.

A second major problem was the inability to identify a function that adequately and consistently reproduced trip distribution behavior. The need for a trip length adjustment variable became apparent in disaggregate validation when estimated models overpredicted long trips. A distance correction variable by district of production was introduced into the utility equation. Its coefficient was developed by trial and error to match observed trip lengths. This is reminiscent of the traditional fitting of older trip distribution functions (friction factors). The consultants interpret this variable as representing travelers' decreasing knowledge of potential attractions as trip length increases. We disagree, since such an adjustment is applied more often to work trips than nonwork trips. It seems that travelers should be much more familiar with work-trip attraction alternatives farther from home than nonwork alternatives.

MTC is performing the 1975 validation by comparing model predictions to estimates from other sources. There was no regional travel survey in 1975, but 1970 census journey-to-work data have been used in this validation. Traffic counts, transit ridership, and numerous special purpose surveys have also been used to check the models.

The performance of the HB work models in 1975 validation indicated that the distribution models require unique adjustment factors similar to traditional K-factors. The K-factor approach was implemented by adding a term to the utility functions for specific i-j pairs (county or district interchanges). This is similar to changes in modal constants of the utility functions. The K-factors were developed first for county interchanges and then for district interchanges.

Since most of the disaggregate research and model development work had been done in the area of work mode choice, the expectations for the performance of these models were high. The results did not measure up to expectations. The mode-choice models also required transit trip production and district interchange adjustment factors.

The results of disaggregate validation showed that the model estimated from 1972 data overpredicted transit in the 1965 disaggregate data set by 80 percent. This required adjustments to the constants of the driving alone and carpooling modes. The prediction error was eliminated by increasing the driving alone constant by 31 percent and increasing the carpooling constant by 26 percent. The reverse happened in aggregate validation. HB work transit prediction using the disaggregately validated model was 42 percent under the 1965 observed regional transit trips. This required a decrease of 43 percent in the constant for the driving alone mode and a decrease of 30 percent in the constant for the carpooling mode.

After pondering these results one wonders which disaggregate data set to believe and if the changes in the constants of the utility functions are more behavioral than shifting empirically derived mode-split diversion curves to match the overall level of transit patronage.

The 1965 aggregate validation of the work mode-split model also showed that even though the total regional transit trips matched the home interview survey data, it was necessary to further adjust the model constants to match transit productions by district. When applying the mode-split model to 1975 estimated conditions, the results showed that different utility adjustment factors than those for 1965 validation were needed. In addition, the model overpredicted intracounty (short) transit trips and underpredicted intercounty (long) transit trips. This meant further intercounty or interdistrict adjustments. The 1975 results were purely mode-choice prediction errors, since the model-simulated person trip table was generated by the distribution model that incorporated K-factors that matched the observed 1975 person trips quite well.

The final aspect of our concern about the models is the resources required to operate them. This includes extensive data processing to prepare a multitude of variables in the proper formats and functions for use by the models. The simplicity of the papers and this discussion belies the extensive directories and cross-references required to use, understand, and interpret these models.

Add to that the problem of expense. To run a complete pass of our model system requires about one week of elapsed time, despite a computer system that gives us the best turnaround in the Bay Area. A complete model run costs over $6000 in computer charges, including all network processing for peak and off-peak transit and highway networks. The demand models themselves cost something over $2700 for a complete run. A complete run includes peak and off-peak networks, four purposes, and all-or-nothing assignment.

Our concerns about these models can be summarized in the following way:

1. The validity and behavioral claims are suspect because of the significant changes in constants and coefficients required to make estimated models reproduce an independent disaggregate data set and an aggregate data set;
2. The particular subsets chosen from the home interview survey data set for disaggregate estimation and validation have too great an effect on estimation results; and
3. The extensive work and reiteration from estimation through validation calls into serious question the claims by some advocates that disaggregate behavioral models are transferable.

# Authors' Closure

We appreciate having this opportunity to respond to the discussions by Dunbar and by Shunk and Kollo. Their review of our papers, and of the MTC model system in its entirety, provides insights from the varying perspectives of the econometrician and tne practitioner. We will respond briefly to the critical comments made by each.

Dunbar's comments deal with potential specification errors in the travel demand models. We note that all of Dunbar's critical comments apply to all existing travel demand models and are not specific to the MTC model system. We will respond briefly to three of these.

Dunbar's comment on cross-section bias may be important only because an ordinary least-squares approach was used to estimate the travel frequency models in the MTC system. In on-going studies, a probabilistic choice structure is used for trip generation models as well as for distribution and mode-choice models. This avoids the type of cross-section bias pointed out in Dunbar's discussion, since estimation of conditional choice probability (i.e., frequency given location) is unbiased.

On value of time Dunbar states that it is inconsistent for a marginal rate of substitution (MRS) between time and cost of travel (i.e., value of time) to decline with respect to time spent in travel. This MRS not only is the result of the effect of time and cost constraints, as Dunbar suggests, but also represents consumers' trade-offs between time and cost that exist when the constraints are not binding. In this case, a declining MRS is perfectly reasonable, matches a wide range of empirical data, and is indeed a property of many existing aggregate and disaggregate travel demand models. The time and cost constraints that do exist are represented indirectly in the MTC model system, as in all others, by limiting predictions of internal travel to the set of zone pairs that exist within some specified analysis area.

Dunbar points out that the zone activity variables used in a number of the MTC destination choice models present potential problems when the models are transferred to other zone systems and other areas. As shown in Dunbar's Equation 25, and using his notation, the total attraction component of a typical model of this type is $E_i^{1-a} Y_i^a$. However, this can also be expressed as:

$$(E_i/Y_i)^{1-a} Y_i \qquad (27)$$

This form supports the interpretation of $Y_i$ as the only size variable, and $E_i/Y_i$ as a rate or density-type variable. The single size variable does lead to the conservation of trips, and any deviations in predictions caused by the rate variable due to changes in zones are no different than deviations in level of service variables—both represent unavoidable aggregation errors inherent in any aggregate application of travel demand models.

Shunk and Kollo comment on questions that have arisen at MTC about the behavioral nature of the models, especially as evidenced by the changes in constants and the distance corrections required for validation, and the cost and effectiveness of model application. We will respond to their major points.

## NATURE OF CONSTANTS IN LOGIT MODELS

Constants in logit models represent more than the unknown and unquantifiable characteristics of travel alternatives mentioned by Shunk and Kollo. When the models are applied at the aggregate (zone) level, these constants must also compensate for whatever biases exist due to the approximations and averages used to characterize the aggregated regional system. The important thing to note is that these adjustments do not change in any way the behavioral validity of the relative weights estimated statistically for the variables of the models. The advantages of disaggregate models in including more relevant variables than is possible in aggregate models, and in requiring fewer observations for model estimation, are also not affected by constants and the need to adjust them in aggregate applications.

TRIP LENGTH ADJUSTMENTS

Shunk and Kollo miss the importance of trip length adjustments when they state that these adjustments are applied more often to work trips than to nonwork trips. The magnitude of these corrections is more relevant than their frequency of application, and the relative magnitudes for work and nonwork purposes vary with trip length. For all trips less than 5 km (3 miles) in length, no correction is applied for either purpose. In the 5- to 24-km (3- to 15-mile) range, work corrections exceed those for nonwork travel. For trips longer than 24 km, the home-based shop corrections are the largest.

COST AND EFFECTIVENESS

Our paper mentions the expanded resource requirements, both in terms of staff understanding and in terms of analysis costs. Shunk and Kollo quote computer costs of \$6000; we maintain that these compare favorably with the costs of traditional aggregate systems, which can be as high as \$10 000 for a full analysis. In addition, it must be noted that these costs apply only to the full network analysis system, MTCFCAST. For many problems faced by MPO's, SRGP can provide the required information at costs per alternative in the \$100-\$200 range, after one-time costs of approximately \$5000, to prepare a data base of household and level-of-service data. It is also worth noting that further work is being done to expand the SRGP approach to be compatible with network assignments. The computer costs of this approach fall in the \$1500-\$2000 range when an iterative procedure is used to predict both demand and network equilibrium, two aspects that cannot be addressed at all for the quoted \$6000 cost.

Fred Reid raised another important question. He asked both Shunk and Ben-Akiva, "If you had the project to do over again, what would you do differently?" This is a question to which we have given considerable thought, because the technical quality and capability of the model system are not being taken full advantage of by the agency for which it has been developed.

One important aspect of the project that would be done differently is that less effort would be spent formulating and estimating additional model components; instead, more effort would be spent on thoroughly testing and validating the fewer model components estimated. This strategy is required to prevent the disillusionment likely to occur when, near the end of the model development process, some component produces unreasonable results under certain input assumptions.

Two other redirections of effort would have increased the usefulness of the modeling work done. First, rather than the almost exclusive emphasis, in the prediction testing and validation portions of the project, on the full network analysis system, MTCFCAST, more effort would have been devoted to demonstrating the value and usefulness of the SRGP program, which is potentially more cost effective for many of the policy questions addressed by an MPO. Second, more emphasis would be placed on ensuring, throughout the project, that the end product be precisely what is needed to meet the agency's planning needs and that the agency staff have full knowledge of the end product and complete facility in using it.

The problems of implementing and successfully using a major new model system require a large amount of cooperative effort by modelers and practitioners to be completely solved.

*Publication of this paper sponsored by Committee on Passenger Travel Demand Forecasting and Committee on Traveler Behavior and Values.*

# Effects of Transportation Service on Automobile Ownership in an Urban Area

Thomas F. Golob,* Consultant, Birmingham, Michigan
Lawrence D. Burns, Transportation and Traffic Science Department,** General Motors Research Laboratories, Warren, Michigan

**A disaggregate automobile ownership choice model is applied to estimating the elasticities of automobile ownership with respect to household income, fixed costs of automobiles, travel times on urban roadways, and public transit service in a case study urban area. Focus is on the aggregate stock of automobiles held by all households and on the distribution of households owning zero, one, two, and three or more autos. Automobile ownership behavior of sociodemographic segments in the total population is also compared. Results indicate that the total number of automobiles owned is approximately three times more sensitive to household income than to automobile travel times. Furthermore, automobile ownership is twice as sensitive to automobile travel times as it is to public transit travel times. Finally, the automobile ownership decisions of inner-city dwellers and older families are more sensitive to all of these factors than are the decisions of suburban dwellers and younger families. It is demonstrated that transportation policies affecting urban traffic efficiency and public transit service are likely to impact on automobile ownership and these impacts will vary with geographical location and population sociodemographic segment.**

The purpose of this research is to estimate the relative sensitivities of urban automobile ownership levels with respect to household income, automobile costs, efficiency of automobiles and public transit travel, and locations of residences and urban activities. Focus is on the aggregate stock of automobiles held by all households and on the distribution of households owning 0, 1, 2, and 3 or more automobiles.

The sensitivity estimates are generated by using a disaggregate automobile ownership model based on a theory proposed by Beckmann, Gustafson, and Golob (1). The theory postulates that a household trades off reduced consumption of goods and services other than transportation for increased accessibility to opportunities when deciding whether or not to own one or more automobiles. The dependent variables are the probabilities that a given household will choose to own a particular number of automobiles. The explanatory variables include those encompassed in the consumption component of the household trade-off function and those encompassed in the accessibility components. The former variables include household disposable income, fixed costs of automobile ownership, automobile operating costs, and public transit fares. The latter variables include travel times by automobile from the household's location to all possible trip destinations, travel times by public transit to those destinations accessible by transit, and the activity level or opportunities at each destination. Household sociodemographic characteristics are used as segmentation variables.

The Beckmann-Gustafson-Golob automobile ownership theory leads directly to a strict utility model of choice in the manner described by Georgescu-Roegen (2, 3) and Halldin (4). Strict utility models are referred to as Bradley-Terry-Luce models in psychology (5, 6) and are applications of the well-known first-choice axiom of Luce (6). As demonstrated by McFadden (7, 8) and Yellott (9), these models are expressed without loss of generality as multinomial logit models for parameter estimation purposes.

The initial logit parameter estimations of the Beckmann-Gustafson-Golob theory were accomplished by Burns, Golob, and Nicolaidis (10). Lerman and Ben-Akiva (11) independently proposed an alternative model that is less rigorous in its underlying choice theory, more restrictive in its assumptions of a hierarchy of travel choices, but able to estimate the effects of a far larger number of explanatory variables. The two models are excellent complements.

The model requires data of the sort typically collected in urban transportation planning system (UTPS) studies. The case study application presented here uses home interview, transportation network, and land-use data from the Detroit Transportation and Land Use Study (TALUS) (12). A maximum likelihood estimation technique was employed. Maximum likelihood estimations of multinomial logit parameters have been shown by McFadden (7, 8) to be statistically consistent, asymptotically efficient, and unique under very general conditions. The estimators are also asymptotically normal, which permits large sample applications of t-statistics and chi-square statistics in significance tests.

Model calibrations were performed by using random subsamples of households interviewed in the 1965 TALUS home interview survey. Since TALUS expended considerable effort to establish a probability sample of households and since large sample sizes were used in the present study, the random subsamples were judged to be representative cross-sections of 1965 Detroit area households. Separate subsamples were used for model calibration and for calculation of goodness of fit.

As a first step in calibrating the models, households were divided into choice-constraint segments based on the maximum feasible number of automobiles they were assumed to consider. This maximum feasible number of automobiles is generally equal to the number of driver-aged household members. However, for some low-income households, the maximum feasible number of automobiles is determined by a constraint on the amount of disposable income available to meet costs of automobile ownership. The rationale of separate calibrations for choice constraint segments has been proposed by Recker and Golob (13). It allows for the possibility that housholds faced with different choice sets weigh the costs and benefits associated with the choices differently in arriving at their final decisions.

In the course of calibrating the models, sensitivity analyses were conducted on several model parameters. These included automobile fixed costs, disposable income definitions, automobile operating costs, and definitions of activities at trip destinations.

Following calibration of the model for the total sample, the households were divided into segments that were homogeneous with respect to their demographic and socioeconomic characteristics. Separate calibrations were then performed for each demographic segment. As discussed by Lovelock (14), Louviere and others (15), and Nicolaidis, Wachs, and Golob (16), such a segmentation allows identification of different sensitivities in the choices of various readily identifiable groups in society.

To study the relative importance of factors affecting household consumption and those affecting household members' accessibilities in household automobile ownership decisions, the effects of changes in these factors must be examined. These effects are captured in a dimensionless measure commonly known as elasticity. Elasticity is a ratio of the resulting percentage change in a dependent variable to the corresponding change in an independent variable. The greater the absolute value of elasticity, the greater the sensitivity of the dependent variable to changes in the independent variable. Expressions were formulated for elasticities of (a) household choice probabilities, (b) expected household automobile ownership, (c) expected aggregate choice frequencies, and (d) expected aggregate automobile ownership with respect to (a) household income, (b) automobile fixed costs, (c) automobile travel times to all destinations, and (d) public transit travel times to destinations reachable by public transit.

These expressions are used in conjunction with results from the model calibrations to determine elasticity values. The values are then interpreted with respect to traffic efficiency and public transit service policies.

## MODEL CALIBRATION

### Total Sample

In the 1965 Transportation and Land Use Study (12) a total of 28 178 households that resided within the boundaries of the 1960 Detroit urban area were interviewed (17). These households were divided into three choice-constraint segments on the basis of the above criteria for determining the maximum feasible number of automobiles for each household.

This division is depicted in Figure 1, where 22.2 percent of the households were postulated to have choices between 0 and 1 automobile; 64.4 percent had choices among 0, 1, and 2 automobiles; and 13.4 percent had choices among 0, 1, 2, and 3 or more automobiles. Again, both number of driver-aged household members and household disposable income were used to segment these households.

Also shown in Figure 1 are the total numbers of households in each choice-constraint segment observed to choose each alternative number of automobiles. These distributions of actual choices affect choice model calibrations and must be taken into account when evaluating the goodness of fit of such models.

Multinomial logit parameter estimates for each choice-constraint segment are shown in Table 1. The sample size for each model is approximately 600. The first three rows in this table list the utility coefficients (the coefficients corresponding to the consumption term and transportation accessibility term are denoted $b_c$ and $b_t$ respectively) and their t-statistics (ratio of coefficient to standard error of coefficient) for the consumption term, transportation accessibility term, and constant. The asymptotic distribution of these t-statistics is Student's t, and therefore they are used to test the null hypotheses $b_c = 0$ or $b_t = 0$.

The 99 percent critical value of the t-distribution for approximately 600 degrees of freedom is 2.33; t-statistics greater than this value have a probability of less than 0.01 of being due to chance.

The consumption term is a function of disposable income and automobile fixed costs. The transportation accessibility term is a function of travel times by automobile from each household's location to all potential trip destinations, the population residing at each destination (a proxy for the attraction of destinations), and travel times by public transit to those destinations accessible by public transit. These terms are described by Burns and Golob (18).

The last row of Table 1 gives a chi-square statistic developed from the ratio of the logarithms of the initial and final likelihoods; it is used to test the joint null hypothesis that $b_c = b_t = 0$ and has three degrees of freedom in the present applications. It can be concluded from these results that, first, the probability that statistics supporting these models are due to chance is extremely low (less than 0.0001) and, second, the relationship between the number of automobiles households choose to own and both the consumption and accessibility variables defined in the present theory is highly statistically significant. Research by Burns and Golob (18) presents additional and encouraging model goodness-of-fit results from previous sensitivity analyses.

The predictive power of these three calibrated models

Figure 1. Choice-constraint segmentation for total sample.

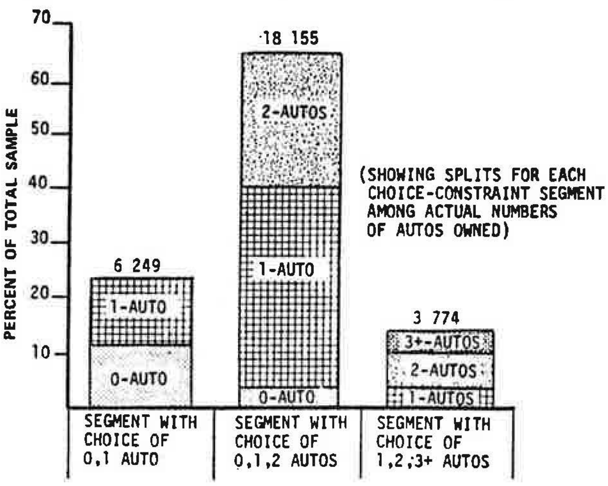



Table 1. Choice model results for total sample.

| Model Parameter | Choice-Constraint Segment: 0, 1 Auto | 0, 1, 2 Autos | 1, 2, 3+ Autos |
|---|---|---|---|
| Coefficient on transportation accessibility term $b_t$ | 0.526 | 0.471 | 0.252 |
| t-Statistic | 13.60 | 22.60 | 16.90 |
| Coefficient on consumption term $b_c$ | 2.13 | 6.96 | 8.02 |
| t-Statistic | 11.60 | 11.40 | 11.70 |
| Constant | 0.598 | 1.43 | 0.816 |
| t-Statistic | 6.37 | 16.90 | 10.20 |
| Choice to which constant assigned | 0 autos | 1 auto | 2 autos |
| Chi-square statistic for likelihood explained by model relative to null hypothesis of choice shares | 127.0 | 116.0 | 96.4 |
| Degrees of freedom | 3 | 3 | 3 |

Table 2. Choice model tests for total sample hold outs.

| Test Statistic | Choice-Constraint Segment: 0, 1 Auto | 0, 1, 2 Autos | 1, 2, 3+ Autos |
|---|---|---|---|
| Aggregate frequencies of households owning particular numbers of automobiles (computed/actual) | | | |
| 0 autos | 518/514 | 77/56 | - |
| 1 auto | 526/530 | 551/562 | 285/285 |
| 2 autos | - | 383/393 | 635/628 |
| 3+ autos | - | - | 339/346 |
| Individual households' choices classified correctly, % | 69 | 60 | 50 |
| Individual households' choices predicted correctly using choice share proportions as aid to random process, % | 50 | 46 | 38 |

Table 3. Choice model results for total equal-proportion sample.

| Model Parameter | Choice-Constraint Segment: 0, 1 Auto | 0, 1, 2 Autos | 1, 2, 3+ Autos |
|---|---|---|---|
| Coefficient on transportation accessibility term $b_t$ | 0.526 | 0.448 | 0.207 |
| t-Statistic | 13.6 | 23.9 | 15.1 |
| Coefficient on consumption term $b_c$ | 2.13 | 10.6 | 7.56 |
| t-Statistic | 11.6 | 17.6 | 11.9 |
| Constant | 0.598 | 0.404 | (constant insignificant) |
| t-Statistic | 6.37 | 4.10 | |
| Choice to which constant assigned | 0 autos | 1 auto | |
| Chi-square statistic for likelihood explained by model relative to null hypothesis of choice shares | 127.0 | 279.0 | 90.7 |
| Degrees of freedom | 3 | 3 | 3 |

Table 4. Sociodemographic factors.

| Factor No. | Percentage of Variance Explained | Component Variables | Correlation Between Factor and Variable |
|---|---|---|---|
| 1 | 25.4 | Marital status of head of household | 0.86 |
| | | Sex of head of household | 0.78 |
| | | No. of licensed drivers | 0.77 |
| | | No. of household members | 0.66 |
| 2 | 19.0 | Age of head of household | 0.83 |
| | | Tenure at address | 0.81 |
| | | Rent or own home | -0.52 |
| | | Education of head of household | -0.49 |
| 3 | 18.0 | Population density of zone of residence | 0.81 |
| | | Race of head of household | 0.80 |
| | | Rent or own home | 0.48 |
| | | Education of head of household | -0.44 |

was next tested on hold-out samples. Probabilities were calculated by using the parameter values calibrated on the original sample and the observed independent variable values for each household in a segment hold-out sample. The aggregate frequencies of households owning each particular number of automobiles were then computed by adding the probabilities for each choice state. In addition, each household was assigned to the choice state with highest calculated probability, and the percentages correctly assigned were tabulated. Results are shown in Table 2.

The descriptive power of each of the models is good. With regard to the less stringent aggregate frequencies test, all computed frequencies were within 2.5 percent of the actual, with the exception of the relatively rare case of the households with choices of 0, 1, and 2 automobiles who chose to own 0 automobiles; these households were over-predicted by 37.5 percent. With regard to the very stringent percent correct classification test, the models each improved classification accuracy by approximately one-third over the best achievable using a priori probabilities based upon the proportions of households choosing to own various numbers of automobiles (i.e., using a random process aided by market share proportions).

The results of the total sample model calibrations are partially dependent on degrees of inequality in choice share proportions. Constants in multinomial logit models adjust for inequalities in choice shares, but, in general, utility coefficients are also affected. Consequently, in order to investigate the relative contributions of the consumption and transportation accessibility terms in explaining the choices of each of the three choice-constraint segments, models were calibrated for samples chosen with equal proportions of each chosen alternative. Results are shown in Table 3. The segment with choice between 0 and 1 automobile had approximately equal choice shares for the total sample, and thus model results are the same as in Table 1.

The presence of constants in a logit model estimated on equal proportion samples indicates that there is a bias in choice toward one or more alternatives not explained by the model variables. The failure to find a constant significantly different from zero is a necessary but not sufficient condition for the full explanatory power of model variables in light of random disturbances. Thus, there is justification in interpreting the results of Table 3 to mean that the choices of households among 1, 2, and 3+ automobiles are more fully explained in terms of the present theory than are the choices of households between 0 and 1 automobile, and among 0, 1, and 2 automobiles. In other words, choices involving the alternative of 0 automobiles are more difficult to explain than choices involving only how many automobiles are to be owned. This conclusion is further strengthened by comparing the chi-square statistics, where degrees of freedom correspond, and t-statistics listed in Table 3.

A second conclusion is that transportation accessibility is more important relative to consumption (disposable income and fixed automobile costs) for households choosing between 0 and 1 automobile than it is for households in the other two choice-constraint segments. This conclusion is based on comparisons of utility coefficients and is only ordinal.

## Sociodemographic Segments

The total sample of Detroit urban area households was segmented on the basis of similarities in patterns of sociodemographic characteristics. The available characteristics measured in the TALUS home interview are listed below.

| Characteristic | Coding |
|---|---|
| Number of household members | Absolute number |
| Number of licensed drivers | Absolute number |
| Rent or own house | 1 = own, 2 = rent |
| Tenure at address | 1 = 7 weeks or less, 2 = 8-51 weeks, 3 = 1-4 years, 4 = 5-10 years, 5 = over 10 years |
| Education of head of household | 1 = 8 years or less, 2 = 9-11 years, 3 = high school, 4 = college |
| Sex of head of household | 1 = female, 2 = male |
| Race of head of household | 1 = white, 2 = nonwhite |
| Age of head of household | Absolute number |
| Marital status of head of household | 1 = unmarried, 2 = married |
| Population density of traffic analysis zone of residence | Persons per hectare |

The segmentation methodology is similar to that described by Golob and Nicolaidis (19). It involves factor analysis and cluster analysis. The factor analysis is used to summarize the interrelationships among the sociodemographic variables by creating linear combinations of the variables (factors) that are independent of one another. Clustering individual households into homogeneous groups is then conducted in the multidimensional space of the factors; this eliminates redundancies in demographic measures and simplifies interpretation of the resulting segments. A random sample of 935 households was used in the factor and cluster analyses. The total sample of 28 178 households were then assigned to the resulting segments by using multiple discriminant analysis classification procedures.

Three sociodemographic factors were found to account for 62.4 percent of the variance in the original ten variables, and additional factors were judged not to add sufficient descriptive power to warrant the loss in efficiency. The factors are described in Table 4, where the percentage of variance accounted for by each factor and the variables that have high correlations (factor loadings) with each factor are listed.

The selection of an appropriate number of segments is accomplished in a fashion similar to the selection of the number of factors: a cut-off point is located in a clustering "compactness" index (i.e., an index simultaneously measuring within-segment homogeneity and between-segment heterogeneity). A good compactness index is judged to be the Wilks $\lambda$-criterion, the ratio of the determinant of the pooled within-segment scatter matrix to the determinant of the total scatter matrix. In this way four sociodemographic segments were found.

The four sociodemographic segments were next plotted in the space of the three factors to facilitate interpretation. The segments were labeled so as to best represent their positions in the factor space. These labels and the proportions of the total sample in each segment are given below. Essentially, there are two large segments, and two segments that are approximately one-half the size of the large segments.

| No. | Label | Percentage of Total Sample |
|---|---|---|
| 1 | Single-person households | 15.7 |
| 2 | Younger families | 39.8 |
| 3 | Inner-city dwellers | 15.2 |
| 4 | Older families | 29.3 |

Division of each of the four segments into choice-constraint segments led to the aggregate splits depicted in Figures 2 through 5. These figures are analogous to Figure 1 for the total sample. However, some choice-

Figure 2. Choice-constraint segmentation for demographic segment 1.

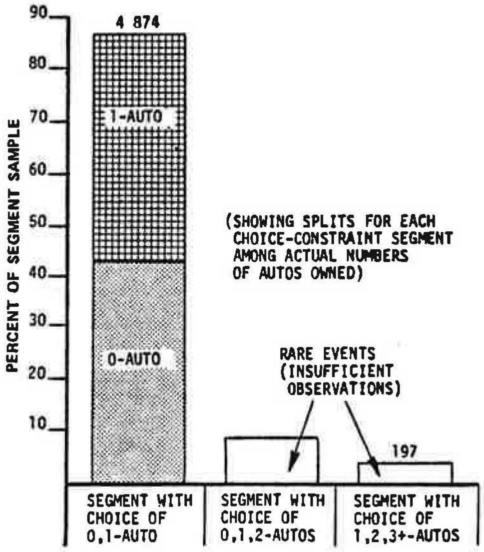



Figure 3. Choice-constraint segmentation for demographic segment 2.

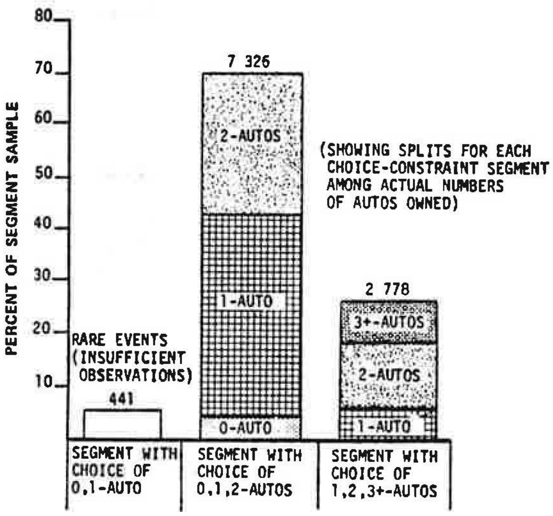
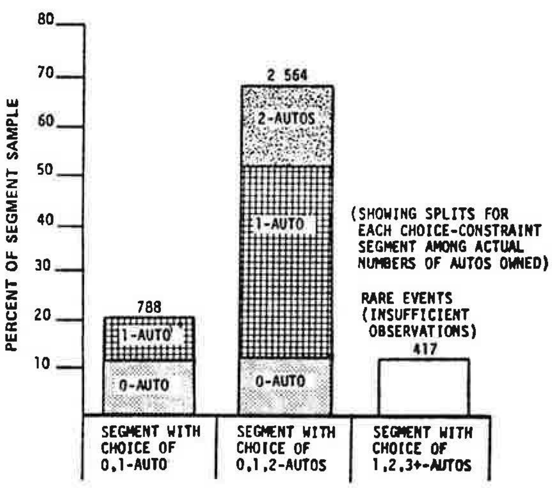



Figure 4. Choice-constraint segmentation for demographic segment 3.




constraint segments essentially did not exist for some sociodemographic segments, because there were insufficient observations to permit choice models to be calibrated for these cases. They are indicated as rare events in Figures 2 through 5.

Interpretation of the figures leads to the conclusion that, for each of the four sociodemographic segments, the distribution of the segment sample into choice-constraint segments is intuitively satisfying, including

Figure 5. Choice-constraint segmentation for demographic segment 4.

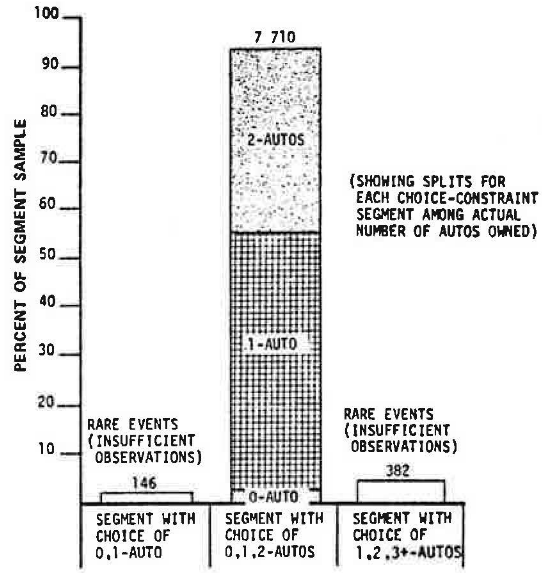



Table 5. Choice model results for sociodemographic segment 1.

| | Choice-Constraint Segment | | |
|---|---|---|---|
| Model Parameter | 0, 1 Auto | 0, 1, 2 Autos | 1, 2, 3+ Autos |
| Coefficient on transportation accessibility term $b_t$ | 0.526 | No model | No model |
| t-Statistic | 11.7 | | |
| Coefficient on consumption term $b_c$ | 2.61 | | |
| t-Statistic | 9.99 | | |
| Constant | 0.615 | | |
| t-Statistic | 5.38 | | |
| Choice to which constant assigned | 0 autos | | |
| Chi-square statistic for likelihood explained by model relative to null hypothesis of choice shares | 109 | | |
| Degrees of freedom | 3 | | |

Table 6. Choice model results for sociodemographic segment 2.

| | Choice-Constraint Segment | | |
|---|---|---|---|
| Model Parameter | 0, 1 Auto | 0, 1, 2 Autos | 1, 2, 3+ Autos |
| Coefficient on transportation accessibility term $b_t$ | No model | 0.433 | 0.215 |
| t-Statistic | | 24.8 | 13.9 |
| Coefficient on consumption term $b_c$ | | 10.9 | 7.80 |
| t-Statistic | | 18.4 | 10.6 |
| Constant | | | |
| t-Statistic | | 0.484 | |
| Choice to which constant assigned | | 5.14<br>1 auto | (constant insignificant) |
| Chi-square statistic for likelihood explained by model relative to null hypothesis of choice shares | | 329 | 70.8 |
| Degrees of freedom | | 3 | 3 |

the occurrence of the rare events.

Choice model results for the sociodemographic segments are given in Tables 5 through 8. The format of these tables is identical to that of Table 3 for the total equal proportion sample. Comparisons among the results shown in Tables 5 through 8 lead to the following conclusions (sizes of the random samples chosen for model calibrations are approximately equal).

First, with regard to choices between 0 automobiles and 1 automobile, significance levels of constants indicate that errors in model specification are expected to be less for demographic segment 3 than for demographic segment 1. That is, choices of inner-city dwellers are more readily explained in terms of the accessibility and consumption variables of the present theory than are the choices of single-person households.

Second, with regard to choices among 0 automobiles, 1 automobile, and 2 automobiles, chi-square statistics and t-statistics indicate that choices of younger families are most effectively explained, then come choices of older families, and finally are choices of inner-city dwellers. In addition, the relative utility weights indicate that accessibility is a more important consideration relative to consumption for inner-city dwellers, which corresponds to an expected higher average level of public transit service for these people.

Comparisons involving choices among 1 automobile, 2 automobiles, and 3+ automobiles were not possible because only one sociodemographic segment had a choice model calibrated for this choice-constraint segment.

Table 7. Choice model results for sociodemographic segment 3.

| Model Parameter | Choice-Constraint Segment 0, 1 Auto | 0, 1, 2 Autos | 1, 2, 3+ Autos |
|---|---|---|---|
| Coefficient on transportation accessibility term $b_t$ | 0.390 | 0.509 | No model |
| t-Statistic | 8.62 | 17.3 | |
| Coefficient on consumption term $b_c$ | 1.87 | 9.81 | |
| t-Statistic | 8.04 | 18.1 | |
| Constant | (constant insig- nificant) | | |
| t-Statistic | | -0.338 | |
| Choice to which constant assigned | | -2.94 | |
| Chi-square statistic for likelihood explained by model relative to null hypothesis of choice shares | 41.2 | 167.0 | |
| Degrees of freedom | 2 | 3 | |

Table 8. Choice model results for sociodemographic segment 4.

| Model Parameter | Choice-Constraint Segment 0, 1 Auto | 0, 1, 2 Autos | 1, 2, 3+ Autos |
|---|---|---|---|
| Coefficient on transportation accessibility term $b_t$ | No model | 0.457 | No model |
| t-Statistic | | 27.5 | |
| Coefficient on consumption term $b_c$ | | 14.2 | |
| t-Statistic | | 20.0 | |
| Constant | | | |
| t-Statistic | | 0.371 | |
| Choice to which constant | | 3.83 | |
| assigned | | 1 auto | |
| Chi-square statistic for likelihood explained by model relative to null hypothesis of choice shares | | 278 | |
| Degrees of freedom | | 3 | |

## ELASTICITY CALCULATIONS

### Total Sample

Elasticities of automobile ownership in the Detroit metropolitan area in 1965 with respect to certain important model explanatory variables are shown in Table 9. These include elasticities for the expected number of households owning alternative numbers of automobiles and the overall elasticity for the aggregate stock of automobiles held by all households.

Automobile ownership was found to be three times more sensitive to changes in disposable income or automobile fixed costs than to uniform percentage changes in automobile travel times throughout the metropolitan area: a 10 percent increase in all incomes would lead to a 3 percent increase in the aggregate stock of automobiles, while a 10 percent increase in automobile travel times would lead to a 1 percent decrease in aggregate stock. Travel time by public transit has one-half the effect of travel time by automobile and almost the same effect as travel time by automobile to destinations located within the city of Detroit.

The net number of households owning 1 automobile is relatively insensitive to changes in disposable income, fixed costs of automobiles, or any travel times. This is because about the same number of households move from 0 automobile to 1 automobile states as move from 2 automobiles to 1 automobile for opposite types of changes. With respect to the relative effects of the consumption term versus the transportation accessibility term, the number of households owning 0 automobiles and the number of households owning 3+ automobiles are most sensitive to income and automobile fixed costs. However, the number of households owning 0 automobiles also has the highest sensitivity to transportation accessiblity variables; the number of households owning 3+ automobiles, together with the number owning 2 automobiles, has only modest sensitivity to transportation accessibility variables.

The results in Table 9 were developed through aggregation of results for each of the three total sample choice-constraint segments. Tables 10 through 12 list these more detailed results for comparison purposes.

Table 9. Elasticities of total sample.

| Variable | Elasticity of Expected No. of Households: Owning 0 Auto | Owning 1 Auto | Owning 2 Autos | Owning 3 Autos | Elasticity of Aggregate Stock of Autos Held by All Households |
|---|---|---|---|---|---|
| Consumption term: | | | | | |
| Disposable income (= negative of fixed costs of all autos) | -0.94 | -0.08 | 0.49 | 0.89 | 0.29 |
| Transportation accessibility term: | | | | | |
| Travel by auto to all destinations | 0.30 | 0.05 | -0.21 | -0.21 | -0.10 |
| Travel by auto to destinations in city of Detroit (only) | 0.11 | 0.02 | -0.08 | -0.08 | -0.04 |
| Travel by public transit to all destinations | -0.17 | -0.01 | 0.10 | 0.09 | 0.05 |
| Travel by public transit to destinations in city of Detroit (only) | -0.10 | -0.01 | 0.06 | 0.06 | 0.03 |

Households faced with choices between 0 and 1 automobile (Table 10) are most sensitive to changes in both consumption and accessibility term variables; households faced with choices among 1, 2, and 3 or more automobiles are least sensitive to such changes. Since households in the latter choice-constraint segment generally have higher incomes, the difference in consumption term elasticities is consistent with traditional economic theories. Households in the latter segment in general are also more suburbanized, so the difference in accessibility term elasticities can be interpreted to reflect the relative unavailability of alternatives to travel by automobile outside the city of Detroit. Many such relationships between elasticities and socioeconomic and location patterns are explored later in this report.

A final issue in this section is the effect of automobile fixed costs. Because of the assumptions underlying the utility theory model, treating automobiles as homogeneous economic goods equates the absolute value of the effect of disposable income and automobile fixed costs. It is beyond the scope of the present theory to distinguish between new and used automobiles. If, however, second and third automobiles held by households are postulated to be affected by exogenous inputs (such as insurance costs) to a different extent than first automobiles are, differences in sensitivities to income and automobile fixed costs can be investigated in terms of the present theory.

Assume that changes in fixed costs of second and third automobiles are less than changes in fixed costs of first automobiles by a fixed percentage. For example, if costs of the first or primary automobile held by households increase by 10 percent, the costs of second and third automobiles increase by 7.5 percent (i.e., fixed cost increases for additional automobiles are 75 percent of fixed cost increases for primary automobiles). Fixed automobile cost elasticities for the total sample have been estimated under such an assumption, and results are graphed as a function of percentage difference between primary and second and third automobile cost differences in Figure 6. Fixed automobile cost elasticities range linearly from on the order of one-half the income elasticity to the income elasticity over the entire domain of possible percentage differences. These results are

Table 10. Elasticities of choice-constraint segment with choice of 0 or 1 auto.

| Variable | Elasticity of Expected No. of Households in the Choice-Constraint Segment: Owning 0 Auto | Owning 1 Auto | Elasticity of Aggregate Stock of Autos Held by Segment Households |
|---|---|---|---|
| Consumption term: | | | |
| Disposable income (= negative of fixed costs of all autos) | -0.72 | 0.72 | 0.72 |
| Transportation accessibility term: | | | |
| Travel by auto to all destinations | 0.21 | -0.21 | -0.21 |
| Travel by auto to destinations in city of Detroit (only) | 0.08 | -0.08 | -0.08 |
| Travel by public transit to all destinations | -0.12 | 0.12 | 0.12 |
| Travel by public transit to destinations in city of Detroit (only) | -0.07 | 0.07 | 0.07 |

Table 11. Elasticities of choice-constraint segment with choice of 0, 1, or 2 autos.

| Variable | Elasticity of Expected No. of Households in the Choice-Constraint Segment: Owning 0 Auto | Owning 1 Auto | Owning 2 Autos | Elasticity of Aggregate Stock of Autos Held by Segment Households |
|---|---|---|---|---|
| Consumption term: | | | | |
| Disposable income (= negative of fixed costs of all autos) | -1.44 | -0.22 | 0.61 | 0.26 |
| Transportation accessibility term: | | | | |
| Travel by auto to all destinations | 0.47 | 0.12 | -0.26 | -0.10 |
| Travel by auto to destinations in city of Detroit (only) | 0.17 | 0.04 | -0.10 | -0.04 |
| Travel by public transit to all destinations | -0.27 | -0.04 | 0.12 | 0.05 |
| Travel by public transit to destinations in city of Detroit (only) | -0.16 | -0.03 | 0.07 | 0.03 |

Table 12. Elasticities of choice-constraint segment with choice of 1, 2, or 3+ autos.

| Variable | Elasticity of Expected No. of Households in the Choice-Constraint Segment: Owning 1 Auto | Owning 2 Autos | Owning 3+ Autos | Elasticity of Aggregate Stock of Autos Held by Segment Households |
|---|---|---|---|---|
| Consumption term: | | | | |
| Disposable income (= negative of fixed costs of all autos) | -1.07 | 0.01 | 0.87 | 0.23 |
| Transportation accessibility term: | | | | |
| Travel by auto to all destinations | 0.22 | -0.01 | -0.21 | -0.05 |
| Travel by auto to destinations in city of Detroit (only) | 0.08 | 0.00 | -0.08 | -0.02 |
| Travel by public transit to all destinations | -0.12 | 0.00 | 0.10 | 0.03 |
| Travel by public transit to destinations in city of Detroit (only) | -0.07 | 0.00 | 0.06 | 0.02 |

Figure 6. Implied elasticity of aggregate stock of automobiles versus change in costs of automobiles.

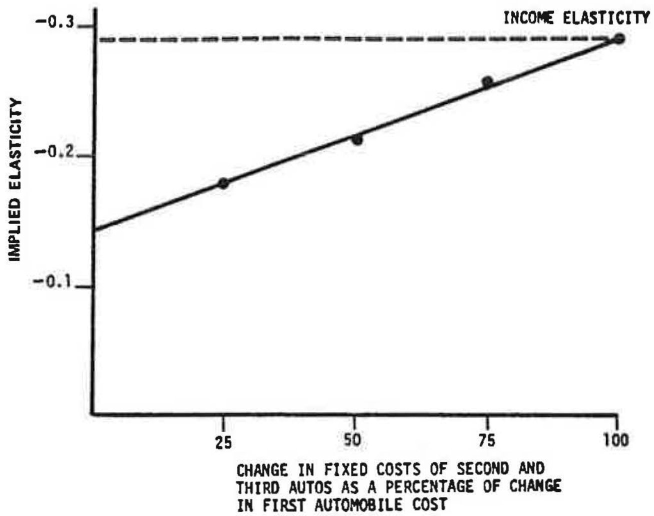



Table 13. Elasticities of aggregate stock of autos.

| Sociodemographic Segments No. | Type | Percentage of Sample | Disposable Income | Areawide Auto Travel Time | Areawide Transit Travel Time |
|---|---|---|---|---|---|
| 1 | Single-person households | 16 | 0.68 | -0.19 | 0.11 |
| 2 | Younger families | 40 | 0.48 | -0.12 | 0.06 |
| 3 | Inner-city residents | 15 | 0.77 | -0.20 | 0.13 |
| 4 | Older families | 29 | 0.66 | -0.16 | 0.07 |
| Total | | 100 | 0.51 | -0.14 | 0.07 |

interpreted to be evidence that the true automobile fixed-cost elasticity lies within this range.

## Sociodemographic Segments

Elasticities of the aggregate stock of automobiles owned by all households in a sociodemographic segment with respect to selected variables are shown for each of the four demographic segments in Table 13. The three variables are disposable income, automobile travel time to all destinations, and public transit travel time to all destinations. Results for the total sample are shown for comparative purposes. These total sample results are different from those listed in Table 9, since they are generated by using equal proportional sampling choice model parameters (Table 3), as opposed to random sampling (Table 1). This is necessary in order to match the sampling underlying the results for the sociodemographic segments.

The choices of inner-city residents are most sensitive to changes in disposable income and both automobile and transit travel times. The choices of younger families are least sensitive to changes in these variables explaining automobile ownership behavior. The effects of the two variables representing the transportation accessibility term are greatest relative to the effect of income for single-person households.

With regard to the two accessibility variables, the choices of inner-city residents and single-person households are most sensitive to public transit travel time relative to automobile travel time. Since these two segments are the least suburbanized, this result is consistent with the conclusion that public transit is a more significant factor in automobile ownership decisions in higher density areas where its service level is higher.

## CONCLUSIONS

The aggregate stock of automobiles held by urban households (all automobiles owned, both new and used) is sensitive to transportation accessibility factors—travel times by automobile and by public transit—as well as to income and automobile ownership cost. Using the Detroit area as a case study, the following sensitivities were found from 1965 cross-sectional data:

1. The elasticity of automobile ownership with respect to automobile travel times is 0.1. It may therefore be inferred that a 10 percent decrease in automobile travel times experienced by all households, as a result perhaps of road or traffic efficiency improvements, would cause a 1 percent increase in automobile ownership, all else being held constant. The elasticity of automobile ownership with respect to public transit travel times is 0.05, or one-half that of automobile travel times.
2. The automobile ownership decisions of inner-city dwellers are more sensitive to travel time accessibility factors (and income) than are those of suburban dwellers. The ownership decisions of young families are less sensitive to these same factors than are those of older families.
3. The elasticity of automobile ownership with respect to disposable income is 0.3, or three times that of automobile travel times and six times that of public transit travel times.

These conclusions are derived from elasticity calculations using cross-sectional data at one point in time. While such calculations are conceptually different from elasticity calculations that use time-series data, they do provide a basis on which to compare relative effects of different variables.

Implications of these results are twofold. First, transportation policies aimed at improving traffic efficiency within urban areas can be expected to increase automobile ownership levels. Second, policies aimed at improving public transit service can be expected to decrease automobile ownership levels, but the absolute value of this effect will be approximately one-half that of the traffic efficiency effect (for an equal percentage change in overall automobile or bus trip time). These effects should not be ignored when assessing the costs and benefits associated with plans affecting traffic efficiency or public transit service.

## ACKNOWLEDGMENTS

Martin J. Beckmann of Brown University and the Technical University of Munich, and a consultant to General Motors Research Laboratories, provided invaluable inputs to the development of the theory underlying this research. Richard L. Gustafson, now at Portland State University, and Gregory C. Nicolaidis, General Motors Research Laboratories, co-authored earlier reports on the project. Ellen M. Racusin, also of General Motors Research Laboratories, prepared much of the data for analyses. We gratefully acknowledge these contributions.

## REFERENCES

1. M. J. Beckmann, R. L. Gustafson, and T. F. Golob. Locational Factors in Automobile Owner-

ship Decisions. Annals of Regional Science, Vol. 7, 1973, pp. 1-12.
2. N. Georgescu-Roegen. Threshold in Choice and the Theory of Demand. Econometrica, Vol. 26, 1958, pp. 157-168.
3. N. Georgescu-Roegen. The Relation Between Binary and Multiple Choices: Some Comments and Further Results. Econometrica, Vol. 37, 1969, pp. 728-730.
4. C. Halldin. The Choice Axiom, Revealed Preferences, and the Theory of Demand. Theory and Decision, Vol. 5, 1974, pp. 139-160.
5. R. A. Bradley and M. E. Terry. Rank Analysis of Incomplete Block Designs: I. The Method of Paired Comparisons. Biometrika, Vol. 39, 1952, 324-345.
6. R. D. Luce. Individual Choice Behavior. Wiley, New York, 1959.
7. D. McFadden. The Revealed Preferences of a Government Bureaucracy. Project for the Evaluation and Optimization of Economic Growth, Institute of International Studies, Univ. of California, Berkeley, Technical Rept. No. 17, 1968.
8. D. McFadden. Conditional Logit Analysis of Qualitative Choice Behavior. In Frontiers in Econometrics (P. Zaremka, ed.), Academic Press, New York, 1974.
9. J. I. Yellot. The Relationship Between Luce's Choice Axiom, Thurstone's Theory of Comparative Judgment, and the Double Exponential Distribution. Journal of Mathematical Psychology, Vol. 15, 1977, pp. 109-144.
10. L. D. Burns, T. F. Golob, and G. C. Nicolaidis. A Theory of Urban Households' Automobile Ownership Decisions. TRB, Transportation Research Record 569, 1975, pp. 56-74.
11. S. R. Lerman and M. Ben-Akiva. Disaggregate Behavioral Model of Automobile Ownership. TRB, Transportation Research Record 569, 1976, pp. 34-55.
12. Growth, Change, and a Choice for 1990: Preliminary Plan, Southeast Michigan, Detroit. Detroit Regional Transportation and Land Use Study, 1965.
13. W. Recker and T. F. Golob. A Behavior Travel Demand Model Incorporating Choice Constraints. Advances in Consumer Research, Vol. 3, 1976, pp. 416-425.
14. H. Lovelock. A Market Segmentation Approach to Transit Planning. Proc., Transportation Research Forum, Vol. 16, 1975, pp. 247-258.
15. J. J. Louviere, L. M. Ostresh, D. H. Henley, and R. J. Meyer. Travel-Demand Segmentation: Some Theoretical Considerations Related to Behavioral Modeling. In Behavioral Travel-Demand Modeling (P. R. Stopher and A. H. Meyburg, eds.), Heath, Lexington, MA, 1976.
16. G. C. Nicolaidis, M. Wachs, and T. F. Golob. Evaluation of Alternative Market Segmentations for Transportation Planning. TRB, Transportation Research Record 649, 1977, pp. 23-31.
17. U.S. Census of Housing: 1960. Bureau of the Census, U.S. Department of Commerce, Vol. 1, States and Small Areas: Michigan, Final Rept. HC(1)-24, 1962.
18. L. D. Burns and T. F. Golob. The Role of Accessibility in Basic Transportation Choice Behavior. Transportation, Vol. 5, 1976, pp. 175-198.
19. T. F. Golob and G. C. Nicolaidis. Comparison of Segmentations for Modeling Consumers' Preferences for Transportation Modes. Paper presented at the American Institute for Decision Sciences Annual National Conference, San Francisco Nov. 10-13, 1976.

*Publication of this paper sponsored by Committee on Passenger Travel Demand Forecasting and Committee on Traveler Behavior and Values.*

**Mr. Golob was with General Motors Research Laboratories when this paper was written.*

***On January 1, 1979, the Transportation and Urban Analysis Department became the Transportation and Traffic Science Department.*

# Perceptual Market Segmentation Technique for Transportation Analysis

Ricardo Dobson*, Charles River Associates, Inc., Cambridge, Massachusetts
Mary Lynn Tischer, Federal Highway Administration

**A primary aim of this research is to illustrate a relatively uncomplicated and effective perceptual market segmentation procedure for transportation policy analysis. This illustration is achieved through a flowchart describing the technique, an empirical application, and tests of the reliability of the derived market segmentation structures across split halves of a data set. The procedure was calibrated on a sample of Los Angeles central business district workers. The segmentation structure, which was derived for the full sample, readily distinguished the perceptual groups and correlated highly with appropriate mode-choice patterns. It was also observed that perceptual segmentation membership was a stronger determinant of mode choice than zone network times and costs. The split sample analyses showed reliable relationships across halves and confirmed the mode-choice linkage of perceptual segments relative to network times and costs. Among the practical implications of the segmentation procedure are its use in developing short-range forecasting models and its potential for developing information aids to target groups of travelers.**

The concept of market segmentation for consumer research issues was introduced into transportation analysis in the early and middle 1970s. Several major reviews address market segmentation and how it can be used to facilitate transportation analysis (1, 2, 3, 4). Transportation analysts have long grouped households on the basis of geographic proximity. However, the introduction of the market segmentation concept has pointed to the fact that spatial arrangement is not the sole basis for agglomeration. Golob and Dobson (5) have suggested that perceptions and preferences may serve as a useful basis for grouping households or individuals.

Three broad areas have received attention in considerations of direct consumer research applications of market segmentation. These are the analysis of behavioral intentions, the estimation of factors underlying objective choices for travel options, and transit marketing by operating agencies.

Studies directed at the analysis of behavioral intentions are most often concerned with innovative or evolutionary transportation systems. Costantino, Dobson, and Canty (6) supported the hypothesis that a more thorough understanding of individuals' preferences for ways of traveling to work or shop could be obtained by stratifying a sample into homogeneous groups rather than by considering respondents as an undifferentiated set. Dobson and Kehoe (7) investigated the validity of market segments based on perceptual dimensions and determined that one market segment's dimensions were better at estimating its satisfactions than other segments' dimensions.

Studies attempting to explain objective travel choices are most often concerned with analyzing the operation of an existing transportation system. The goals of these analyses are frequently the diagnosing of current operations in a search for system enhancements and/or the devising of a better forecasting model. Hensher (8) found that employment commitment, size and composition of household, and age distribution of travelers in a household were related to shopping trip frequency. Recker and Golob (9) found that market segments based on perceived constraints for the ability to easily use a way of traveling emphasize alternate classes of system attributes as determinants of a way of getting to work. Dobson and Tischer (10, 11) reported that choice models based solely on system perceptions were statistically significant beyond the 0.01 level.

There are a limited number of actual applications of market segmentation strategies to transit marketing. Some of these are documented in a marketing publication issued by the Urban Mass Transportation Administration entitled Pricing—Transit Marketing Management (12). For example, the San Diego Transit Commission raised the general fare to a total of 35 cents. However, fares for senior citizens and handicapped patrons were set at 15 cents, and the student fare remained at 25 cents. While general patronage declined slightly in response to this segmentation policy, there were substantial increases in the degree of bus use among groups spared the fare increase. Furthermore, these patronage adjustments were accompanied by an overall revenue increase. Other cities that instituted different segmentation policies include New York and Chicago.

This paper is designed to extend the literature on market segmentation for transportation analysis. One principal goal of the report is to illustrate a relatively simple market segmentation procedure for use in consumer research on traveler choices. This technique was previously discussed from theoretical and research perspectives by Dobson (13) and Horowitz and Sheth (14).

## STUDY DESIGN

### Sample

The sample comprised 874 individuals who work in the Los Angeles downtown area and who live within 3 km (2 miles) of a freeway that feeds radially into that area. Further details on the sample characteristics are given elsewhere (11, 15).

### Data

Only three of the many types of data collected from home interviews are presented here: frequency of mode use, perceived system attributes, and socio demographic data on the individual and his or her household. An additional data set composed of network time, distance, and cost was derived for the sample at the geographic zone level and is assigned to individuals depending on their origin and destination zones.

Individuals were asked how frequently per month they used each of three modes—the single-occupant auto, the bus, and the carpool—to travel to work. The frequency of use for each mode was subtracted from every other mode to obtain comparisons of mode use. These constructed variables, frequency of auto use minus frequency of bus use, frequency of auto minus carpool use, and frequency of bus minus carpool use, are used throughout the analysis as the dependent variables.

Beliefs about 19 attributes for each of the modes were collected with a semantic differential format. The individual is asked to locate his or her perception of each mode on a seven-point scale anchored at each extreme by opposite descriptors of an attribute. Marital status, number of people in the household, dwelling type, income, age, race, sex, life cycle, and auto ownership were obtained from the respondent.

The network time and cost data were constructed to follow standard procedures. These data are estimated for small geographic units or zones as they relate to the transportation network of arterials, highways, and transit routes. The auto network data were supplied by the California Department of Transportation. Carpool data were calculated by adjusting times for the picking up of passengers and assuming 2.28 passengers per carpool. The transit data were developed from schedules and maps provided by the Southern California Rapid Transit District and other relevant transit-operating agencies. Additional information regarding the computation and definitions for the network data can be found in Dobson and Tischer (11).

### Research Procedures

A conceptually simple and easily implemented procedure to segment respondents on their perceptions of transportation alternatives is discussed. The segmentation structure that results from the application of the procedure is compared with engineering data to determine its merits, separately or in conjunction with network data, for predicting the choice of mode to work. Finally, the segmentation structure is tested for stability across halves.

The segmentation procedure follows the outline in Figure 1. The first step can be predicated on a large number of belief judgments (e.g., 57 for the current data set) and the expectation that many beliefs would be related as elements of broader concepts. Principal components factor analyses with varimax rotation are performed on the belief judgments about each mode. The belief factors explaining the largest percentage of shared variance for beliefs about each mode (e.g., carpools, single-occupant autos, and buses) are saved, and vari-

Figure 1. Segmentation procedure.

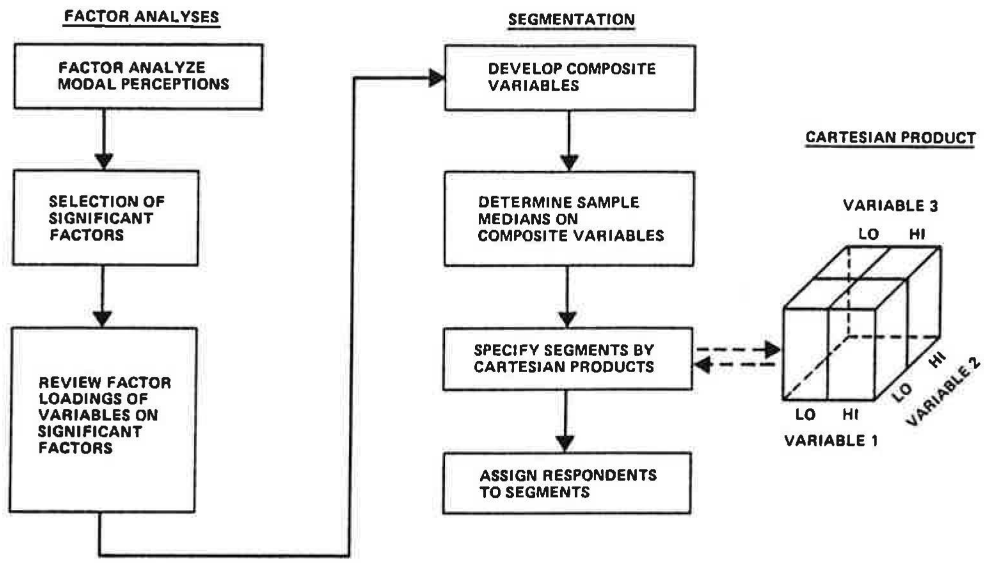



Table 1. Factors selected for segmentation procedure.

| | Definition for Sample | | |
|---|---|---|---|
| Factor Label | (N = 874) | (N = 435) | (N = 439) |
| Bus convenience | Convenience<br>Ease of use<br>Reliability<br>On-time arrival<br>Ease to destination<br>Wait for vehicle | Convenience<br>Ease of use<br>Reliability<br>On-time arrival<br>Ease to destination<br>Wait for vehicle<br>Parking cost | Convenience<br>Ease of use<br>Reliability<br>On-time arrival<br>Ease to destination<br>Wait for vehicle<br>Flexible schedule |
| Carpool convenience | Convenience<br>Ease of use<br>On-time arrival<br>Ease to destination<br>Crowding<br>Wait for vehicle<br>Weather<br>Extra time | Convenience<br>Ease of use<br>On-time arrival<br>Reliability | Convenience<br>Ease of use<br>On-time arrival<br>Ease to destination<br>Crowding<br>Wait for vehicle<br>Weather<br>Extra time |
| Single-occupant auto comfort and convenience | Comfort<br>Relaxing<br>Convenience<br>Flexible schedule<br>Ease of use<br>On-time arrival<br>Ease to destination | Comfort<br>Convenience<br>Flexible schedule<br>Ease of use<br>On-time arrival<br>Ease to destination | Comfort<br>Weather<br>Crowding<br>Ease of use<br>On-time arrival<br>Ease to destination<br>Wait for vehicle<br>Extra time |

ables with high factor loadings on those factors are combined to form composite variables. The composite variables are merely the average of the semantic differential scores for the attributes that load highly on a factor. If it is known from a priori considerations which variables are important, these variables can be used to estimate composite variables without analyses and varimax rotations. The latter statistical techniques are not a mandatory component of the perceptual segmentations procedure.

The sample is assigned categorical scores on the basis of position above or below the median on each composite variable. The three variables are then crossed by using a Cartesian product process as represented in Figure 1; eight cells result. Respondents are segmented into eight groups. For example, one group is characterized as being below the median on all three composite variables. Another group is above the median on one composite variable and below the median on the other two.

The comparative utility of segmented belief data, in place of or in addition to network data, is tested by using a general variance component analysis procedure. Through a regression technique, the variance of mode use explained by the two data sets and their interaction are tested separately. The additional contribution of each independent variable set to explaining the dependent variable can be determined by calculating $R^2$ with and without the particular variable set included.

First, a regression is performed for all independent variables, as in the following equation:

$$Y = a_0 + a_1 X_1 + a_2 X_2 \tag{1}$$

Then a second regression that omits the variable of interest is performed for example $X_2$.

$$Y = b_0 + b_1 X_1 \tag{2}$$

The $R^2$ resulting from Equation 2 is subtracted from Equation 1, leaving the amount of variance in Y explained by $X_2$ above and beyond that explained by $X_1$. Testing for interaction follows the same procedure as that described above. Two regression equations are necessary, but a new component, the interaction term, is added. The significance of each type of variable can be determined with standard statistical procedures.

The stability of the segmentation structure is tested by randomly splitting the sample in half, redefining the groups, and applying one half's definition of the belief composites to the other half. If the perceptual segmentation structure transfers from one half to the other, then more general transferability is demonstrated.

## FINDINGS

### Full Sample Analysis

The first major step of the market segmentation procedure outlined in Figure 1 is a factor analysis of perceptual judgments for alternative transport modes. Across all three modes, 14 factors were extracted. The number of factors retained for buses, carpools, and single-occupant autos were 5, 4, and 5 respectively. In no case was the gap between the largest and second largest factor with respect to explained variance less than eight percentage points, and the largest factors accounted for 31-39 percent of the variance for their respective modes.

The factors with the largest percentage of variance for the judgments about a mode were used for the market segmentation analysis. Three factors were retained, one factor per mode. The largest factor for buses was labeled "convenience." The largest carpool factor was also convenience. The name assigned to the single-occupant auto's largest factor was "comfort and convenience." The definitions for all three factors are shown in Table 1. Composite variables were computed and market segments were derived according to the method described above.

Market segments are labeled according to their relative scores on the three composite variables defined above. Because there are three composite variables and two classes (i.e., above or below the median), eight mutually exclusive and exhaustive groups can be derived from the Cartesian product of the variables. The group of commuters that was above the median on all three composite variables is designated B+, CP+, SOA+. Commuters who viewed carpools in a relatively positive way, but were comparatively negative in their perceptions of the other two modes, are labeled B-, CP+, SOA-. Finally, the appropriate designation for commuters who judged all modes in a relatively negative fashion is B-, CP-, SOA-. The number of commuters in each market segment is as follows:

1. N (B-, CP-, SOA-) = 127,
2. N (B-, CP-, SOA+) = 138,
3. N (B-, CP+, SOA-) = 78,
4. N (B-, CP+, SOA+) = 104,
5. N (B+, CP-, SOA-) = 105,
6. N (B+, CP-, SOA+) = 61,
7. N (B+, CP+, SOA-) = 146, and
8. N (B+, CP+, SOA+) = 115.

Figure 2 presents average attribute scores on the 19 original variables for three characteristic segments: B-, CP-, SOA+; B-, CP+, SOA-; and B+, CP-, SOA-. There is a separate frame for each mode. Apart from a couple of minor exceptions, groups segmented to favor a mode have uniformly more positive perceptions toward that mode than other groups. The first frame of Figure 2 shows segment B+, CP-, SOA- to have the largest bus ratings for 18 of 19 attributes. Similar patterns for the favored mode are exhibited by segments B-, CP+, SOA- and B-, CP-, SOA+ in the second and third frames respectively.

For the single-occupant auto and to a lesser degree for buses and carpools, the differences in the perceptions of groups can be characterized by degree rather than kind. The shapes of the profiles across segments are similar for the same mode. Furthermore, it can be noted that the perceptual profiles reflect widespread characterizations of system alternatives. For example, single-occupant autos are viewed as being compatible with a flexible schedule and ease of use, although it is recognized that autos are expensive. On the other hand, buses are seen as being inexpensive, but they expose passengers to unpleasant conditions such as inclement weather and crowding. The high degree of comparability for the modes across segments and the intuitive validity of characterizations emphasize the vertical nature of perceptual judgments and support their widespread use in urban transportation planning.

Figure 3 presents profiles of behavioral and sociodemographic data for the following three representative segments: B-, CP+, SOA-; B+, CP-, SOA-; and B-, CP-, SOA+. Among the most central questions about the perceptually based market segments are those about their relationship to mode use. Do commuters identified as favoring a mode usually travel by that mode? The first two sets of bars of Figure 3 answer this question affirmatively.

Sex, income, and auto ownership are associated with mode perceptions in corresponding ways. For example, segments with positive views of high-occupancy vehicles have a disproportionately large percentage of females, low-income workers, and individuals from households with zero autos. The lowest income and auto ownership characteristics and highest percentage of females are found in B+, CP-, SOA-. A positive view of carpooling is associated with households of three or more individuals, but not particularly with marital status. Apparently, children create an extra emphasis on keeping an auto at home. Carpooling makes a family auto more readily available to nonworking family members.

Mode choice is often accounted for by network data in urban transportation planning analyses. Therefore, perceptual segments should account for mode choice better or at least add extra explanatory power to a choice model calibrated with network data. In order to help evaluate the usefulness of the perceptual segments, three types of choice models are calibrated. The types of models correspond to the differences in how often per month commuters use alternate modes. The three dependent variables for the models are auto minus bus, auto minus carpool, and bus minus carpool. The network independent variables consist of impedance and cost differences for the corresponding models. Seven dummy variables represent the perceptual segment membership. With these variables, each commuter could be coded as belonging to one of the eight perceptual segments.

The table below shows the percentage of variance accounted for in the three dependent variables by four types of independent variables.

| | Dependent Variables | | |
|---|---|---|---|
| Independent Variables | A-B | A-CP | B-CP |
| Network (NT) | 0.03 | 0.00 | 0.02 |
| Beliefs (B) | 0.27 | 0.12 | 0.16 |
| NT + B | 0.28 | 0.12 | 0.16 |
| NT + B + NT * B | 0.29 | 0.14 | 0.21 |

These four types of independent variables are network data alone (NT), belief data alone (B), network and belief data combined in an additive model (NT + B), and network and belief data combined in additive and multiplicative fashion (NT + B + NT * B). With one exception, NT for A-CP, all modes are significant at beyond the 0.01 level. However, the most interesting finding is that the perceptual segment data always explain mode choice better than network data. Furthermore, there is at best only a slight improvement when network data are additively introduced to a belief model of mode choice. When network and belief data are combined in additive and multiplicative fashions, the results again represent only a slight improvement over a mode-choice model calibrated solely with respect to the belief segments.

The following table reports the significance level of the main and interaction terms for the network and belief data with respect to mode choice.

| | Variables Held | Dependent Variables | | |
|---|---|---|---|---|
| Data Types | Constant | A-B | A-CP | B-CP |
| NT | B + NT * B | $p < 0.01$ | $p > 0.05$ | $p < 0.05$ |
| B | NT + NT * B | $p < 0.001$ | $p < 0.01$ | $p > 0.05$ |
| NT * B | B + NT | $p > 0.05$ | $p < 0.05$ | $p < 0.001$ |

The technique used to compute statistical significance was referenced earlier in the research procedures subsection of the study design. Because of the way that the variance accounting properties of the belief data can be absorbed into the interaction terms between network and belief data, the belief data are found not to be statistically significant when tested against a model that includes a significant interaction set of terms for the bus minus carpool dependent variable. In any event, the belief data are superior to the network data for two of three dependent variables. The results shown in the above tables indicate that perceptual segment data account for mode choice better than network data and that belief data can add substantially to the explanatory power of a

Figure 2. Beliefs about buses, carpools, and single-occupant autos for selected market segments.

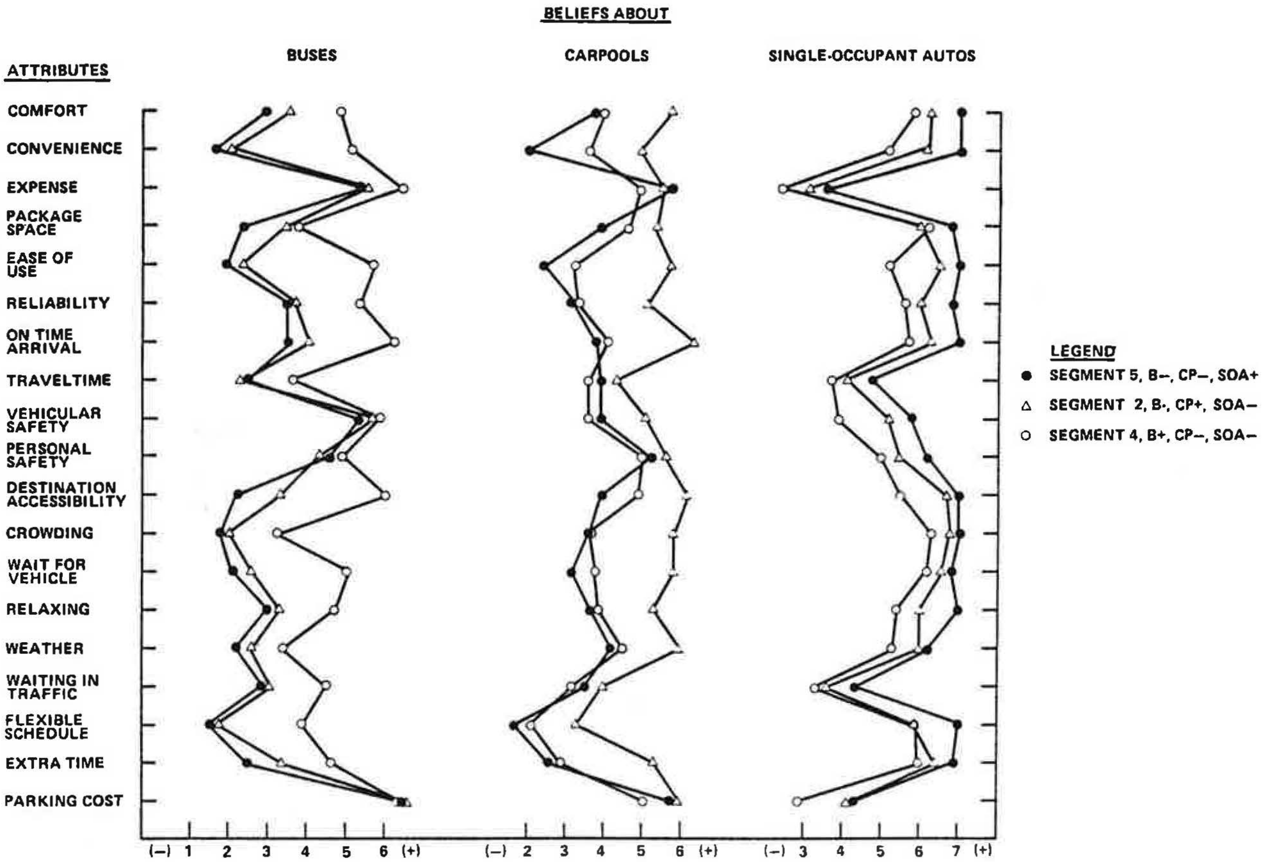
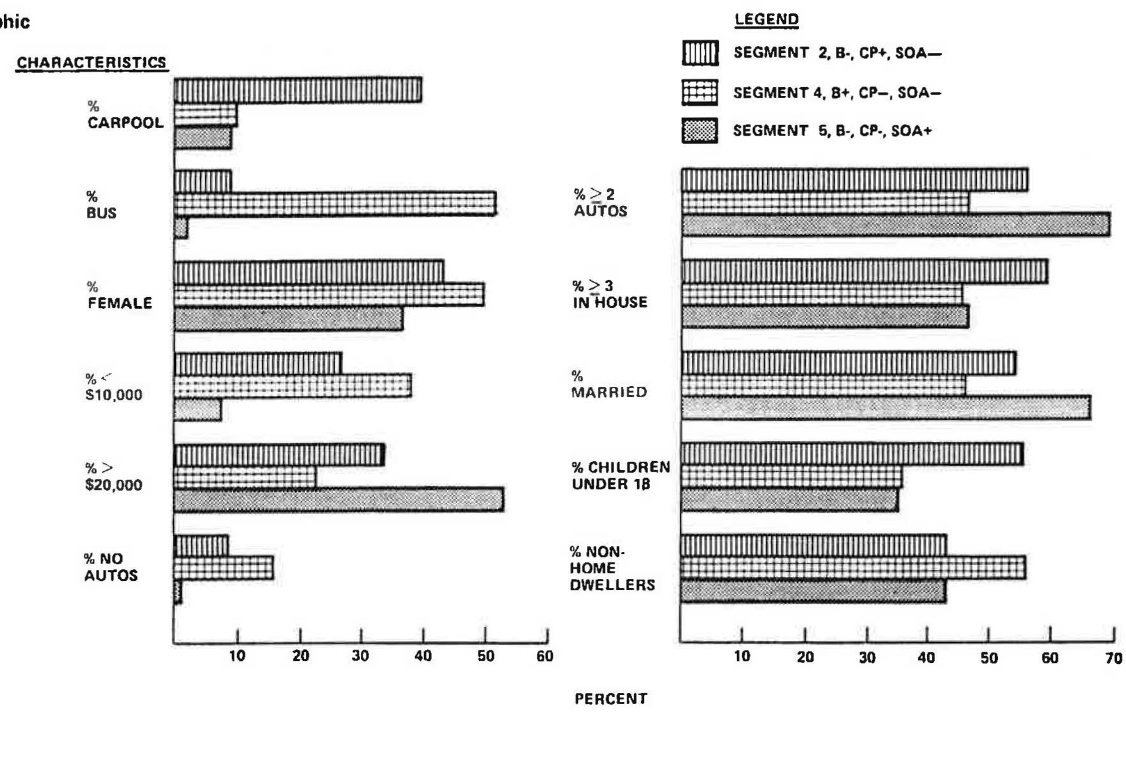



Figure 3. Sociodemographic profiles of selected market segments.



model calibrated in network data through either the main or the interaction terms.

## Split Sample Analysis

The full sample of 874 commuters was randomly divided into two parts: one sample contained 439 travelers, the other 435. This section of the findings compares and contrasts the 439 and 435 samples and the full sample to assess the reliability and robustness of the market segmentation strategy.

Factor analyses were performed on beliefs about each mode for each split half. In general, there is a high degree of comparability despite the fact that the number of retained factors for each mode is not the same for all the samples. The first bus factor was interpreted as convenience across all samples. In addition, the first factors for the carpool and single-occupant auto modes were interpreted, respectively, as convenience, and comfort and convenience. The first table listed specific attributes for the generic factor labels mentioned above. The bus convenience factor is defined in an almost identical fashion for all three samples. There are six attributes used by all three modes. There is more heterogeneity for the carpool convenience factor, but three attributes are used by all three modes. The single-occupant auto factor is also more heterogeneous across samples than the bus convenience factor.

With the composite variable definitions, separate market segmentation structures were derived for the 435 and 439 samples in the manner described by the text discussing Figure 1. The segmentation structure is similar across split samples with respect to the 19 attribute ratings. The following table shows three intercorrelation matrixes among three segments for the nineteen attributes across split samples. These segments are: B+, CP-, SOA-; B-, CP+, SOA-; and B-, CP-, SOA+. There is a separate matrix for each mode.

| | Product-Moment Correlations | | |
|---|---|---|---|
| Beliefs | B-, CP+, SOA- | B+, CP-, SOA- | B-, CP-, SOA+ |
| About carpools | | | |
| B-, CP+, SOA- | 0.64 | 0.61 | 0.50 |
| B+, CP-, SOA- | 0.66 | 0.85 | 0.74 |
| B-, CP-, SOA+ | 0.47 | 0.89 | 0.93 |
| About SOA | | | |
| B-, CP+, SOA- | 0.85 | 0.86 | 0.97 |
| B+, CP-, SOA- | 0.70 | 0.97 | 0.92 |
| B-, CP-, SOA+ | 0.83 | 0.93 | 0.97 |
| About bus | | | |
| B-, CP+, SOA- | 0.93 | 0.71 | 0.97 |
| B+, CP-, SOA- | 0.62 | 0.94 | 0.59 |
| B-, CP-, SOA+ | 0.94 | 0.66 | 0.98 |

Although the split samples have slightly different composite variable definitions and therefore slightly dissimilar segmentation structures, there is a high degree of comparability between the 435 and 439 samples. The diagonal cells would always be larger than other cells from the row and column for which the diagonal cell forms an intersection if a segment correlated most highly with its matching segment in the other sample. This is generally the case, with only six violations out of 36 possible split sample comparisons for the three intercorrelation matrixes. This pattern is statistically significant at beyond the 0.01 level by a sign test.

Because of space limitations, the sociodemographic patterns for the 439 and 435 samples are not shown. The distribution of sociodemographic variables across segments in the two split halves is comparable to that shown for the 874 sample in Figure 3. For example, it is overwhelmingly the case that commuters who favor a mode usually travel by that mode. Another example is that income and auto ownership are negatively associated with positive perceptions of buses and carpools.

Tables 2 and 3 present model calibration and statistical analyses for the perceptual segmentation of the split samples and the engineering data with respect to the three dependent variables. These tables show similar patterns to those reported for the full sample. For example, Table 1 shows that modal beliefs by themselves are consistently better estimators of the dependent variables than the network data. When network data are added in a linear, independent fashion to beliefs, there is no, or at best, only a slight improvement in percentage of variance accounted. However, the addition of network data in a multiplicative fashion to beliefs results in a more substantial increase. Table 2 shows the interaction terms to be statistically significant for every dependent variable in both split samples. These findings support the conclusion from the first two tables that perceptual segment data account for mode choice better than network data and that belief data can add substantially to the explanatory power of a model calibrated in network data either through the main or interaction terms.

Table 4 shows cross-validation squared correlation coefficients for the frequency models calibrated on the split samples. The cross validation was implemented by using the composite variable definitions and model coefficients of one half to estimate the frequency difference between modes in the other half. However, travelers were assigned to perceptual segments based on their composite scores and the distribution of composite scores in the cross-validation sample.

Models based on the perceptual segmentation data

Table 2. Coefficients of determination for mode difference frequency model calibration analysis for split samples.

| | Dependent Variables | | | | | |
|---|---|---|---|---|---|---|
| | (N = 435) | | | (N = 439) | | |
| Independent Variables | A-B | A-CP | B-CP | A-B | A-CP | B-CP |
| NT | 0.03 | 0.01* | 0.01* | 0.03 | 0.00* | 0.03 |
| B | 0.21 | 0.08 | 0.16 | 0.27 | 0.15 | 0.14 |
| NT + B | 0.22 | 0.08 | 0.16 | 0.28 | 0.15 | 0.16 |
| NT + B + NT * B | 0.32 | 0.15 | 0.23 | 0.34 | 0.19 | 0.21 |

*$p > 0.05$.

Table 3. Statistical significance of data types for split sample frequency.

| | Dependent Variables | | | | | |
|---|---|---|---|---|---|---|
| | (N = 435) | | | (N = 439) | | |
| Data Types | A-B | A-CP | B-CP | A-B | A-CP | B-CP |
| NT | $p > 0.05$ | $p > 0.05$ | $p > 0.05$ | $p < 0.05$ | $p > 0.05$ | $p < 0.01$ |
| B | $p < 0.001$ | $p < 0.001$ | $p > 0.05$ | $p < 0.001$ | $p < 0.05$ | $p > 0.05$ |
| NT * B | $p < 0.001$ | $p < 0.001$ | $p < 0.05$ | $p < 0.001$ | $p < 0.05$ | $p < 0.05$ |

Table 4. Cross-validation of frequency models for split samples.

| | Dependent Variables | | | | | |
|---|---|---|---|---|---|---|
| | (N = 435) | | | (N = 439) | | |
| Independent Variables | A-B | A-CP | B-CP | A-B | A-CP | B-CP |
| B | 0.24 | 0.10* | 0.13 | 0.24 | 0.11 | 0.14 |
| B + NT | 0.23 | 0.10 | 0.12 | 0.25 | 0.11 | 0.16 |
| B + NT + NT * B | 0.11 | 0.05 | 0.11 | 0.08 | 0.01[b] | 0.08 |

*$0.01 > p > 0.001$. [b]$p > 0.05$.

alone result in cross-validation squared correlation coefficients that are approximately as large as those from the belief and network data combined in an additive model. In other words, network data do not add to the ability of the perceptual segmentation data to account for the model frequency differences in the other half. Furthermore, both the belief model and the additive belief and network data model are consistently superior to a model that combines multiplicative with additive agglomerations of network and belief data.

## SUMMARY AND CONCLUSIONS

Two principal goals of this report are to illustrate an effective and yet uncomplicated perceptual segmentation paradigm and to compare the perceptual segments resulting from the procedure with engineering indexes of systems performance as determinants of mode choice. Figure 1 presents an outline of important steps in the market segmentation procedure. This presentation is process oriented as opposed to model oriented, and it is, therefore, compatible with the recommendations of Hensher (2, 8) regarding the acceptability of market segmentation to transportation analysts.

While the empirical discussion of the process treats a variety of reliability and validity issues related to the segmentation strategy, its actual implementation involves little more than a factor analysis and the grouping of respondents based on median scores. Furthermore, as noted in the discussion of Figure 1, factor analysis is not a mandatory component of the procedure.

It was empirically demonstrated that this uncomplicated segmentation paradigm yielded perceptual segments that correlated better with mode choice than standard engineering data on systems performance. This finding is a basis for the more widespread use of this segmentation paradigm.

The reliability of the market segmentation procedure highlighted in Figure 1 is assessed in the findings. When the full sample of 874 commuters was split into two samples of 435 and 439, there was a large degree of comparability between the two parts. Furthermore, the superiority of the perceptual data relative to the engineering data was clearly evidenced in both split samples, and this superiority was also demonstrated in a cross-validation test.

The perceptual segmentation paradigm described here can be used for transportation planning in a two-step process. The first step involves the calibration of a market segmentation structure. As mentioned above, this step involves little more than an identification of important attributes and grouping respondents based on median composite scores, and its principal elements are described in the discussion of Figure 1. The resulting segmentation structure will isolate travelers with common viewpoints about alternative transport modes. In addition, the criteria used to identify who belongs to particular segments will be readily available. These criteria will comprise selected semantic differential scales that are the basis for composite variables for alternative transport modes. Example scale designations are listed in Table 1.

In the second step, the calibrated market segmentation structure is applied to respondents in a different geographic region or for some planning horizon associated with a future date. If it is possible to survey the sample to which the segmentation structure is to be applied, then a reduced questionnaire form based on the composite variable definitions can be used. This will save interviewing time and simplify the interviewing instrument. In any event, no factor analysis is required in the second step because the important attributes will have been identified in the first phase. Where reinterviewing is impossible, estimates of the distribution of travelers in the distinct perceptual segments must be based on the best available alternative data sources.

When a perceptual segmentation procedure is implemented as described above, several useful consequences follow. For example, the factor analysis step can be used to identify and define salient generalized attribute variables, such as those discussed by Spear (16) or Nicolaidis (17). Furthermore, these definitions serve as the bases for the resulting perceptual segmentation structure. The specific attributes that underlie the segmentation structure are likely to be very important with respect to consumer choice, and they thus deserve high-level consideration in the setting of policies with respect to system operation, new service planning, and innovative transport facilities.

Within the calibration step of the segmentation process, it is possible to examine the correlation of the perceptual segments to various behavioral and sociodemographic indexes. This correlation element will aid in the validation of the segmentation structure. For example, those with a positive viewpoint toward a mode should use that mode more than those without. By combining the perceptual and sociodemographic information, it is possible to specify different messages for alternative segments of the traveling public with different sociodemographic characteristics. By uncovering the linkages between sociodemographic characteristics and perceptual segments, it is possible to identify target groups for directing information.

The significance of the study reported herein can be judged by the new knowledge it contributes to the body of market segmentation research on transportation planning and the number of new applications and analyses generated by it. The current investigation is itself a synthesis and extension of earlier analyses and discussions presented by Golob and Dobson (5), Hensher (2, 8), and Costantino, Dobson, and Canty (6). The fundamental behavioral tenet around which this report centers is that travelers have different viewpoints of transportation facilities and services and that these viewpoints have strong implications for the selection of modal alternatives. The segmentation procedure applied here was shown to be highly correlated with mode choice, especially in comparisons with network data. This correlation supports the fundamental tenet of the paper and supports the diffusion of perceptual segmentation paradigms.

## REFERENCES


1. R. Dobson. Market Segmentation: A Tool for Transportation Decisionmaking. Proc., 3rd International Conference on Behavioral Travel Modelling (in press).
2. D. A. Hensher. Use and Application of Market Segmentation. In Behavioral Travel-Demand Models (P. R. Stopher and A. H. Meyburg, eds.), Heath, Lexington, MA, 1976, pp. 271-282.
3. J. L. Louviere, L. M. Ostresh, D. H. Hensley, and R. J. Meyer. Travel Demand Segmentation: Some Theoretical Considerations Related to Behavioral Modelling. In Behavioral Travel-Demand Models (P. R. Stopher and A. H. Meyburg, eds.), Heath, Lexington, MA, 1976, pp. 259-270.
4. R. R. Reed. Market Segmentation for Public Transportation. Dept. of Industrial Engineering, Stanford Univ., Stanford, CA, 1973.
5. T. F. Golob and R. Dobson. The Assessment of

Preferences and Perceptions Toward Attributes of Transportation Alternatives. TRB, Special Rept. 149, 1974, pp. 58-81.
6. D. P. Costantino, R. Dobson, and E. T. Canty. An Investigation of Modal Choice for Dual Mode Transit, People Mover, and Personal Rapid Transit Systems. General Motors Research Laboratories, Warren, MI, GMR 1587, 1974; abstract available in TRB, Special Rept. 170, 1976, p. 67.
7. R. Dobson and J. F. Kehoe. Disaggregated Behavioral Views of Transportation Attributes. TRB, Transportation Research Record 527, 1974, pp. 1-15.
8. D. A. Hensher. Market Segmentation as a Mechanism in Allowing for Variability of Traveller Behavior. Transportation, Vol. 5, 1976, pp. 257-284.
9. W. W. Recker and T. F. Golob. A Behavioral Travel Demand Model Incorporating Choice Constraints. Proc., Association for Consumer Research Meeting, Vol. 3 (in press).
10. R. Dobson and M. L. Tischer. Beliefs About Buses, Carpools, and Single-Occupant Autos: A Market Segmentation Approach. Proc., Transportation Research Forum, Vol. 17, 1976, pp. 200-209.
11. R. Dobson and M. L. Tischer. A Comparative Analysis of Determinants for Central Business District Worker Mode Choices. TRB, Transportation Research Record 649, 1977, pp. 7-14.
12. Pricing—Transit Marketing Management. Urban Mass Transportation Administration, 1976.
13. R. Dobson. The General Linear Model Analysis of Variance: Its Relevance to Transportation Planning and Research. Socio-Economic Planning Sciences, Vol. 10, 1976, pp. 231-235.
14. A. D. Horowitz and J. N. Sheth. Ride Sharing to Work: An Attitudinal Analysis. TRB, Transportation Research Record 637, 1977, pp. 1-8.
15. M. L. Tischer and R. Dobson. An Empirical Analysis of Behavioral Intentions to Shift Ways of Traveling to Work. Paper presented at the 56th Annual Meeting, TRB, 1977.
16. B. Spear. Generalized Attribute Variable for Models of Mode Choice Behavior. TRB, Transportation Research Record 592, 1976, pp. 6-11.
17. G. C. Nicolaidis. Quantification of the Comfort Variable. Transportation Research, Vol. 9, 1975, pp. 55-66.

*Publication of this paper sponsored by Committee on Passenger Travel Demand Forecasting and Committee on Traveler Behavior and Values.*

**Mr. Dobson was at the Federal Highway Administration when this research was completed, but revisions to the text were incorporated while he was with Charles River Associates, Inc.*

# Testing for Significant Induced Trip Making and Travel in Providence, Rhode Island

Michael E. Smith and George E. Schoener, Federal Highway Administration, U.S. Department of Transportation

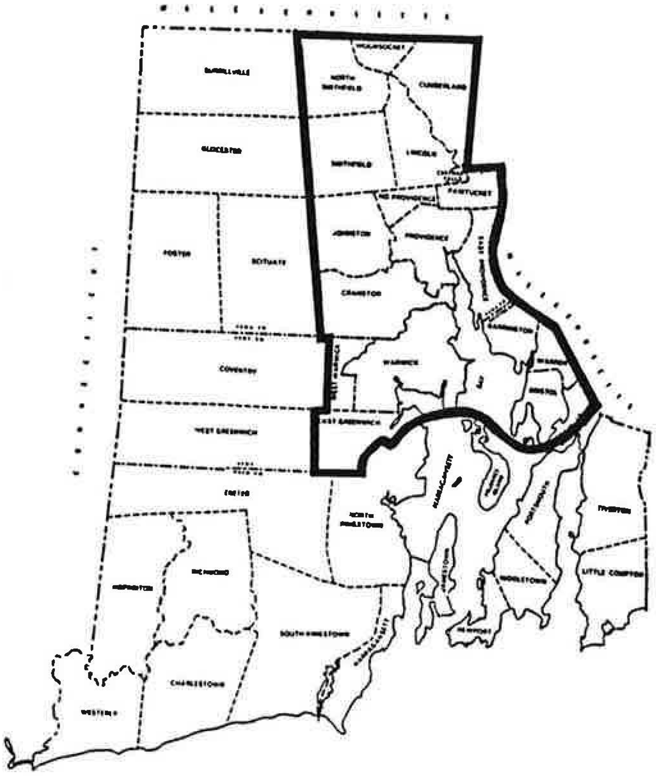
**The research reported in this paper was conducted in order to determine whether or not increased highway supply causes increased travel or trip making or both. In order to make this determination, data from origin-destination travel surveys conducted by the Rhode Island Department of Transportation in Providence for the years 1961 (before construction of I-95) and 1971 (after I-95) were used. For each year the origin-destination survey data from the Providence area were divided into two groups—samples representing households inside the I-95 corridor and samples representing those outside it. For the resulting four groups of households, cross-classification matrixes were developed using household size and auto ownership as independent variables; the dependent variables were vehicle-kilometers of travel (VKMT) per household, vehicle-hours of travel (VHT) per household, and auto driver trips per household. The comparison of the resulting matrixes revealed that the highway did not increase trips or VHT, but it did increase VKMT. This allows the tentative conclusion that travelers increase their VKMT until they use up a given amount of travel time. This conclusion supports the standard system-insensitive approach to trip generation as well as the use of travel time as an impedance in trip distribution.**

A frequent statement advanced by transportation professionals is that highway improvements, by inducing travel, create more congestion than they eliminate. Although few data exist to support this statement, it has gained legitimacy by sheer repetition. Another frequent observation is that new highway facilities imply an increase in vehicle-kilometers of travel (VKMT). The legitimacy of this is crucial to the evaluation of the energy and air quality impacts of new facilities. The purpose of this paper is to provide some empirical evidence to either validate or invalidate these two observations.

The data we shall use were gathered during two household surveys—one in 1961 before construction of I-95 and the other in 1971 after construction—in Providence, Rhode Island (Figure 1). The two origin-destination (O-D) travel surveys conducted in Providence were of the home interview type. The 1961 survey consisted of 11 467 samples, while the 1971 survey contained 855 samples.

A naive way to determine whether or not a highway induces trip making and travel (TM/T) is to measure TM/T before and after a highway is built. If the "after" TM/T is greater, then one might conclude that the highway did indeed induce it. The problem with this approach is that the increase in TM/T may be due to changes other than those on the new highway facility. For example, between the two surveys incomes probably rose, auto availability rose, and land-use changes occurred. Any

Figure 1. Providence metropolitan area.



of these changes could cause increased TM/T, regardless of whether or not the highway had been constructed.

In this paper such changes are controlled for in two ways. First, the study area is divided into two parts—the portion inside the influence of the new highway and the portion outside it. Determining the influence area will be discussed later. Second, trip-making behavior is studied by dividing the population into different groups, based on family size and auto ownership. The trip rates of these groups rather than those of the population as a whole are analyzed.

Many previous studies have shown that a correlation exists between aggregate highway supply per capita and VKMT per capita (1, 2, 3, 4). The existence of such a correlation, however, does not guarantee the existence of a causal relationship between the two variables.

Another study design used in the past and similar to the one used here is the before-and-after variety. The main objection to this type of study is the presence of confounding variables (5, 6), which we have attempted to eliminate in this research by establishing a control group. This technique has also been criticized because it is impossible to specify a control group of identical, or even nearly identical persons who are completely unaffected by the change (5). Because of this, we analyzed differences between the control and survey groups to see if these differences remained constant.

Although the control group was admittedly not completely immune to effects from the new highway, they were less affected by it. Therefore, if a change in the differences between the groups occurs, it can be concluded that the highway did change travel behavior. However, because the line between the two groups is fuzzy at best, it will be impossible to detect the structure or magnitude of the change. This research, then, began by being confined only to finding out whether or not a highway itself (directly or indirectly) generates a significant number of trips or is responsible for a significant increase in VKMT or vehicle hours of travel (VHT). Certain relationships emerged from the findings of this research that allowed some tentative conclusions about the magnitude and structure of the change, but most of the discussion is oriented toward TM/T generation.

Other research has been conducted on the premise that new traffic on a highway can be divided into different categories such as developmental (from land-use changes), natural growth (from socioeconomic changes), diverted (from other streets or highways), induced (new trips made because of the highway), transferred (from other modes), and shifted (to new destinations) (7).

The methodology of this research should control for diverted and natural growth traffic because the dependent variables are analyzed at the household level, with the households divided into homogeneous groups. Induced, transferred, and shifted traffic are to be considered highway generated and will all be captured in this study. Developmental traffic will also be discussed in this study, but, since there is a control group, only those trips attracted to highway-induced development will be measured. Again, we did not set out to measure the magnitude of the change, only to decide whether or not a significant change occurred.

## METHODOLOGY

To determine whether the presence of the highway caused additional TM/T in the Providence metropolitan area (PMA), a two-part procedure was applied. First, the PMA was divided into two parts: first the influence area, defined as those zones influenced by the presence of the highway, and second the rest of the area. The influence area is referred to as inside the corridor, or, simply, inside. The rest of the area will be referred to as outside the corridor, or outside. The household and trip data from the 1961 and 1971 surveys were divided into two groups—one for inside the corridor and the other for outside the corridor—in order to develop comparative data.

### Definition of the Corridor

In order to find which zones were inside as opposed to outside, a selected link analysis program, LINKUSE, was run with a 1971 calibrated network and 1971 O-D survey information. Given any origin zone, LINKUSE will compute the number of trips originating in that zone that use a given set of links. By coding I-95 as the given set of links, it was possible to identify all zones containing trip origins whose paths followed I-95 for some part of the trip.

Each of the zones so identified was marked on a 1971 PMA traffic zone map. Most of these zones were within 3 or 4 km (1.5 or 2 miles) of the highway. Zones farther out were seldom marked. Therefore, the marked zones provided a rough, or "first-cut" corridor. In order to identify a continuous corridor, "holes" were filled in; that is, an unmarked zone, when surrounded by marked zones, was included in the corridor. Conversely, a marked zone surrounded by unmarked zones was excluded from the corridor. The first-cut corridor, plus these adjustments, was the corridor used for this study (Figure 2).

To determine which traffic zones in the 1961 survey were within the corridor, a traffic-zone equivalency table was developed. Thus, the 1971 traffic zones that were identified as being within the corridor were converted to equivalent 1961 traffic zones. Similarly, this conversion was performed for traffic zones outside the corridor. This procedure made it possible to identify four data sets for analysis purposes.

## Development of Comparative Data

One way to estimate the effect of the highway on trip making would be to compute the overall trip rates for persons inside the corridor and for persons outside the corridor for both 1961 and 1971. If the trip rates increased for persons inside the corridor more than it increased for persons outside it, one might conclude that the highway actually did induce trips. This approach was rejected for the following two reasons: first, variables other than the presence of the highway that are correlated with location may be the cause of trip rate changes, and, second, by taking into account the effect of other variables through the use of a standard trip generation model, the model's stability can be tested.

A cross-classification model was chosen as the best way to represent trip generation behavior. Although cross-classification by auto ownership and income is usually recommended (8), income was not available in the Providence survey. Therefore, cross-classification by auto ownership and family size was used instead. Family size has been used successfully in other trip generation studies.

Because we were concerned with vehicle trips and travel only, we deleted from the data set all trips that were not auto driver trips. The auto driver trips were then divided into the following four data sets:

1. 1961 IN—all 1961 auto driver trip records with zones of residence inside the corridor,
2. 1961 OUT—all 1961 auto driver trip records with zones of residence outside the corridor,
3. 1971 IN—same as 1961 IN but with 1971 data, and
4. 1971 OUT—same as 1961 OUT but with 1971 data.

We amended each data set by adding trip length and trip time to each trip record. The trip time was obtained by locating the minimum time path from the given origin to the given destination on the calibrated network. The trip distance was defined as the road kilometers along the minimum time path. This provides a means for measuring VKMT as well as trip rates.

The next step was to cross-classify the data in each of the four data sets. Table 1 shows the number of observations in each cell as a result of this cross-classification. In addition, each dependent variable (VKMT per household, VHT per household, and vehicle trips per household) was cross-classified by auto ownership and family size. The results are shown in Tables 2, 3, and 4. It should be noted that there are a rela-

Figure 2. I-95 corridor in Providence metropolitan area.

Table 1. Number of observations in each family size and auto ownership category.

| Data | Number of Autos | Number of People 1 | 2-3 | 4+ |
|---|---|---|---|---|
| 1961 IN | 0 | 637 | 792 | 188 |
| | 1 | 294 | 2404 | 1865 |
| | 2+ | 9 | 526 | 799 |
| 1961 OUT | 0 | 248 | 295 | 110 |
| | 1 | 98 | 1263 | 1087 |
| | 2+ | 6 | 358 | 504 |
| 1971 IN | 0 | 57 | 49 | 6 |
| | 1 | 41 | 153 | 45 |
| | 2+ | 2 | 116 | 100 |
| 1971 OUT | 0 | 20 | 9 | 4 |
| | 1 | 10 | 57 | 25 |
| | 2+ | 2 | 61 | 66 |

Table 2. Cross-classification of auto driver trips per household by family size and auto ownership.

| Data | Number of Autos | Number of People 1 | 2-3 | 4+ |
|---|---|---|---|---|
| 1961 IN | 0 | 0.013 | 0.078 | 0.293 |
| | 1 | 2.656 | 3.090 | 3.583 |
| | 2+ | - | 5.055 | 5.885 |
| 1961 OUT | 0 | 0.012 | 0.075 | 0.300 |
| | 1 | 2.663 | 3.193 | 3.626 |
| | 2+ | - | 4.687 | 5.591 |
| 1971 IN | 0 | 0.0 | 0.041 | 0.0 |
| | 1 | 3.171 | 3.778 | 4.511 |
| | 2+ | - | 6.431 | 8.060 |
| 1971 OUT | 0 | 0.0 | 0.0 | 0.0 |
| | 1 | 4.000 | 3.421 | 4.280 |
| | 2+ | - | 5.361 | 7.500 |

Table 3. Cross-classification of VKMT per household in tenths of kilometers by family size and auto ownership.

| Data | Number of Autos | Number of People 1 | 2-3 | 4+ |
|---|---|---|---|---|
| 1961 IN | 0 | 0.6 | 6.1 | 20.9 |
| | 1 | 146.3 | 186.2 | 217.7 |
| | 2+ | - | 333.9 | 391.4 |
| 1961 OUT | 0 | 1.4 | 10.0 | 26.2 |
| | 1 | 165.3 | 220.3 | 254.6 |
| | 2+ | - | 387.8 | 444.2 |
| 1971 IN | 0 | 0.0 | 2.1 | 0.0 |
| | 1 | 196.2 | 267.1 | 218.1 |
| | 2+ | - | 510.2 | 610.4 |
| 1971 OUT | 0 | 0.0 | 0.0 | 0.0 |
| | 1 | 192.0 | 213.4 | 225.8 |
| | 2+ | - | 435.0 | 573.6 |

Table 4. Cross-classification of VHT per household in minutes by family size and auto ownership.

| Data | Number of Autos | Number of People 1 | 2-3 | 4+ |
|---|---|---|---|---|
| 1961 IN | 0 | 0.11 | 0.89 | 2.96 |
| | 1 | 23.47 | 29.62 | 34.39 |
| | 2+ | - | 51.23 | 60.74 |
| 1961 OUT | 0 | 0.18 | 1.29 | 3.63 |
| | 1 | 23.94 | 30.58 | 35.51 |
| | 2+ | - | 52.65 | 60.71 |
| 1971 IN | 0 | 0.0 | 0.33 | 0.0 |
| | 1 | 27.71 | 36.07 | 32.64 |
| | 2+ | - | 67.83 | 80.35 |
| 1971 OUT | 0 | 0.0 | 0.0 | 0.0 |
| | 1 | 28.20 | 29.07 | 30.52 |
| | 2+ | - | 59.48 | 76.83 |

Table 5. Test for differences between cell means for trips per household.

| Data | Number of Autos | Number of People 1 | 2-3 | 4+ |
|---|---|---|---|---|
| 1961 IN versus 1971 IN | 0 | 0.642 | 0.425 | 0.715 |
| | | 303 | 342 | 114 |
| | 1 | -0.979 | -2.85 | -2.27 |
| | | 137 | 1414 | 1130 |
| | 2+ | - | -5.06 | -5.78 |
| | | - | 472* | 602* |
| 1961 OUT versus 1971 OUT | 0 | 0.346 | 0.440 | 0.414 |
| | | 655 | 799 | 190 |
| | 1 | -1.67 | -1.11 | -1.43 |
| | | 302 | 2459 | 1888 |
| | 2+ | - | -0.777 | -3.172 |
| | | - | 585 | 863* |
| 1961 IN versus 1961 OUT | 0 | -0.083 | -0.062 | 0.046 |
| | | 883 | 1085 | 296 |
| | 1 | 0.025 | 1.347 | 0.463 |
| | | 390 | 3665 | 2950 |
| | 2+ | - | -1.931 | -1.373 |
| | | - | 882* | 1301 |
| 1971 IN versus 1971 OUT | 0 | 0.0 | 0.432 | 0.0 |
| | | 75 | 56 | 8 |
| | 1 | -0.630 | 0.668 | 0.242 |
| | | 49 | 208 | 68 |
| | 2+ | - | 1.621 | 0.684 |
| | | - | 175 | 164 |

*Significant at the 10 percent level.

Table 7. Test for differences between VHT per household among the four data sets.

| Data | Number of Autos | Number of People 1 | 2-3 | 4+ |
|---|---|---|---|---|
| 1961 IN versus 1961 OUT | 0 | -0.526 | -0.661 | -0.353 |
| | | 883 | 1085 | 296 |
| | 1 | -0.153 | -0.973 | -0.918 |
| | | 390 | 3665 | 2950 |
| | 2+ | -. | -0.520 | +0.011 |
| | | - | 882 | 1301 |
| 1961 OUT versus 1971 OUT | 0 | +0.342 | 0.345 | 0.525 |
| | | 266 | 302 | 112 |
| | 1 | 0.372 | 0.372 | 0.729 |
| | | 106 | 1318 | 1110 |
| | 2+ | - | -1.172 | -2.720 |
| | | - | 417 | 568* |
| 1961 IN versus 1971 IN | 0 | -0.5533 | -0.499 | -0.429 |
| | | 692 | 839 | 192 |
| | 1 | 0.927 | 2.740 | -0.378 |
| | | 333 | 2555* | 1908 |
| | 2+ | - | 3.894 | 3.585 |
| | | - | 640* | 897* |
| 1971 IN versus 1971 OUT | 0 | 0.0 | 0.4349 | 0.0 |
| | | 75 | 56 | 8 |
| | 1 | -0.0319 | 1.258 | 0.317 |
| | | 49 | 208 | 68 |
| | 2+ | - | 1.052 | 0.332 |
| | | - | 175 | 164 |

*Significant at the 10 percent level.

Table 6. Test for differences between VKMT per household among the four data sets.

| Data | Number of Autos | Number of People 1 | 2-3 | 4+ |
|---|---|---|---|---|
| 1961 IN versus 1971 IN | 0 | 0.569 | 0.474 | 0.390 |
| | | 692 | 839 | 192 |
| | 1 | -1.464 | -4.52 | -0.009 |
| | | 333 | 2555* | 1908 |
| | 2+ | - | -5.32 | -5.20 |
| | | - | 640* | 897* |
| 1961 OUT versus 1971 OUT | 0 | 0.346 | 0.315 | 0.509 |
| | | 266 | 302 | 112 |
| | 1 | -0.404 | 0.203 | 0.470 |
| | | 106 | 1318 | 110 |
| | 2+ | - | 1.001 | -2.719 |
| | | - | 417 | 568* |
| 1961 IN versus 1961 OUT | 0 | -0.878 | -0.801 | -0.364 |
| | | 883 | 1085 | 296 |
| | 1 | -0.850 | -4.386 | -1.421 |
| | | 390 | 3665* | 1950 |
| | 2+ | - | -2.508 | -2.606 |
| | | - | 882* | 1301* |
| 1971 IN versus 1971 OUT | 0 | 0.0 | 0.430 | 0.0 |
| | | 75 | 56 | 8 |
| | 1 | 0.426 | 1.171 | 0.151 |
| | | 49 | 208 | 68 |
| | 2+ | - | 1.182 | 0.416 |
| | | - | 175 | 164 |

*Significant at the 10 percent level.

tively small number of cells in each table. This is because of the small number of samples obtained in the 1971 survey, which made it impossible to construct a large table with stable cell means. Also, the one-person, multicar household cell was deleted due to lack of data.

## ANALYSIS

In order to decide whether or not the highway had generated trips and to determine model stability, several t-statistics were constructed to test the significance of differences between the cell means in each of the four matrixes shown in Table 2. The following pairs of matrixes were compared: 1961 IN versus 1971 IN, 1961 OUT versus 1971 OUT, 1961 IN versus 1961 OUT, and 1971 IN versus 1971 OUT. The t-statistics, along with the degrees of freedom (which are, here, the numbers of observations) for each one, are shown in Tables 5, 6, and 7.

Because we are trying to compare matrixes, t-statistics on the individual elements of each matrix are of little value in isolation. What is needed is a way to evaluate the significance of a matrix of t-values. This was done in the following way. The probability of a t-statistic's being significant at the 10 percent level by pure chance is 0.10. Therefore, the chance probability that one t-value in a set of eight will be significant is computed as follows:

$$p(1) = [8!/1!\,(7!)]\ (0.9)^7\ (0.1)^1 = 0.383 \tag{1}$$

Other probabilities are computed as follows:

$$p(2) = [8!/2!\,(6!)]\ (0.9)^6\ (0.1)^2 = 0.149 \tag{2}$$

$$p(3) = [8!/3!\,(5!)]\ (0.9)^5\ (0.1)^3 = 0.033 \tag{3}$$

$$p(n > 3) = \sum_{i=3}^{8} p(i) = 1 - \sum_{i=0}^{2} p(i) = 0.048 \tag{4}$$

We had previously decided to reject the null hypothesis (i.e., that the two matrixes are equivalent) when a significance level of 10 percent was reached. Therefore, this probability analysis shows that, when comparing any two matrixes, up to two significant t-statistics may be generated and we will still fail to reject the null hypothesis that the matrixes are equivalent. If, on the other hand, three or more significant t's are found in the comparison matrix, then we must conclude that the matrixes being compared are significantly different.

Using this criterion, the relationships between the matrixes were developed. These relationships are shown schematically in Figures 3, 4, and 5. Figure 3 shows that the entries in the trips per household matrix for the 1961 IN data set are significantly smaller than

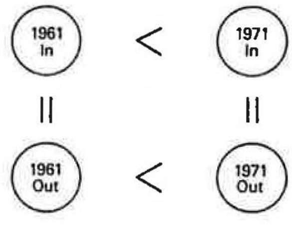


**Figure 3. Schematic of trip rate differences.**

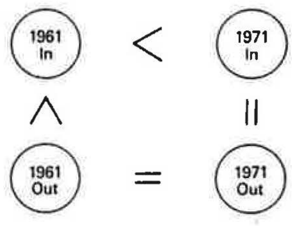


**Figure 4. Schematic of VKMT per household differences.**

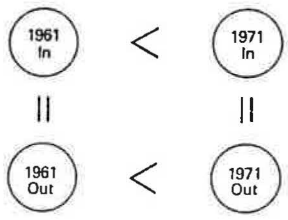


**Figure 5. Schematic of VHT per household differences.**

corresponding entries in the 1971 IN data set. The same relationship holds between 1961 OUT and 1971 OUT. However, there was no significant difference in trip rates for those inside the corridor as opposed to those outside for either year. Therefore, it appears that, although trip rates did increase significantly, the highway was not the cause.

Figure 4 shows that, in 1961, there was significantly less VKMT per household inside the corridor than outside it. Over the years this difference disappeared. Apparently people living close to the highway are now making longer trips, though they have not increased their trip rate.

Figure 5 shows how VHT per household varies temporally and spatially. The pattern is precisely the same as that in Figure 3. Therefore, one could conclude that the highway did not increase the daily vehicle hours of travel per household.

Although Table 7 shows only one significant difference between 1961 OUT and 1971 OUT, there were enough data in these two data sets to show a significant difference between the two one-person, multicar household cells. We had previously decided to reject the null hypothesis only when three or more significant differences had been detected. However, rejecting the null hypothesis in this case preserves transitivity as indicated in Figure 5. Not rejecting the null hypothesis would violate this transitivity.

## CONCLUSIONS AND DISCUSSION

The first conclusion reached was that, according to limited data collected, the highway did not generate new trips. This conclusion is based on the comparisons schematized in Figure 3. Since the trip rate matrixes for the two geographic areas (inside and outside the corridor) were equal at both points in time, the highway could not have been responsible for any increase in trip rates.

The results shown in Figure 3 also permit another conclusion. Since the trip rates increased over time, it must be concluded that the model form is temporally unstable. It must be noted, however, that family size is not a preferred variable for the cross-classification. A model using income may have proved to be more stable. Also, the number of cells in the matrixes is far less than ideal. This is because of the lack of data. An expansion of each matrix would have thinned out the data to the point where there may have been no data in some of the cells.

Although analysis of Figure 3 leads to the conclusion that the highway had generated no new trips, Figure 4 shows that the highway did create a demand for longer trips (in terms of kilometers). Although the data in this case are insufficient to measure the amount of extra VKMT generated by the highway to any practical level of precision, the data are sufficient to show that the increase is statistically significant. This demand for longer trips is indicated by the fact that in 1961 households inside the corridor traveled fewer kilometers than similar households outside the corridor. By contrast, in 1971 the households inside the corridor were traveling just as many kilometers as those outside the corridor. The fact that VKMT per household rates did not change over time in the outside group indicates that when there are no major transportation system changes the cross-classification model for predicting VKMT may be temporally stable. More data may indicate otherwise, however.

As indicated by Figure 5, we cannot reject the null hypothesis that the highway did not generate extra vehicle hours of travel or reduce them. If we accept the null hypothesis previously not rejected, then a very interesting conclusion can be drawn: The highway led to extra VKMT only to the extent that a trip could be completed in a given amount of time. For example, a 10-min shopping trip remains 10 min; however, it is now a 12-km (7-mile) trip rather than an 8-km (5-mile) trip.

To analyze the effect of this concept we used the following variables: h = VHT per household, m = VKMT per household, and s = average travel speed. Obviously, $m/h = s$, or $m = sh$. Therefore, a 2 percent increase in travel speed would result in a 2 percent increase in VKMT. This result suggests a procedure for predicting VKMT for a given transportation network.

First, devise a cross-classification model that directly predicts VKMT per household. Any change in the transportation system that increases speed by a certain factor should be dealt with by increasing the predicted VKMT by the same factor.

If this is indeed a true model of travel behavior, then transportation improvements that significantly increase travel speed will result in a proportional increase in VKMT. However, improvements that are designed mostly to smooth out traffic flow but do not significantly increase travel speed will result in no significant increase in VKMT.

Another conclusion that may be gleaned from this research is relevant to the transportation planning process as it currently exists. Most travel demand models predict trips produced independently of the characteristics of the transportation system. This research supports such an approach. Since, as far as the data show, a radical improvement in the transportation system did not significantly increase trip making, it appears reasonable to assume that trip generation is indeed independent of the transportation system.

This research also supports the use of travel time as a measure of spatial separation in the gravity model. A decrease in travel time, when put into the standard gravity model, will result in increased trip lengths and therefore increased VKMT. The fact that people affected by the new transportation facility in Providence may have increased their VKMT directly in response to increased speed shows that travel time may be the major impediment to travel. This result further supports the use of travel time as the measure of spatial separation in distribution models such as the gravity model and the intervening opportunities model.

Although the analyses performed in this study indicate that new highways result in more VKMT with no increase in VHT or the number of trips, the data were insufficient to measure the amount of change in the VKMT. In addition, the paucity of the data makes small changes difficult to detect. There may have been small changes in VHT and trips that escaped detection in this study. Therefore, the conclusion that VKMT increases by the same amount as average speed increases must be made with appropriate reservations.

Further research should address the question of how many extra VKMT are produced by new highways as well as the relationship to VHT, number of trips, average trip length, and average speed. Such a quantification would provide a major breakthrough in the field of highway planning. In addition, this analysis was limited to the comparison of residential trip productions; further research should investigate the effect of system supply changes on nonresidential trip generation (i.e., trip attractions).

ACKNOWLEDGMENTS

We appreciate the assistance of Roland Frappier of the Rhode Island Statewide Planning Program in providing much of the data used in the analysis.

REFERENCES

1. A System Sensitive Approach for Forecasting Urban Area Travel Demands. Alan M. Voorhees and Associates, McLean, VA, Rept. AMV-18-524; Federal Highway Administration, 1971.
2. F. Koppelman. A Model for Highway Needs Evaluation. HRB, Highway Research Record 314, 1970, pp. 123-134.
3. F. Koppelman and I. Shalkowitz. Allocation of Resources for Construction of Tri-State Regional Highways. HRB, Highway Research Record 399, 1972, pp. 51-61.
4. S. Berwager and G. Wickstrom. Estimating Auto Emissions of Alternative Transportation Systems. Metropolitan Washington Council of Governments; U.S. Department of Transportation, 1972.
5. Cambridge Systematics and JHK and Associates. The Relationship of Changes in Urban Highway Supply to Vehicle-Miles of Travel. U.S. Department of Transportation, Federal Highway Administration, NCHRP Project 8-19, 1977.
6. M. Ben-Akiva. Measurement of Traveler Response to Changes in Transportation System Supply. In Design of Procedures to Evaluate Traveler Responses to Changes in Transportation System Supply, Conference Summary and White Papers. NTIS, Springfield, VA, 1974, PB-240003/PDD.
7. S. Zimmerman, M. West, and T. Kozlowski. Urban Highways as Traffic Generators. U.S. Department of Transportation, Federal Highway Administration, 1974.
8. Trip Generation Analysis. U.S. Department of Transportation, Federal Highway Administration, 1975.

*Publication of this paper sponsored by Committee on Passenger Travel Demand Forecasting and Committee on Travel Behavior and Values.*

# Destination Choice Behavior for Non-Grocery-Shopping Trips

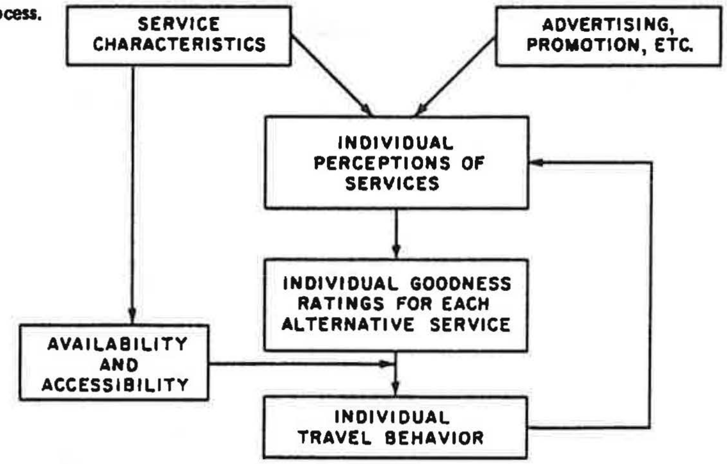
Frank S. Koppelman, Department of Civil Engineering, and John R. Hauser, Department of Marketing, Transportation Center, Northwestern University, Evanston, Illinois

**This paper combines attitude and market research and disaggregate behavioral demand modeling to produce a diagnostic and predictive model of destination choice for non-grocery-shopping trips. The analysis is based on perception and preference models to measure attractiveness and logit choice models to link attractiveness and accessibility to frequency of destination choice. Alternative analytic techniques were compared to identify the most effective technique for each step in the process. Factor analysis was found to be superior to nonmetric scaling to identify consumer perceptions of shopping location attractiveness because it is more understandable and predicts better. Statistical preference models (first preference logit, preference regression) provided consistent predictions and similar interpretations. For choice prediction, revealed preference (standard logit approach) and intermediate preference models provided complementary insight into the consumer behavior process. Use of both models leads to insights that would have remained hidden had either model been used alone. The results indicated that attractiveness of trip destination can be effectively measured with attitudinal models; that the five basic (measured) constructs of attractiveness are variety, quality, satisfaction, value, and parking; that of these quality is consistently the most important and prestige of store appears to be the most important aspect of quality; and that both attractiveness and accessibility are important determinants of destination choice. Any destination choice model should include both.**

The focus of this study is on the trip maker's choice of destination for non-grocery-shopping trips. This research was undertaken in the belief that improved understanding of this destination choice process would provide insight into the general process of destination choice behavior. This research also develops effective analytic models that can be used for the analysis of destinations other than shopping areas.

Travel choice behavior can be represented by a simple

Figure 1. Consumer response process.




evaluation and selection process. Each individual evaluates each alternative that is known and available to him or her and chooses the alternative he or she values most highly. Because the value of an alternative to an individual cannot be precisely specified, the choice process is represented by a probabilistic choice model in terms of those aspects of value that can be identified. That is, based on a partial valuation of each alternative, the model predicts the probability that the individual will select each of the available alternatives. The individual probabilities can be aggregated across individuals to provide predictions of group behavior. The structure of the consumer response process used in this study extends the approach described above in two significant ways.

First, the characteristics of the alternatives are described by what the individual perceives rather than by engineering measures. This approach extends the range of attributes to include those that cannot be measured by direct engineering means; it accounts for differences of individual perceptions of identical alternatives; and it gives useful insight into how consumers actually perceive alternatives.

Second, the substantive aspects of the destinations—their attractiveness—are modeled separately from the spatial aspects of these alternatives—their accessibility. This approach follows the research direction suggested earlier by Hanson (1).

The resulting model structure consists of three integrated components that describe individual perceptions of shopping locations, individual evaluations of shopping location attractiveness based on relative preferences for perceived characteristics, and choice of shopping location based on attractiveness ratings and accessibility measures. This model is based on a consumer behavior model formulated by Hauser and Urban (2) and independently by Shocker and Srinivasan (3) and modified for transportation by Hauser and Koppelman (4).

## OBJECTIVES OF THE RESEARCH AND APPROACH

The research had two objectives. The first was to increase understanding of the process by which individual consumers select locations for non-grocery-shopping trips. The second was to develop and critically evaluate alternative empirical models of the consumer choice process.

These objectives were achieved by developing and interpreting alternative models of perception and preference and integrating them with a choice model. The alternative models provide different perspectives on the consumer process and contribute to an overall understanding of that process. Comparing models provides a basis for selecting those that will be most useful in particular situations. The primary criteria are their ability to provide useful insights into consumer behavior and to predict accurately consumer preferences and choice behavior.

The model structures examined include three models of perceptions and three models of consumer preference combined with the multinomial logit choice model. The models of perception include fundamental attributes, factor analysis, and nonmetric scaling. Fundamental attributes represent perceptions in terms of an extensive list of attributes. Factor analysis and nonmetric scaling identify the underlying cognitive dimensions consumers use to evaluate products or services.

The preference models considered are preference regression, first preference logit, and revealed preference logit. Preference regression and first preference logit select relative importance weights for attributes in order to best explain rank-order preferences or first preferences, respectively. Revealed preference models select relative importance weights for both the attractiveness attributes and the accessibility characteristics by analysis of observed choices. These importances identify those dimensions that most affect the consumer choice process and thus help managers identify which characteristics to stress in the formulation of strategy or policy.

The details of the research are described in the remainder of this paper. The second section reviews the theory and models used; the third describes the empirical setting and experimental design; the fourth and fifth evaluate the different models with respect to interpretability and predictive accuracy, respectively; and the conclusion presents a summary of results and indicates directions for further research.

## THEORY AND MODELS OF SHOPPING LOCATION CHOICE BEHAVIOR

The process by which consumers evaluate and choose among a set of alternatives can be described in different ways. In this study, we represent the consumer response process by the sequence of distinct but interrelated stages described in Figure 1. This simplified representation of perception, attractiveness, accessibility, and choice is a part of a more complex market process that describes interaction among individuals, information diffusion, changes in behavior based on experience, differences between market segments, etc. (2, 3, 4). Nonetheless, this representation provides a useful framework for the analysis of destination choice behavior and provides a critical link in any behavioral model of the transportation consumer.

Methodologies based on combining perception, preference, and choice have proved extremely productive in other contexts (2, 3). It is reasonable, therefore, to posit that these methodologies will enjoy similar success in transportation. They do not replace the disaggregate behavioral demand models now in common use but augment them, enhance their predictive abilities, and make them more responsive to the planning needs of today's managers.

## Modeling Consumer Perceptions

Consideration of consumer perceptions rather than direct (engineering) measures of alternatives makes it possible to include attributes or characteristics for which direct measures do not exist and to account for differences between consumers' subjective evaluation of alternatives and objective reality. The usefulness of incorporating nonengineering measures in travel choice behavior has been demonstrated in studies by Spear (5), Nicolaidis (6), and Dobson and Kehoe (7). Differences in perceptions among individuals or differences between perceived and engineering measures or both have been identified by Burnett (8) and Dobson and Tischer (9).

We shall examine three alternative perceptual models in this study. The simplest and most obvious method of representing consumer perceptions is by individual ratings for an exhaustive list of attributes. These attribute scales, called fundamental attributes, provide a complete description of consumer perceptions and are conceptually easy to use because no further data collection or analysis is required. Use of the complete list assumes that the individual simultaneously evaluates a long list of attributes in formulating preferences among alternatives. Alternatively, one can assume that underlying cognitive dimensions exist and that consumer ratings of attributes include a common component attributable to these cognitive dimensions, an attribute-specific component, and some measurement error. The common components or cognitive dimensions can be found by factor analysis of the attribute ratings across alternatives and consumers (2). The advantage of factor analysis over fundamental attributes is that it identifies a simpler perceptual structure that can provide clearer insight into how consumers perceive alternatives.

Finally, one can identify cognitive dimensions by analysis of perceived similarities between products or services. Nonmetric scaling positions alternatives in n-dimensional space so that the distance between pairs of alternatives corresponds as closely as possible to the reported similarity between them (10, 11). The advantage of nonmetric scaling over factor analysis is that it does not assume the ratings scales are metric, because scales are determined independently of the attributes and can uncover dimensions not represented in the attributes. However, nonmetric scaling requires additional, hard-to-collect data on similarity judgments; also, the scaling procedures are difficult and expensive to use. Furthermore, the assumptions built into these algorithms are very restrictive behavioral postulates and could, therefore, restrict the model. For example, a commonly used algorithm called INDSCAL assumes that all consumers have perceptions that are homogeneous subject to a scale transformation. Finally, the number of dimensions that can be identified is severely constrained by the number of stimuli (10, 12).

## Modeling Consumer Preferences by Attractiveness

We describe the consumer response process as one of perception, preference, and choice. The purpose of separating the preference and choice steps is to avoid confounding performance or attractiveness characteristics, which influence both preference and choice, with other characteristics such as availability, awareness, and accessibility, which primarily influence choice. In this study, this two-step process is tested by comparing importance weights and predictive ability of models, including an intermediate preference step, with revealed preference models that exclude the intermediate preference step.

The analysis of consumer preferences is directed toward finding a function that maps consumer perceptions into a preference rating, or attractiveness index. The preference models considered in this study determine the relative importance of the fundamental attributes or cognitive dimensions by estimating a linear compensatory model of the form

$$P_{ij} = \sum_k w_k d_{ijk} \qquad (1)$$

This model states that consumer i's preference or attractiveness index for product j, $P_{ij}$, is the weighted sum of his or her perceptions, $d_{ijk}$, of alternative j for attribute or dimension k where the estimated importance weights are average insights for the population. Three models are evaluated.

Preference regression statistically estimates the importance weights by using rank-order preference as the dependent variable and the consumers' perceptions as independent variables. Ordinary least squares is used to estimate importance weights, despite the implicit metric assumption, because it has been shown that these results are similar to those that would be obtained by more expensive monotonic regression (13). Preference regression uses full rank-order information in the estimation of importance weights.

Preference logit assumes that the true preference, $P^T_{ij}$, is composed of an observable part, $P_{ij}$, as in Equation 1, plus an error term, $e_{ij}$:

$$P^T_{ij} = P_{ij} + e_{ij} \qquad (2)$$

Assuming an independent Weibull-distributed error term makes it possible to derive a functional form for the probability $L_{ij}$ that consumer i ranks j as his or her first preference (14). This probability is given by

$$L_{ij} = \exp(P_{ij}) \Big/ \sum_l \exp(P_{il}) \qquad (3)$$

where the sum is over all alternatives, l. The importance weights are estimated by maximum likelihood techniques (14). The appeal of the logit model is that it explicitly models stochastic behavior (15) and that it makes no metric assumptions about preference rankings. Although it uses only first-preference information, it can be extended to use rank-preference information (16) with similar results.

Revealed preference assumes that the underlying preference weights must be obtained by analysis of choice behavior. It assumes that each individual selects an alternative that has the greatest utility to him or her. The importance weights, $w_k$, are estimated jointly with the importance of nonpreference characteristics such as the time, effort, or cost of obtaining a selected alternative (see Equation 6, below).

The advantage of the revealed preference model is that it does not rely on reported preference data but on observed choice behavior. However, the estimates of cognitive importance weights may be biased if the nonpreference choice elements are not carefully specified.

### Modeling Consumer Choice Behavior

The consumer response process is designed to explain and predict individual choice based on a model of perceptions and preferences. The choice model postulates that individual consumers associate a value $v_{ij}$ with each available alternative and select the alternative that has the greatest value. Our estimate of the individual value $\hat{v}_{ij}$ is a linear combination of the preference index, $P_{ij}$, and situational variables, $Z_{ijm}$, influencing choice behavior.

$$\hat{v}_{ij} = \beta_0 P_{ij} + \sum_m \beta_m Z_{ijm} \qquad (4)$$

The true value is equal to the estimated value plus a random component that represents unobserved influence and specification errors. Using the same distribution assumption as for preference logit, we obtain the multinomial logit choice model, which describes the probability of individual i's choosing alternative j on a single occasion by

$$C_{ij} = \exp(\hat{v}_{ij}) \Big/ \sum_l \exp(\hat{v}_{il}) \qquad (5)$$

When the preference index has not been estimated, the value function is formulated in terms of the fundamental attributes or cognitive dimensions,

$$V_{ij} = \beta_0 \sum_k w_k d_{ijk} + \sum_m \beta_m Z_{ijm} \qquad (6)$$

and the revealed preference importance weights, $w_k$, are estimated simultaneously with the other parameters of the choice model.

## EMPIRICAL SETTING

The empirical focus of this study is on non-grocery-shopping trips. Historically, researchers in both transportation (18) and marketing (19, 20) have emphasized the importance of accessibility or distance from the consumer's residence to the shopping center. Some studies have included measures of shopping location size, usually retail floor space or number of retail employees (21). Although size of shopping locations, which also represents the range of opportunities available to the shopper, is a relevant measure of attractiveness, it is unlikely to capture all the aspects of attractiveness that influence shopping location choice behavior. In order to understand the construct of shopping location attractiveness from the perspective of consumers, we must determine the cognitive dimensions of shopping location attractiveness, their relative importances in forming preferences, and their importance relative to accessibility in influencing choice behavior.

The models estimated in this study are based on data collected at shopping locations in the North Shore suburbs of Chicago. The process of sampling individuals at their chosen destination requires the use of choice-based estimation procedures to obtain consistent estimators of parameters (17). The data collected describe attitudes toward and use of seven shopping locations. The data used in this analysis include rank-order preference for each shopping location, similarity judgments for all pairs of shopping locations, direct ratings of each shopping location for 16 attributes, and frequency of trips to each location. The 16 attributes (see the list below) chosen to describe the general characteristics of shopping locations and the questionnaire were developed through extensive literature review, preliminary surveys, and analysis of developmental questionnaires (22). The 16 attributes are

1. Layout of store,
2. Ease of returning or servicing merchandise,
3. Prestige of store,
4. Variety or range of merchandise,
5. Quality of merchandise,
6. Availability of credit,
7. Reasonable price,
8. Availability of sale items (specials),
9. Free parking,
10. Stores located in a compact area,
11. Store atmosphere (heating, cooling, noise, crowds, etc.),
12. Ability to park where you want,
13. Shopping center atmosphere (pedestrian-only area, flowers and shrubs, covered walkways, etc.),
14. Courteous and helpful sales assistants,
15. Availability of a specific store, and
16. Number and variety of stores.

The analysis is based on a random sample of 500 consumers who reported familiarity with all seven shopping locations. All models were then tested for ability to predict on a saved-data sample of an additional 500 consumers. Predictions were quite good for all preference and choice models on factor analysis and fundamental attributes. Furthermore, all relative model comparisons were supported by the saved-data analysis (23).

The data collected did not include information on the costs (time, out-of-pocket cost, physical effort, etc.) of traveling to each of the shopping locations. Only the residential location of the shopper was obtained. For this reason accessibility is represented by the distance between each shopping location and the shopper's residence.

## MANAGERIAL INTERPRETABILITY

The primary goal of this study was to understand and explain consumer response in the selection of destinations for non-grocery-shopping trips. Thus initial analysis of and comparison between models was based on managerial interpretability. The model interpretations, which provide basic insight into the behavioral process, serve as a first test of model usefulness and validity.

### Perception Models

A plot of the average ratings of each shopping location for the 16 fundamental attributes (Figure 2) reveals a number of insights into the existing pattern of perceptions. However, the complexity of the figure and excessive amount of data in it make it difficult to focus on critical areas.

The alternative perceptual models, factor analysis and nonmetric scaling, produce simpler perceptual representations. Although the methods of analysis differ, each of these perceptual models identifies cognitive dimensions by structure matrixes that relate them to the 16 fundamental attributes. These structure matrixes are used to identify the cognitive perceptual dimensions.

**Figure 2. Map of fundamental attributes ratings for seven shopping locations.**

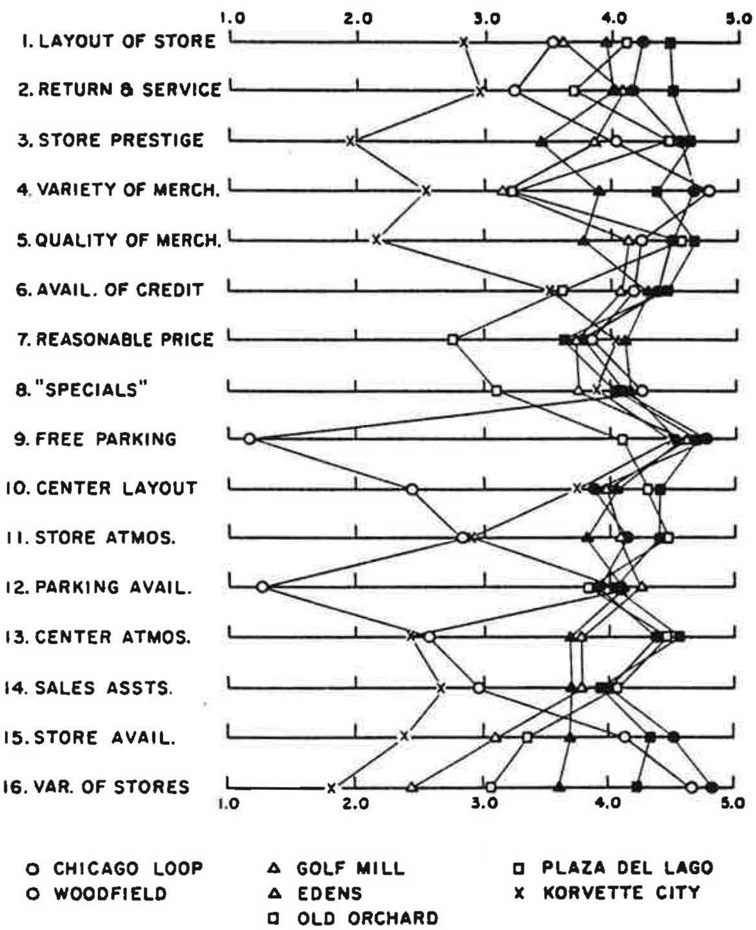



**Table 1. Structure matrixes for perceptual models.**

| | Factor Analysis, Factor Loadings | | | | Nonmetric Scaling Directional Cosines | | |
|---|---|---|---|---|---|---|---|
| Attribute Number | Variety | Quality and Satisfaction | Value | Parking | Variety | Quality Versus Value | Parking and Satisfaction |
| 1 | 0.27 | 0.58 | 0.16 | 0.20 | 0.22 | 0.50 | 0.84 |
| 2 | 0.10 | 0.53 | 0.34 | 0.26 | 0.32 | 0.12 | 0.94 |
| 3 | 0.34 | 0.82 | -0.00 | -0.06 | 0.30 | 0.80 | 0.52 |
| 4 | 0.67 | 0.33 | 0.31 | -0.19 | 0.93 | 0.36 | 0.08 |
| 5 | 0.31 | 0.81 | 0.04 | -0.07 | 0.30 | 0.81 | -0.51 |
| 6 | 0.16 | 0.34 | 0.49 | 0.05 | 0.88 | -0.09 | -0.47 |
| 7 | 0.07 | -0.06 | 0.60 | 0.11 | 0.49 | -0.85 | 0.19 |
| 8 | 0.22 | 0.07 | 0.74 | 0.01 | 0.79 | -0.59 | 0.17 |
| 9 | -0.15 | 0.07 | 0.04 | 0.81 | -0.29 | -0.55 | 0.78 |
| 10 | 0.03 | 0.31 | 0.07 | 0.56 | -0.45 | 0.04 | 0.89 |
| 11 | 0.08 | 0.66 | 0.03 | 0.40 | -0.20 | 0.45 | 0.87 |
| 12 | 0.15 | 0.11 | 0.11 | 0.84 | -0.46 | -0.48 | 0.75 |
| 13 | 0.24 | 0.69 | 0.04 | 0.40 | -0.10 | 0.48 | 0.87 |
| 14 | 0.17 | 0.56 | 0.15 | 0.32 | -0.05 | 0.41 | 0.91 |
| 15 | 0.62 | 0.32 | 0.20 | 0.03 | 0.87 | 0.43 | -0.24 |
| 16 | 0.83 | 0.29 | 0.16 | -0.17 | 0.92 | 0.39 | -0.05 |

Based on statistical rules and intuitive interpretations, the best results were obtained by using a three-dimensional perception space for nonmetric scaling and a four-dimensional space for factor analysis. The perceptual structures for each model are presented in Table 1.

Both models identify combinations of five basic constructs of variety, quality, satisfaction, value, and parking. These constructs are consistent with earlier studies by Singson (24), Monroe and Guiltinan (25), and Jolson and Spath (26). However, the grouping of these constructs is different between models. Factor analysis groups quality with satisfaction, while nonmetric scaling groups quality versus value and groups parking with satisfaction. The reverse directionality of quality and value in the nonmetric scaling solution undermines interpretability because it is impossible to identify a clear direction of goodness along this scale. The appropriateness of these alternative models will be examined in terms of their predictive abilities.

These models are used to develop perceptual maps of shopping locations based on the underlying cognitive dimensions. These maps are shown in Figure 3. It is immediately apparent that the maps are simpler for managers to interpret, but one can hypothesize that this simplicity comes at a cost of reduced detail. We must compare the predictive ability of these perceptual models with that obtained from the fundamental attributes.

Figure 3. Perceptual maps for models of consumer perceptions.

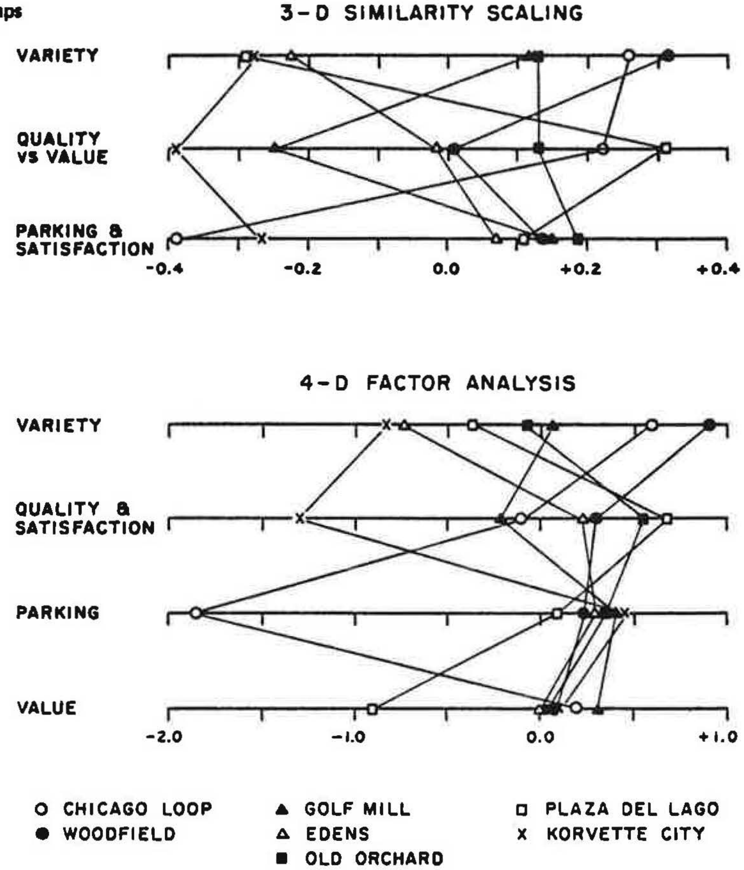



Table 2. Normalized importance weights for reduced models.

| | Factor Analysis | | | | Nonmetric Scaling | | |
|---|---|---|---|---|---|---|---|
| Consumer Model | Variety | Quality and Satisfaction | Value | Parking | Variety | Quality Versus Value | Parking and Satisfaction |
| Preference regression | 0.39 | 0.57 | 0.03* | 0.01* | 0.26 | 0.43 | 0.31 |
| First preference logit | 0.30 | 0.41 | 0.23 | 0.06* | 0.26 | 0.49 | 0.26 |
| Revealed preference | 0.05 | 0.48 | 0.29 | 0.19 | 0.22 | 0.76 | 0.03* |

*Not significant at the 5% level.

## Preference Models

The normalized importance weights for the various models on the two reduced perception structures are shown in Table 2. The most important dimension for each perception structure includes quality as a component. The importance weights estimated by preference regression and preference logit are similar for each of the perception structures. This robustness of direct preference models is important because it suggests that the choice of perception model is more crucial to the identification of strategically important policies than is the choice of preference model. The similarity of the importance weights between preference regression and preference logit for each perception model is partially retained in the revealed preference model. Quality-related dimensions remain the most important. However, the revealed preference weights differ for the other dimensions. For factor analysis, parking gains in importance at the expense of variety. In nonmetric scaling, quality versus value gains at the expense of parking. These shifts are strategically important as they produce insight into the strength of feeling about the aspects of shopping destinations.

For example, suppose (as will be shown later) that factor analysis is the preferred model for this data set. The increase in parking importance can be explained because it is partially confounded with accessibility. The decrease in variety importance can be explained in this data set by the characteristics of the destinations available to the residents of the North Shore suburbs of Chicago. That is, the two destinations highest in variety are least accessible. Thus, in the revealed preference choice model, variety and accessibility are highly correlated and the relative weights may not be stable. These correlations are reduced in the two-step choice mode.

Other hypotheses for the differences observed can be developed but not tested with the available data (23). The difference in results illustrates the importance of using both a revealed preference choice model and a two-step preference choice model. This use of convergent models is a powerful tool that can lead to insights not obtainable by either model alone. Note that in this case interpretations based on either model alone might miss the interactions between variety and accessibility.

Preference models estimated by using fundamental attributes consistently identify prestige of store, which is closely related to quality, as the most important

Table 3. Prediction tests.

| Consumer Model | Preference: First Preference Recovery | Preference: Rank Preference Recovery | Choice: Percent Correctly Predicted | Choice: Information |
|---|---|---|---|---|
| Base models | | | | |
| Equally likely | 14.3 | 14.3 | 14.3 | 0.0 |
| Market share | 26.7 | - | 18.5 | 24.7 |
| Distance only | - | - | 31.9 | 32.6 |
| Best fundamental attributes model | 55.6 | 37.9 | 32.7 | 39.2 |
| Theoretically best model | - | - | 38.7 | 100.0 |
| Factor analysis | | | | |
| Preference regression | 50.6 | 32.9 | 32.7 | 36.4 |
| First preference logit | 55.0 | 37.0 | 32.9 | 37.3 |
| Revealed preference | - | - | 32.6 | 38.5 |
| Nonmetric scaling | | | | |
| Preference regression | 36.6 | 25.1 | 31.8 | 33.5 |
| First preference logit | 34.8 | 24.4 | 31.8 | 33.7 |
| Revealed preference | - | - | 32.2 | 34.1 |

attribute. However, there is a high degree of instability in the estimated values for other importance weights due to multicolinearity. Thus, while there is consistency in preference estimation with reduced perceptual dimensions, there is a high degree of instability in the estimation of preference weights for fundamental attributes (23).

## PREDICTIVE ABILITY

This section tests the ability of each model to predict preference and choice on the estimation data sample. Separate predictions on a saved-data sample of 500 observations support the results reported here (23).

### Prediction Formation

Individual predictions are made by applying the alternative model structures to each individual's ratings on the fundamental attributes and distance for each shopping location. The prediction process consists of the following sequence of steps.

1. Perception measures are obtained by applying the perception models to the fundamental attribute ratings or individual similarity measures to obtain perception scores for the cognitive dimensions.
2. Perception scores formulated are combined with the estimated or measured importance weights to obtain individual preference (attractiveness) measures for each shopping location.
3. Preference measures are rank ordered to obtain individual preference ranks used in the analysis of preference prediction.
4. Preference and accessibility measures are used in the choice models to predict overall ratings and choice probabilities for each shopping location. These predictions are used in the analysis of choice prediction.

Preference predictions are made with each of the perceptual and preference models, and choice predictions are made with each of the linked choice models. These predictions are compared to a variety of base models that serve as bounds on prediction and help indicate the power of each set of models. The lower bounds include random (equally likely) preference choice in proportion to market share with distance as the only variable. Perfect prediction of choice frequency serves as the upper bound.

For a detailed discussion of these bounds see Hauser (27). Prediction with the best fundamental attributes models identifies the loss in predictive ability, which may result from reduction in cognitive data through faster analysis or nonmetric scaling.

### Preference Prediction Results

Preference prediction results for each perceptual and preference model are reported in Table 3. Factor analysis dominates nonmetric scaling with respect to first preference recovery and rank preference recovery. Furthermore, factor analysis does as well as fundamental attributes, indicating that there is no loss in predictive ability due to the simplification of perception structure. Preference logit is slightly superior to preference regression, but the most important differences among models is in the choice of perception model.

### Choice Prediction Results

Choice prediction results are presented in Table 3 in terms of percentage of correct predictions and percentage of information explained (27). The model comparisons are the same as for the preference predictions, but the differences are not as great. Revealed preference does better on choice but not significantly better than the two-step preference and choice models.

The overall predictions are quite good. The best model (factor analysis with preference logit) correctly predicts 32.9 percent of the choice occasions as opposed to the 38.7 percent that is theoretically possible in this population.

The goal was to predict frequency of choice for each individual. Since situational variables were not included, the model cannot predict for each choice occasion. Furthermore, the maximum information is quite respectable compared to previous results with similar models. The factor analysis models do well compared to the equally likely and market share proportional models and are almost as good as the fundamental attributes model.

Of the perceptual models tested, factor analysis is the superior predictive model for both preference and choice. It does as well as the fundamental attributes but provides important simplification of the perceptual process. Thus for this data set it appears that factor analysis is most representative of the consumers' cognitive process. It is interesting to note that these results have since been replicated on another data set (28).

The differences among the preference models are less dramatic. This further supports the observation that the preference models are relatively robust and that the selection of preference model is less crucial than the selection of perceptual model. The predictive similarity of revealed preference and two-step preference and choice models supports the conjecture that the use of these models in parallel is an important managerial tool.

## CONCLUSIONS

The focus of this research is on the behavioral modeling of destination choice. The models developed use state-of-the-art techniques in marketing and transportation to provide strategic, policy-sensitive models for the explanation and prediction of destination choice behavior for non-grocery-shopping trips. The products of this research are a behavioral model of destination choice and an identification of the most accurate and useful techniques to analyze destination choice behavior.

### Behavioral Model of Destination Choice

The interpretations and insight about consumer behavior come from the combined analysis. This process of convergent analysis provides insights that might not be obtained from a single model structure. In summarizing these results we look for consistent results when the models converge and "best model" results when they diverge. The primary results are that

1. Attractiveness can be measured with combined perceptual and preference models, and this measure predicts well and provides useful insight into consumer behavior;
2. Five basic constructs best measured by the factor analysis perceptual model describe shopping destination attractiveness: variety, quality, satisfaction, value, and parking;
3. Quality is consistently the most important construct of shopping destination attractiveness, and prestige of store appears to be the most important aspect of quality; and
4. Both attractiveness and accessibility are important determinants of travel behavior, and any destination choice model should contain good measures of both.

### Comparison of Model Structures

A number of alternative techniques are tested to select the best models for the analysis of destination choice. The results suggest that factor analysis is the best perceptual model for identifying a concise set of dimensions to describe the consumers' cognitive process and that the statistical preference models (first preference logit and preference regression) are reasonably robust in providing consistent predictions and similar interpretations.

Based on these results but subject to confirmation in other empirical studies, we recommend that statistical analyses of consumer destination choice be based on factor analysis to identify perception, preference regression or first-preference logit or both to identify importance weights, and convergent analysis with revealed preference and two-step preference and choice models to analyze choice behavior.

## ACKNOWLEDGMENTS

This research was supported in part by the U.S. Department of Transportation, Office of University Research. Technical assistance was provided by Joseph Prashker, Bruce Bagamary, and Steve Shugan. The analysis is based on data previously collected under the direction of Peter R. Stopher and Peter L. Watson.

## REFERENCES

1. S. Hanson. On Assessing Individuals' Attitudes Towards Potential Travel Destinations: A Research Strategy. Transportation Research Forum, Vol. 13, 1974, pp. 363-370.
2. J. R. Hauser and G. L. Urban. A Normative Methodology for Modeling Consumer Response to Innovations. Operations Research, Vol. 25, No. 4, July-Aug. 1972, pp. 579-619.
3. A. D. Shocker and V. Srinivasan. A Consumer Based Methodology for the Identification of New Product Ideas. Management Science, Vol. 20, Feb. 1974, pp. 921-938.
4. J. R. Hauser and F. S. Koppelman. Designing Transportation Services: A Marketing Approach. Transportation Research Forum, Vol. 16, 1977, pp. 628-651.
5. B. D. Spear. The Development of a Generalized Convenience Variable for Models of Mode Choice. Cornell Univ., Ithaca, NY, PhD thesis, June 1974.
6. G. C. Nicolaidis. Quantification of the Comfort Variable. Transportation Research, Vol. 9, No. 1, 1975, pp. 55-56.
7. R. Dobson and J. F. Kehoe. Disaggregated Behavioral Views of Transportation Attributes. HRB, Highway Research Record 527, 1974, pp. 1-15.
8. K. P. Burnett. The Dimensions of Alternatives in Spatial Choice Processes. Geographical Analysis, Vol. 5, No. 3, 1973, pp. 181-204.
9. R. Dobson and M. L. Tischer. Beliefs About Buses, Carpools, and Single Occupant Autos: A Market Segmentation Approach. Transportation Research Forum, Vol. 15, 1976, pp. 200-209.
10. P. E. Green and V. Rao. Applied Multidimensional Scaling. Holt, Rinehart and Winston, New York, 1972.
11. J. B. Kruskal. Multidimensional Scaling by Optimizing Goodness of Fit to a Nonmetric Hypothesis. Psychometrika, Vol. 29, Mar. 1964, pp. 1-27.
12. D. Klahr. A Monte Carlo Investigation of the Statistical Significance of Kruskal's Non-Metric Scaling Procedure. Psychometrika, Vol. 34, No. 3, Sept. 1969, pp. 319-330.
13. F. J. Carmone, P. E. Green, and A. K. Jain. The Robustness of Conjoint Analysis: Some Monte Carlo Results. Univ. of Pennsylvania, Working Paper, May 1976.
14. D. McFadden. Conditional Logit Analysis of Qualitative Choice Behavior. In Frontiers in Econometrics (P. Zarembka, ed.), Academic Press, New York, 1970, pp. 105-142.
15. F. Bass. The Theory of Stochastic Preference and Brand Switching. Journal of Marketing Research, Vol. 11, Feb. 1974, pp. 1-20.
16. R. D. Luce and P. Suppes. Preference Utility and Subjective Probability. In Handbook of Mathematical Psychology (R. D. Luce, R. R. Bush, and E. Galanter, eds.), Wiley, New York, Vol. 3, 1965, pp. 249-410.
17. S. R. Lerman, C. F. Manski, and T. J. Atherton. Non-Random Sampling in the Calibration of Disaggregate Choice Models. Federal Highway Administration, Final Rept. No. PO-6-3-0021, Feb. 1976.
18. M. Ben-Akiva. Structure of Passenger Travel Demand Models. TRB, Transportation Research Record 526, 1974, pp. 26-65.
19. L. P. Bucklin. The Concepts of Mass in Intra-

Urban Shopping. Journal of Marketing, Vol. 31, Oct. 1967, pp. 37-42.
20. D. L. Huff. A Probabilistic Analysis of Shopping Center Trade Areas. Land Economics, Vol. 39, No. 1, Feb. 1963.
21. J. B. Mason. Retail Market Area Shape and Structure: Problems and Prospects. Advances in Consumer Research, Vol. 2, 1975, pp. 173-181.
22. P. R. Stopher, P. L. Watson, and J. M. Blin. A Method for Assessing Pricing and Structural Changes on Transport Mode Use. Transportation Center, Northwestern University, Evanston, IL, Interim Rept. to the Office of University Research, U.S. Department of Transportation, Sept. 1974.
23. F. S. Koppelman and J. R. Hauser. Consumer Travel Choice Behavior: An Empirical Analysis of Destination Choice for Non-Grocery Shopping Trips. Transportation Center, Northwestern University, Evanston, IL, Research Rept. 414-09, July 1977.
24. R. Singson. Multidimensional Scaling Analysis of Store Image and Shopping Behavior. Journal of Retailing, Vol. 51, No. 2, 1975.
25. K. B. Monroe and J. P. Guiltinan. A Path-Analytic Exploration of Retail Patronage Influences. Journal of Consumer Research, Vol. 2, June 1975.
26. M. A. Jolson and W. F. Spath. Understanding and Fulfilling Shoppers' Requirements. Journal of Retailing, Vol. 49, No. 2, summer 1973.
27. J. R. Hauser. Testing the Accuracy, Usefulness, and Significance of Probabilistic Choice Models: An Information Theoretic Approach. Transportation Center, Northwestern University, Evanston, IL, working paper, April 1976.
28. P. Simmie. Alternative Perceptual Models: Reproducibility, Validity and Data Intergrity. Proc., American Modeling Association's Educators' Conference, Aug. 1978.

*Publication of this paper sponsored by Committee on Passenger Travel Demand Forecasting and Committee on Traveler Behavior and Values.*

# Trip Distribution in Subregional Analysis

Stephen M. Howe, North Central Texas Council of Governments, Arlington
Yehuda Gur, John Hamburg and Associates, Philadelphia

**The paper describes the formulation and calibration of the access and land development trip distribution gravity model (ALDGRAV) for use in highway planning at a subregional level. The model is being used as an element of the thoroughfare analysis process (TAP), which, in turn, is one module of the thoroughfare planning system (TPS). TPS has been developed by the North Central Texas Council of Governments, in close cooperation with the local governments, to answer present planning needs, in particular to provide tools for orderly, inexpensive, and fast response evaluation of small- and medium-scale strategies. TAP provides the analysis capabilities of the system. The paper introduces the hierarchy of objectives, design requirements, and the resulting design decisions of TPS, TAP, and the ALDGRAV trip distribution model. A detailed description of the latter is given.**

The North Central Texas Council of Governments (NCTCOG), together with the participating local governments, has developed the thoroughfare planning system (TPS). The system is designed to answer many of the recent needs in the field that have arisen primarily from shifting stress from large-scale, capital-intensive projects to subregional projects. Major objectives of TPS include providing tools for the planning of the principal and minor arterial network that supports the freeway system in the region and tools for evaluating projects such as the annual capital improvement programs of individual communities, on a local scale, and providing support, cost effectively, for the analysis of small- and medium-scale projects within the framework of the regional thoroughfare plan.

TPS is described in detail elsewhere (1). Its major elements include: (a) an approved regional thoroughfare plan complete with design standards, (b) a base inventory of the thoroughfare system with procedures for continuous updates, (c) a thoroughfare information system (TIS) that facilitates the storage and easy access of both inventory data and analysis results, (d) a thoroughfare analysis process (TAP) to evaluate the impact of alternative strategies, and (e) a methodology for evaluating transportation system management (TSM) strategies.

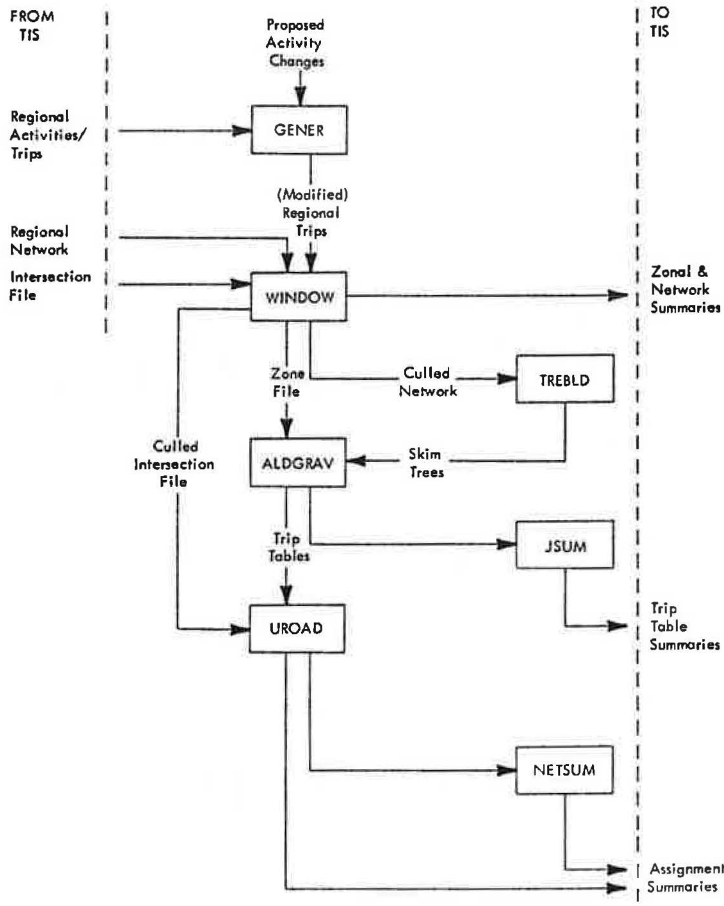
## THOROUGHFARE ANALYSIS PROCESS

TAP is the travel simulation component of the TPS and has the following specific design requirements. It must be able to analyze a wide range of potential strategies, such as the effect of land-use changes (e.g., a new shopping center), the effect of major TSM strategies, and small- and medium-scale capital projects, and it should develop and maintain the regional long-range plan and analyze small-scale problems quickly and inexpensively. The structure of TAP is described in Figure 1. Its logic closely follows that of the conventional urban transportation planning system. Major innovations in the system include windowing and streamlined processing.

Windowing means that by using computerized procedures subfiles for analysis are built from base data files that describe the zones and networks of the region in much detail. Typically, these subfiles include detailed presentation for the area of interest; the level of detail decreases gradually with distance from the area of interest. Different subfiles are built for practically every analysis.

In streamlined processing, through both the selection of models and the use of computerized procedures, it is possible to go through the whole analysis process for one alternative in one or two computer jobs. At the same

**Figure 1. TAP program sequence.**




time, the user can easily change the structure of the process in response to special analysis requirements.

## CONSTRUCTION OF THE TRIP TABLE

The procedure for constructing the trip table is a major part of TAP. It determines, to a large extent, the structure of the whole process in response to special analysis requirements. The following is a short description of the process as implemented. This description is followed by a discussion of alternative approaches and the reasoning behind the selected approach.

A major input to TAP is the zone data file, which includes (for each of the 7000 traffic survey zones) estimates of activity such as population, employment, service employment, and median income and productions and attractions of vehicle trips by five trip purposes. Such a file is prepared, externally to TAP, for each of the likely planning years, for instance, for each 5-year period until the year 2000.

Through the program GENER, the user can introduce changes in the activity measures for individual traffic survey zones. Using a combined trip generation, mode-split, and auto occupancy model, the program recalculates trip productions and attractions. Thus it is possible to introduce into the process and analyze the effect of proposed land-use changes.

The zone data file is then put into the program WINDOW. Depending on the user-specified window structure, WINDOW aggregates the zone data to form the zone structure used in the analysis. The resulting analysis zones might include individual traffic survey zones within the area of interest, with aggregation of the zones elsewhere according to a five-level hierarchy. Typically, 150 to 300 analysis zones are created by aggregating the original 7000 traffic survey zones. WINDOW also processes the network by culling links from the base network according to their importance in a five-level hierarchy and their distance from the area of interest. Next, the zones are connected to the network through approach links, or directly by load nodes.

Minimum impedance skim trees are built, using the program TREBLD. Trees are built on a prestressed network; i.e., speeds are calculated by considering average expected link volumes. The network might be prestressed to consider average daily loads and/or peak period loads. Impedance is calculated as a linear combination of time, operating costs, and tolls.

The trip file, together with the skim trees, is put into the program ALDGRAV for trip distribution. Trips are distributed separately by five trip purposes using the corresponding skim trees. A detailed description of the trip distribution process is given below.

By using the program UMATRIX, the purpose-specific trip tables are combined and, if necessary, transposed to create the final trip table. At this point, special trips (through and airport trips) are also added to the table.

## MAJOR DESIGN CONSIDERATIONS

### Trip Generation

It was possible, obviously, to make the trip generation model an integral part of TAP and to require as input only the estimated zone activity levels. This approach was rejected, because in the analysis of small-scale problems there is no need to repeat the entire trip generation for each analysis. Selective updates through GENER are sufficient for analysis of localized land-use effects.

Another major problem was the method for mode-split analysis (or, more accurately, the estimation of the number of auto person trips, given the number of total person trips). Available procedures for mode-split estimation are rather costly and require the input of skim trees by mode, as well as various zone data. It is clearly infeasible to go through such a process for the whole region for every analysis of a small-scale project. Thus, the base trip generation model includes a full mode-split analysis.

Mode split for activity updates (within TAP) is performed by using the resulting zone-split factors. It is implicitly assumed that in most projects of the type analyzed by TAP, system changes are not large enough to cause significant changes in mode split. Obviously, whenever this assumption is not justified, a full-scale mode-split analysis (outside TAP) has to be made.

### Trip Distribution

Conceptually, it is possible to treat trip distribution in the same way as trip generation is treated, namely, to distribute the trips outside TAP and to aggregate the resulting trip table for each window. This approach has been rejected, since the cost of a 7000 × 7000-zone trip distribution would be prohibitive. Moreover, connecting the 7000 traffic survey zones to the network (for purposes of skim tree building) would result in an impossibly large network.

In TAP, trip distribution is done after constructing the window by using the 150 to 300 analysis zones typically resulting from the windowing phase. It is possible to use this approach only if the performance of the trip distribution model is not overly sensitive to area aggregation. Nihan and Miller (2) have shown that a properly formulated gravity model possesses this attribute. In a number of applications in New York it was shown that the ALDGRAV model, in particular, gives very stable results under a wide range of aggregation schemes. The ALDGRAV model and its use in TAP are described in detail in the following sections.

## TRIP DISTRIBUTION MODEL

The trip distribution model used in TAP is ALDGRAV, a gravity model formulation adapted from the access and land development (ALD) model originally developed by Schneider (3, 4, 5) and further discussed more recently by Kaplan (6).

### ALDGRAV Concepts

In TAP, the ALDGRAV model is used to distribute trips from production end to attraction end. Trip productions and attractions, for each traffic survey zone, are calculated apart from TAP and then aggregated according to the creation of analysis zones by program WINDOW within TAP. The trips are then distributed by the TAP version of ALDGRAV, which embodies the following basic assumptions.

1. Probability maximization is applicable to the distribution of trips;
2. For a given group of trip makers, the sensitivity of travel to the disutility is not a single value but ranges over a continuum;
3. The total disutility of travel, incurred by the trips produced from a given zone, must be finite; and
4. For a given zone, the input number of attractions is a surrogate measure of the attractiveness of that zone to trip makers.

The application of probability maximization, with the constraint implicit in assumption 3, yields the ALDGRAV model form. For a more rigorous discussion, the reader is referred to Kaplan (6), from which much of the ensuing discussion is also excerpted. The basic gravity model formulation may be expressed as

$$V_{ij} = V_i[G(F_{ij})A_j/\Sigma_r G(F_{ir})A_r] \quad (1)$$

where

$V_{ij}$ = number of trips produced by zone i and attracted to zone j,
$V_i$ = total number of trips produced by zone i,
$G(\cdot)$ = travel (decay) function representing the rate at which attractiveness declines with increasing travel disutility,
$F_{ij}$ = disutility of travel from zone i to zone j, and
$A_j$ = attractiveness (number of attractions) for zone j.

Equation 1 can be interpreted as a share formula that allocates the total productions from zone i, $V_i$, among alternative attraction zones, according to their relative attractiveness weighted by their corresponding decay values.

Specific gravity formulations are distinguished by different forms of the travel function $G(\cdot)$. Examples include

1. Inverse power function $G(F) = F^{-a}$,
2. Negative exponential function $G(F) = \exp(-aF)$,
3. Combined inverse power and negative exponential function $G(F) = F^{-a}[\exp(-bF)]$, and
4. Gamma density function, $G(F) = F^{a-1}\exp(-F)/\Gamma(a)$.

The travel function used in ALDGRAV is somewhat more complex than the above functions but can be related to the negative exponential function 2 as follows.

If basic assumption 2 is replaced by the simplified assumption of a single value, a, for traveler sensitivity, one derives the gravity model form (Equation 1) with negative exponential travel function 2. This model has been derived from entropy maximization principles by Wilson (7). However, the ALDGRAV formulation is based on the theoretically more complete assumption 2 that leads to integration over a range of sensitivity values and results in the gravity form with the ALDGRAV travel function

$$G(F) = K_2(2\sqrt{aF})/4aF \quad (2)$$

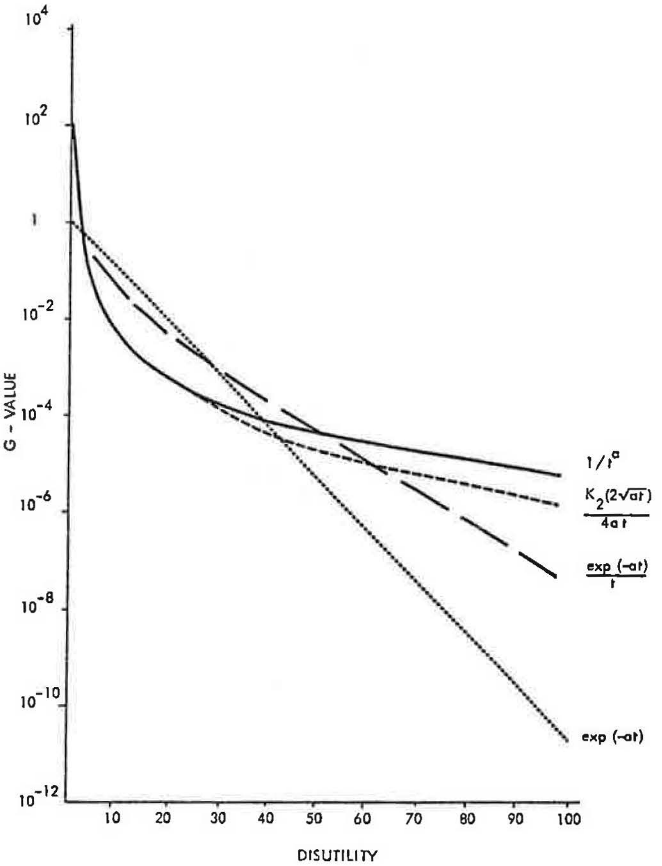
where $K_2$ is the modified Bessel function of second kind and second order, and a is a value representing an average traveler sensitivity.

For comparison, alternative travel functions are plotted in Figure 2. The value of the a constant was chosen to ensure comparability of the four functions, as follows:

1. $G(t) = t^{-a}$; $a = 2.625$,
2. $G(t) = e^{-at}$; $a = 0.260$,
3. $G(t) = \frac{e^{-at}}{t}$; $a = 0.143$, and
4. $G(t) = K_2(2\sqrt{at})/4at$; $a = 0.100$.

Regardless of the a-values, the Bessel function will always have a faster decay rate than the negative exponential at very small disutility values and a slower decay rate over large disutility values. No such general statement can be made about the comparison with the inverse

Figure 2. Comparison of alternative travel functions.




power or the combined functions.

In its application in TAP, the ALDGRAV model is doubly constrained; i.e., Equation 1 is iteratively adjusted to balance the trips received by each zone to the input number of attractions.

## Travel Disutility

For interzone trips from zone i to zone j the travel disutility measure used in TAP is

$$F_{ij} = T_{ij} + TP(AT_i) \tag{3}$$

where

$T_{ij} = W_1 * t_{ij} + W_2 * c_{ij} + W_3 * f_{ij}$ is the total travel impedance from zone i to zone j,
$t_{ij}$ = travel time from zone i to zone j,
$c_{ij}$ = operating cost from zone i to zone j,
$f_{ij}$ = tolls from zone i to zone j, and
$W_1, W_2, W_3$ = weighting parameters (uniform for the whole region).

$TP(AT_i)$ is a fixed penalty assessed according to the area type, $AT_i$, of zone i.

Conceptually, the fixed penalty reflects factors such as cost of owning the car, parking costs, and walking time from the parking to the final destination. The fixed penalty may have a different value for each of the four area types used in TAP (numbered by Urban Mass Transportation Administration (UMTA) convention and corresponding roughly to a trip-end density classification)

1 = Central business district (CBD) (20 000+ daily one-way person trip ends per square mile),
2 = CBD fringe (5000-19 999 trips per square mile),
3 = Suburban (1000-4999 trips per square mile), and
5 = Rural (0-999 trips per square mile).

For intrazone trips the calculation of travel disutility differs somewhat. Generally, intrazone travel time cannot be obtained directly from conventional skim trees. Within ALDGRAV, therefore, the intrazone travel time is estimated as a function of the radius of the zone. This quality is, in turn, divided by the intrazone speed to estimate the average intrazone travel time, to which the fixed penalty is added to yield the intrazone disutility. Note that, within TAP, the calculation of intrazone disutility is of special importance because of the large variations in zone sizes due to windowing. Proper calculation of the intrazone disutilities plays a major role in ensuring stable model performance under varying aggregation schemes.

A special treatment has been established for the distribution of external-local trips. These trips are somewhat unique, due to the fact that they are generally longer than internal trips and that only the within-region portion of these trips is described by the skim trees.

The fixed penalty (TP in Equation 3) can be interpreted, in the case of external trips, as the average impedance of that part of the trip outside the region. This interpretation is fully compatible with the theory of ALDGRAV. By treating the external-local trips as a special trip type, it is possible to assign to them an appropriate fixed penalty, as required.

Another unique attribute of these trips is that their distribution is not as dependent on the value of the intraregional impedance as that of the other trip types. Thus, in order to ensure their smooth distribution within the region, the trips are "flopped"; i.e., the internal end of the trip is considered the production end, while the external station is considered the attraction end.

## Calibration of Model Parameters

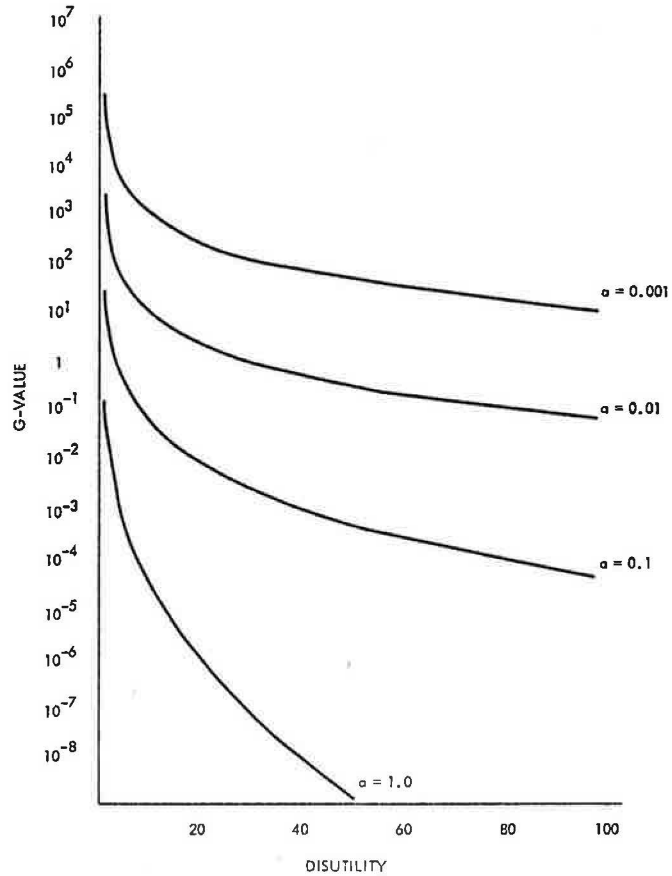
The ALDGRAV parameters requiring calibration, for use in TAP, are (a) the relative weights of the different impedance components, (b) the multiplier constant a, (c) the fixed time penalty TP associated with each area type, and (d) the average speed SI associated with intrazone trips within each area type. These parameters affect the simulation of trip distribution patterns through their effects on travel function and calculation of travel disutility.

Figure 3 shows how the shape of the travel function G(•) is affected by the value of a. With all fixed time penalties set at 0, an increase in a-value determines a sharper rate of declining attractiveness with increasing disutility and hence a shorter mean trip length (disutility).

Differences in mean trip length by area type can be simulated by adjusting the fixed penalty TP associated with each area type. Consider, in Figure 3, the curve associated with a = 0.001: the addition of TP to the disutility moves the ordinate to the right, i.e., shifts the curve to the left. Thus, the rate of decline in attractiveness becomes more gradual over the range of interest, and the mean travel disutility increases.

The a-value and the TP values determine the mean interzone travel disutility. With a and TP fixed, the intrazone speed, SI, can then be adjusted to determine the intrazone percentage of trips for each area type. By decreasing SI the intrazone disutility is increased, and hence the intrazone percentage is decreased. Adjustments in SI do not affect the interzone disutilities, $F_{ij}$, and therefore have no direct effect on interzone trip

Figure 3. G-values versus disutility by a-values.




lengths, although in practice the latter may be slightly affected because of the competition for trip ends implicit in the balancing of trips received with input attractions.

## Calibration Procedure and Criteria

The calibration procedure is based on accurate approximations of observed average interzone trip length (or travel time) by area type and of the interzone percentage of trips, by area type. The rationale for these approximations is that vehicle kilometers of travel (VKMT), or vehicle hours of travel (VHT), by area type are thereby accurately approximated, since VKMT equals mean trip length times percentage of interzone times number of trips and VHT equals mean travel times percentage of interzone times number of trips.

As an additional check on the validity of the model parameters, of course, simulated and observed volumes are compared for closeness of fit, particularly on major interchanges.

The calibration itself is conducted in an essentially stepwise cut-and-try fashion. From an initial set of parameter values, the multiplier a is first adjusted to roughly approximate regional mean trip length (or travel time), but with allowance for adjustment within each area type. The fixed penalties, TP, are then adjusted to approximate the mean within each area type. Finally, the intrazone speeds, SI, are set to give the correct intrazone percentage. Although the effects of the parameters are interrelated, with the aid of manual calculations the calibration procedure thus organized can accomplish the basic criteria in three to five test runs.

# CALIBRATION RESULTS

This section documents the calibration of ALDGRAV for distribution of vehicle trips in TAP for each of the following trip purposes: home-based work, auto driver (HBW); home-based non-work, auto driver (HNW); non-home-based, auto driver (NHB); truck and taxi (T&T); and internal versus external (I/E).

## Calibration Results

The base data for calibration were taken from origin-destination survey data compiled in a 1964 home interview survey conducted in the Dallas-Fort Worth area. The trip data were redefined in production-attraction format and expanded to form vehicle trip tables for each of the five trip purposes. The zone structure used in calibration consists of 504 regional analysis areas (RAAs), plus 18 external stations. For analysis purposes, the RAAs are aggregated into 39 jurisdiction districts. The districts and external stations are shown in Figure 4. The calibration effort utilized travel-time skim trees compatible with the trip tables in zone structure, base year, and peak versus off-peak conditions.

Shown in Tables 1, 2, and 3 are the calibration results for HBW, HNW, and NHB trip purposes, which collectively constitute more than 80 percent of vehicle trips in the region. A comparison of observed and simulated trip patterns with respect to the basic criteria, interzone percentage, and average interzone time, is shown. Both observed and simulated trip tables were aggregated (squeezed) for comparison of district-district interchanges, and a classification of major interchanges by percent error is shown for each trip purpose. Also shown are the calibrated parameters.

To briefly evaluate the calibration results, the interzone percentage and the average interzone travel time have generally been matched quite closely, even when broken out by area type. For major interchanges, the accuracy summaries are encouraging, particularly in view of the fact that these results were obtained without the use of K-factors. (Aside from the usual questions of behavioral validity and temporal stability, K-factors present additional problems for planning with a flexible zone structure.)

The errors did not appear to be systematic except in the case of HNW and, to a lesser extent, NHB. For these trip purposes, the simulated within-district percentages tended to be lower than observed. As noted above, however, simulation results were accurate in the interzone percentages, as well as in the average interzone impedance. The implication is that there is a propensity, particularly in HNW travel, to go either to a neighboring zone or to a distant one, which is not fully captured in the model. In other words, for interzone trips, the observed impedance distribution curve is flatter, or less peaked, than the simulated curve. Possible solutions would be to go to a long and a short stratification (this creates problems in definition) or to separate home-based shopping from other HNW purposes. This is one of the issues to be addressed in future research.

# NEED FOR FURTHER RESEARCH AND DEVELOPMENT

In spite of the generally satisfactory performance of TAP and its procedure for constructing trip tables in particular, there are still a number of areas where more study is needed and likely to be highly cost effective. The following list of subjects to be studied reflects our

**Figure 4. District definition and external stations.**

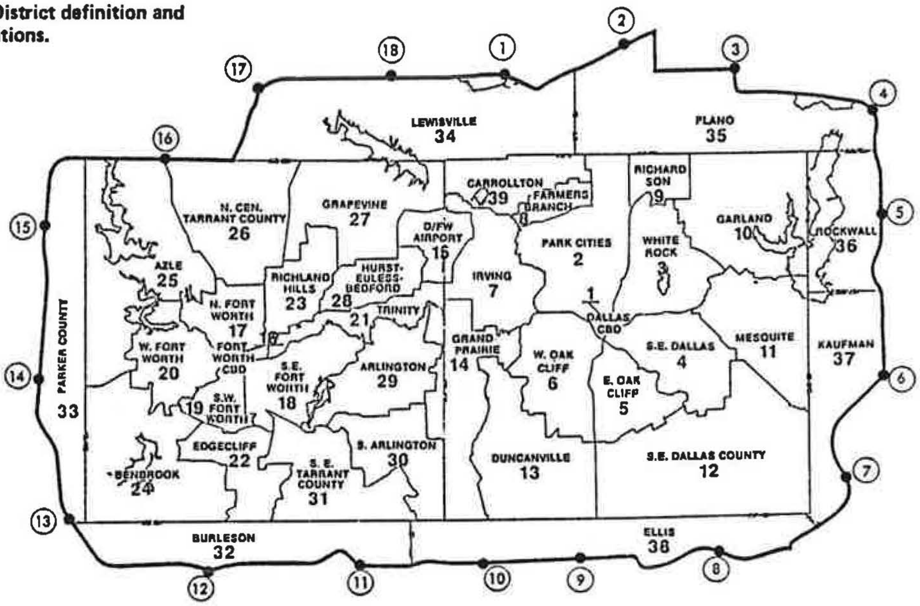



**Table 1. HBW auto driver trips.**

| Area Type | No. of Trips | Percentage of Total | Observed: Percentage of Interzone | Observed: Average Travel Time (min) | Simulated: Percentage of Interzone | Simulated: Average Travel Time (min) |
|---|---|---|---|---|---|---|
| 1. CBD | 1 748 | 0.2 | 100.0 | 8.9 | 98.8 | 8.2 |
| 2. CBD fringe | 140 936 | 20.1 | 95.3 | 12.7 | 95.9 | 12.1 |
| 3. Suburban | 514 652 | 73.4 | 92.6 | 16.5 | 93.1 | 16.6 |
| 5. Rural | 44 184 | 6.3 | 79.4 | 24.2 | 84.8 | 24.2 |
| Total | 701 520 | 100.0 | 92.4 | 16.1 | 93.1 | 16.1 |

Note: Percentages of error for all district-district interchanges greater than 5000 trips were 0-10 for 14, 10-20 for 11, 20-30 for 5, 30-40 for 4, 40-50 for 0, and 50-60 for 1; the parameters used were a = 0.54 and SI 150 for TP 12, SI 400 for TP 12, SI 600 for TP 10, SI 600 for TP 3, and SI 600 for TP 3.

**Table 2. HNW auto driver trips.**

| Area Type | No. of Trips | Percentage of Total | Observed: Percentage of Interzone | Observed: Average Travel Time (min) | Simulated: Percentage of Interzone | Simulated: Average Travel Time (min) |
|---|---|---|---|---|---|---|
| 1. CBD | 3471 | 0.2 | 98.4 | 8.0 | 98.7 | 8.0 |
| 2. CBD fringe | 295 006 | 17.0 | 79.5 | 9.7 | 82.4 | 9.2 |
| 3. Suburban | 1 322 180 | 76.2 | 67.5 | 10.6 | 66.8 | 10.5 |
| 5. Rural | 113 702 | 6.6 | 56.0 | 16.5 | 55.6 | 16.6 |
| Total | 1 734 359 | 100.0 | 68.8 | 10.7 | 68.8 | 10.6 |

Note: Percentages of error for all district-district interchanges greater than 10 000 trips were 0-10 for 9, 10-20 for 13, 20-30 for 3, but none above; the parameters used were a = 1.20 and SI 150 for TP 13, SI 420 for TP 5, SI 420 for TP 3, SI 500 for TP 0, and SI 500 for TP 0.

**Table 3. NHB auto driver trips.**

| Area Type | No. of Trips | Percentage of Total | Observed: Percentage of Interzone | Observed: Average Travel Time (min) | Simulated: Percentage of Interzone | Simulated: Average Travel Time (min) |
|---|---|---|---|---|---|---|
| 1. CBD | 76 582 | 8.5 | 98.6 | 11.5 | 96.1 | 11.3 |
| 2. CBD fringe | 204 907 | 22.8 | 86.8 | 10.3 | 89.1 | 9.8 |
| 3. Suburban | 586 337 | 65.1 | 75.9 | 11.6 | 76.2 | 11.5 |
| 5. Rural | 32 688 | 3.6 | 57.6 | 17.5 | 58.2 | 17.5 |
| Total | 900 514 | 100.0 | 79.6 | 11.4 | 80.2 | 11.2 |

Note: Percentages of error for all district-district interchanges greater than 5000 trips were 0-10 for 16, 10-20 for 14, 20-30 for 11, but none above; the parameters used were a = 1.2 and SI 100 for TP 25, SI 450 for TP 9, SI 450 for TP 7, SI 550 for TP 2, and SI 550 for TP 2.

present major concerns; it is not intended to be comprehensive or exhaustive.

### Multimodal Windowing

The extension of the present capabilities of TAP to multimodal analysis seems, naturally, to be the next order of business. The unimodal capabilities of TAP are clearly insufficient for modern planning. The problems of windowing for transit analysis might be rather complicated; specifically, the structure of transit networks will require more involved network culling techniques, compared to the techniques used for highway networks. Moreover, conventional mode-choice models are rather sensitive to area aggregation (because of the importance of access-egress impedance). They might perform poorly within the framework of windowing, where skim trees are available only for the aggregated zones, which might be rather large.

### Trip Distribution for Microassignment

Within TAP, the ALDGRAV model produces trip tables that can be used for microassignment. In some instances, in order to attain sufficient precision in the microanalysis, analysis zones are very small, only a few blocks. There is, as yet, very little experience with the performance of ALDGRAV (and practically all trip distribution models) in such small-scale analysis. A careful study of this issue is much needed.

### Need and Justification for Precision

There are a number of areas in which certain increases in the complexity and costs of the analysis might make the results of the analysis more precise. Examples include

1. Making the relative weights of travel cost and time a function of income in impedance calculations,
2. Relating fixed impedance penalties to measurable zone attributes such as cost and availability of parking,
3. Further stratifying home-based non-work trips to short and long in order to attain better duplications of observed trip-length distributions, and
4. Using a number of paths rather than only one path for calculating impedance (the ALDGRAV theory, for example, suggests that both the minimum time path and the minimum cost path should be considered).

In spite of the long experience in travel forecasting, it seems that these issues have never been studied thoroughly. Various assertions, based primarily on so-called behavioral and theoretical considerations, on these subjects have been made; however, there is a need to study these issues by comparatively analyzing them with observed data, as well as by weighing the potential increase in the precision of the results versus increasing the cost of acquiring data and the complexity of the analysis.

## REFERENCES


1. Program Concept, Development of a Thoroughfare Plan and Inventory. North Central Texas Council of Governments, Arlington, Apr. 1975.
2. N. Nihan and D. G. Miller. The Subarea Focusing Concept for Trip Distribution in the Puget Sound Region. TRB, Transportation Research Record 610, 1976, pp. 37-43.
3. M. Schneider. Access and Land Development. HRB, Special Rept. 97, 1968, pp. 164-177.
4. M. Schneider. Direct Estimation of Traffic Volume at a Point. HRB, Highway Research Record 165, 1967, pp. 108-116.
5. Creighton, Hamburg. Transportation and Land Development—A Third Generation Model: Theory and Prototype. Federal Highway Administration, U.S. Department of Transportation, Rept. Contract No. FH-11-6792, 1969.
6. M. P. Kaplan. Calibration of the Access and Land Development (ALD) Model Travel Function: A Multimodal, Multidimensional Travel Function for Use in Urban Travel Demand Models. Department of Civil Engineering, Northwestern Univ., Evanston, IL, master's thesis, Aug. 1976.
7. A. G. Wilson. A Statistical Theory of Spatial Distribution Models. Transportation Research, Vol. 1, No. 3, 1967, pp. 253-269.

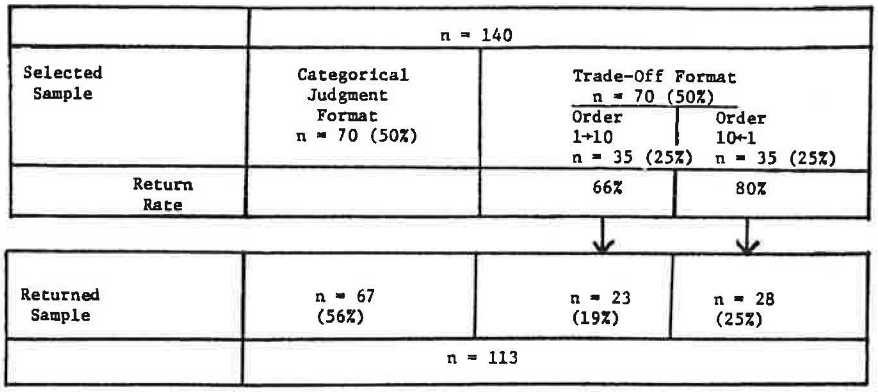

*Publication of this paper sponsored by Committee on Passenger Travel Demand Forecasting and Committee on Traveler Behavior and Values.*

# Recent Structural and Empirical Findings in Trade-Off Analysis

Patricia M. Eberts,* Kellogg Corporation, Battle Creek, Michigan
K.-W. Peter Koeppel,* Institute of Administration and Management, Union College, Schenectady, New York

**This paper reports on three recent investigations by the New York State Department of Transportation's Planning Research Unit into empirical and theoretical aspects of trade-off analysis, a multidimensional attitude scaling procedure. First, the possible influence of the length of the questionnaire was investigated. Fatigue bias was found to be substantial, and use of abbreviated questionnaires and a random order of items is suggested. Second, tests were made for a degradation in response accuracy, with substantially shortened questionnaires. No significant loss of information was found in reductions of up to 50 percent of a 10-matrix design. Third, the effects of different utility integration rules were studied. Some differences were found but they are too small to be of practical importance. The research concludes that the trade-off procedure is a powerful, robust approach that can be used with confidence.**

**Figure 1. Split-half partitioning of the sample.**




Numerous methods of attitude and opinion scaling have been used with increasing frequency by transportation researchers in recent years. A newly developed approach, trade-off analysis, has been used extensively by Market Facts (1, 2), and a similar algorithm was subsequently operationalized by the Planning Research Unit of the New York State Department of Transportation (NYSDOT) (3).

Using data about rank-order preferences of combinations of attribute levels, this procedure estimates the utilities each respondent places on the attributes describing a policy. In a second step, these individual utilities are then aggregated to estimate market shares for proposed objects or policies, defined in terms of these attributes. NYSDOT has used this method extensively in its studies of travel preferences and behavior (4, 5, 6).

However, due to the relative newness of the procedure, until recently little analysis has been conducted on the properties of the procedure itself. This paper summarizes a number of such studies by NYSDOT. Specifically, the following problems are dealt with:

1. Respondent fatigue because of the questionnaire length ("item position bias"),
2. Effects of reductions in the questionnaire size on the relative accuracy of the scaling procedure, and
3. Sensitivity of the procedure to various functional forms of the utility integration model.

The research summarized in this paper is described separately in more detail in other reports (7, 8).

## DATA

The data for these analyses were obtained from a study of white-collar employee attitudes toward alternative work schedules (5). A random sample of 140 employees of the main office of NYSDOT was administered a questionnaire on travel patterns, general attitudes toward work schedule changes, perceived impacts of these changes, and detailed attitudes toward five characteristics of work schedules. The five attributes and their levels are as follows:

1. Work week:
   - 4 days/week—Monday - Thursday,
   - 4 days/week—Tuesday - Friday, and
   - 5 days/week—Monday - Friday;
2. Hours per day:
   - 7 hours/day,
   - 8 hours/day, and
   - 9 hours/day;
3. Times worked:
   - Fixed (everyone starts and stops work at the same time),
   - Individual-specific (fixed for each person but allowing for differences between persons), and
   - Variable (start and stop work whenever the employee wants, subject to working a full schedule each day);
4. Parking location:
   - Unassigned spaces in assigned lots,
   - Special location for carpools, and
   - Assigned place in assigned lots; and
5. Cost of parking:
   - Free,
   - $1/month, and
   - $1/week.

The present working arrangement at NYSDOT's main office consists of (a) 5 days, Monday-Friday; (b) 7.5 hours/day; (c) fixed schedule; (d) unassigned parking in assigned lots; (e) free parking.

To investigate the structure of the trade-off model, a split-half approach was used to partition the sample, in which each respondent was randomly assigned one of these three versions of the questionnaire: categorical judgment format, trade-off format (10 matrixes in order 1→10), and trade-off format (10 matrixes in order 10→1). Figure 1 shows the distribution of the selected and returned trade-off sample (n = 51). Characteristics of respondents in each group were statistically similar to the population on three characteristics.

## ITEM POSITION BIAS

The questionnaire required by trade-off analysis is very lengthy, and thus we should expect to observe fatigue on the part of the respondents. Two different versions of the trade-off questionnaire, in which the order of matrixes was reversed, were administered to test this hypothesis.

The idea underlying this procedure is that, if an attribute included early in one questionnaire but rather late in the other questionnaire is changed, the result should be two different estimates of the preference shares for the same future. The difference in preference should be most pronounced for attributes presented at the extreme ends of the questionnaire and less pronounced for attributes presented in the middle of the questionnaire.

To determine the position of each attribute in the questionnaire, a ranking of from 1 to 10 was computed as below.

| Attribute | Original Rank (Group 2) | Rank Form Reversed (Group 3) | Difference in Rank |
|---|---|---|---|
| 1 | 2.5 | 8.5 | -6 |
| 2 | 4.75 | 6.25 | -1.5 |
| 3 | 6 | 5 | 1 |
| 4 | 6.75 | 4.25 | 1.5 |
| 5 | 7.5 | 3.5 | 4 |

Ten test policies were then constructed, changing one attribute level at a time. Differences in the predicted preferences for group 2 versus group 3 were plotted against average differences in rank for each attribute. Results are shown in Figure 2. The largest difference in the preference shares does indeed exist for attribute 1 (the attribute with the largest difference in rank), followed by attributes 2 and 3. Attributes 4 and 5, however, do not follow the hypothesis, probably because state workers are extremely sensitive to them and any change from the present status would be uniformly rejected.

Thus, the result of this study is that lengthy applications of trade-off analysis are likely to contain position bias and that appropriate measures (e.g., reversing order) should be taken to remove position bias. One additional method of avoiding such a bias, reducing questionnaire length, is examined in the next section.

**Table 1. Mean utilities for attribute levels under full and reduced data sets.**

| Attribute | Percentage of Data Remaining 100 | 60 | 50 Stepwise | 50 Circular | 40 |
|---|---|---|---|---|---|
| 1. Work week | | | | | |
| 4 days, M-Th | 0.3806 | 0.3862 | 0.3772 | 0.3798 | 0.3726 |
| 4 days, T-F | 0.3697 | 0.3747 | 0.3584 | 0.3693 | 0.3597 |
| 5 days, M-F | 0.2496 | 0.2390 | 0.2644 | 0.2509 | 0.2677 |
| 2. Hours per day | | | | | |
| 7 | 0.4321 | 0.4583 | 0.4199 | 0.4477 | 0.4470 |
| 8 | 0.5292 | 0.3275 | 0.3253 | 0.3254 | 0.3085 |
| 9 | 0.2387 | 0.2142 | 0.2548 | 0.2269 | 0.2446 |
| 3. Work schedule | | | | | |
| Fixed | 0.2917 | 0.2810 | 0.2857 | 0.2962 | 0.2916 |
| Specific | 0.3148 | 0.3140 | 0.3107 | 0.3151 | 0.3122 |
| Variable | 0.3935 | 0.4049 | 0.4036 | 0.3887 | 0.3962 |
| 4. Parking location | | | | | |
| Anywhere in lot | 0.3899 | 0.4025 | 0.3924 | 0.3949 | 0.4145 |
| Specific if carpool | 0.2743 | 0.2657 | 0.2723 | 0.2731 | 0.2548 |
| Assigned | 0.3358 | 0.3317 | 0.3353 | 0.3320 | 0.3306 |
| 5. Parking fee | | | | | |
| Free | 0.5823 | -[a] | 0.5636 | 0.5919 | 0.5632 |
| $1 per month | 0.2988 | -[a] | 0.2862 | 0.2799 | 0.2753 |
| $1 per week | 0.1189 | -[a] | 0.1503 | 0.1281 | 0.1614 |

[a]Since all matrixes relating to attribute 5 were excluded, no mean utility was calculated.

## REDUCTION IN QUESTIONNAIRE LENGTH

Not only could the effects of respondent fatigue be reduced by a shorter questionnaire (i.e., reduction of the number of matrixes presented), but substantial economic savings in survey administration and processing could result from such a reduction as well.

To test the sensitivity of the trade-off technique to data reduction, four reductions of 40, 50, 50, and 60 percent were performed by eliminating matrixes from the existing data set. A test policy was selected in which the time schedule (attribute 3) was changed from fixed to completely variable. The mean utilities obtained are given in Table 1. Overall, the differences in mean utilities between the full set of data and the reduced sets are small. Expectedly, this leads to the same predicted preference patterns for the full and all reduced data sets, as shown in Figure 3. In general, even a 50 percent reduction in questionnaire length does not lead to a significant loss of information.

A circular design in the selection of the attribute pairings is probably superior to other designs, in the absence of prior knowledge about the dominance of any attribute. The 60 percent reduction possibly does not lead to a significant loss of information, but its performance depends critically on the dominance of the pivotal element(s); in the absence of prior knowledge, such a design is not advisable.

## FUNCTIONAL FORMS OF THE MODEL

All the preceding tests were done by using a multiplicative utility integration rule:

Figure 2. Differences in policy preferences versus differences in attribute location.

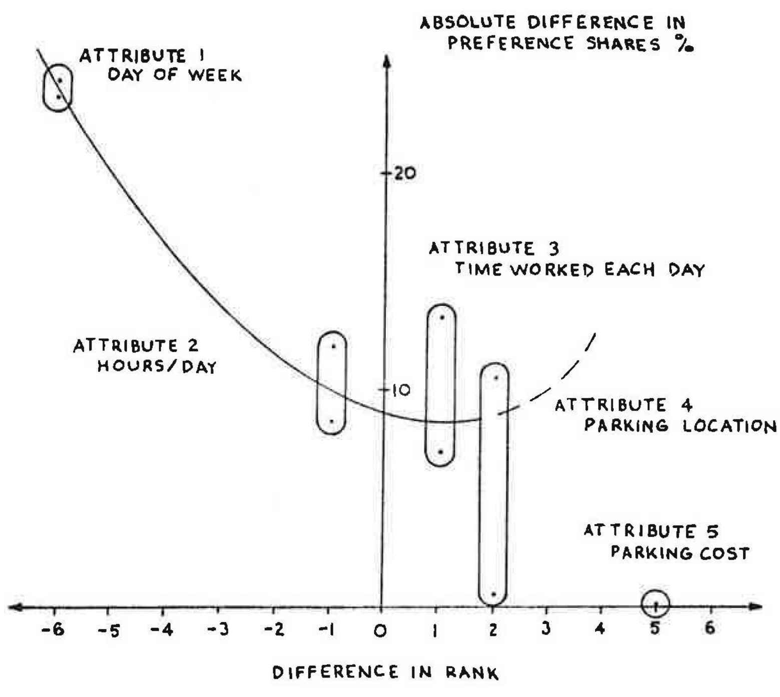


**Figure 3. Preference predictions under full and reduced data sets.**

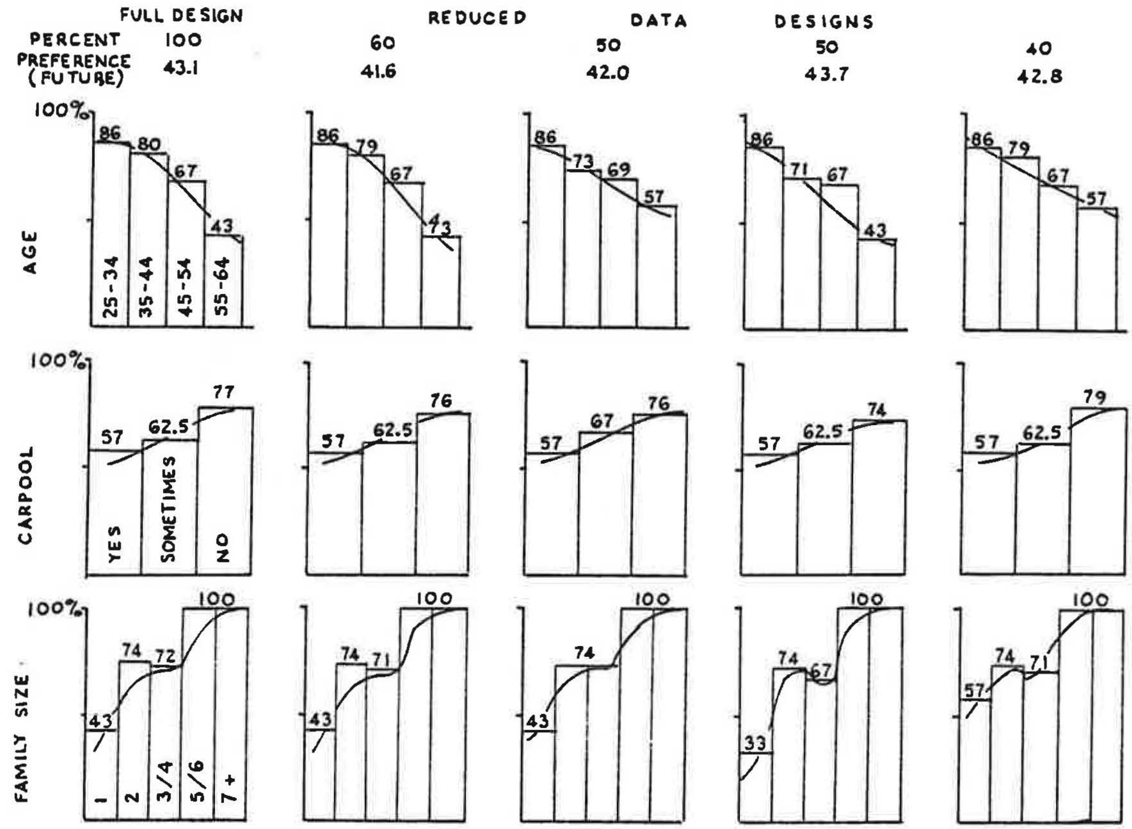



**Table 2. Mean utilities under different functional forms.**

| Attribute | Level | Utilities: Additive Mean | Utilities: Multiplicative Mean |
|---|---|---|---|
| 1. Work week | 4 days, M-Th | 0.3903 | 0.3806 |
| | 4 days, Tu-Fr | 0.3795 | 0.3697 |
| | 5 days, M-Fr | 0.2302 | 0.2496 |
| 2. Hours per day | 7 | 0.4458 | 0.4321 |
| | 8 | 0.3396 | 0.3292 |
| | 9 | 0.2147 | 0.2387 |
| 3. Work schedule | Fixed | 0.2865 | 0.2917 |
| | Specific | 0.3173 | 0.3148 |
| | Variable | 0.3966 | 0.3935 |
| 4. Parking location | Unassigned | 0.3950 | 0.3899 |
| | Prefer for carpool | 0.2675 | 0.2743 |
| | Assigned | 0.3376 | 0.3358 |
| 5. Parking fee | Free | 0.6203 | 0.5823 |
| | \$1 per month | 0.3094 | 0.2988 |
| | \$1 per week | 0.0703 | 0.1184 |

$$U_p = u_{p,1}(u_{p,2} \ldots u_{p,n}) \tag{1}$$

where

$U_p$ = utility of policy p and
$u_{p,i}$ = utility of attribute i as defined for policy p.

A number of objections to this model have been raised, the most important being the requirement of ratio scales, which are difficult to establish from the preference information in the questionnaire. Two other models, therefore, were compared to the multiplicative model: an additive model

$$U_p = u_{p,1} + u_{p,2} + \ldots + u_{p,n} \tag{2}$$

and the additive exponential model

$$U_p = \exp(u_{p,1} + u_{p,2} + \ldots u_{p,n}) \tag{3}$$

In each case, the probability of policy choice is then computed as a Luce share model:

$$\text{Prob}_P = U_P / \sum_q U_q \tag{4}$$

Obviously, these models are related by the exponential transformation, since

$$u_{p,1}(u_{p,2}) \ldots u_{p,n} = \exp(v_{p,1} + \ldots + v_{p,n}), \text{ with } u_{p,1} = \exp(v_{p,1}) \tag{5}$$

The additive exponential model takes the form of the multinomial logit model when combined with the Luce share model.

Since the exponential transportation is not admissible for ratio or interval scales, if we want to interpret these scales in other than an ordinal manner we can make the following predictions. First, utilities for attribute levels and policies would change slightly, comparing the three forms; second, market preferences based on first choices would not change at all; and, third, market preferences based on preference shares would change slightly.

In the tests done to confirm the above hypotheses, the computer program was modified to allow for the direct estimation of the additive utilities. These same additive utilities were then used to calculate preferences under the logit model, due to restrictions inherent in the programs. The results (Table 2) confirm the above hypotheses. The differences in the mean utilities are indeed so small as to be of no practical value.

For the evaluation of market preferences, the 5 days, variable hours future was used. The resulting market shares are given in Figure 4. As expected, there is no

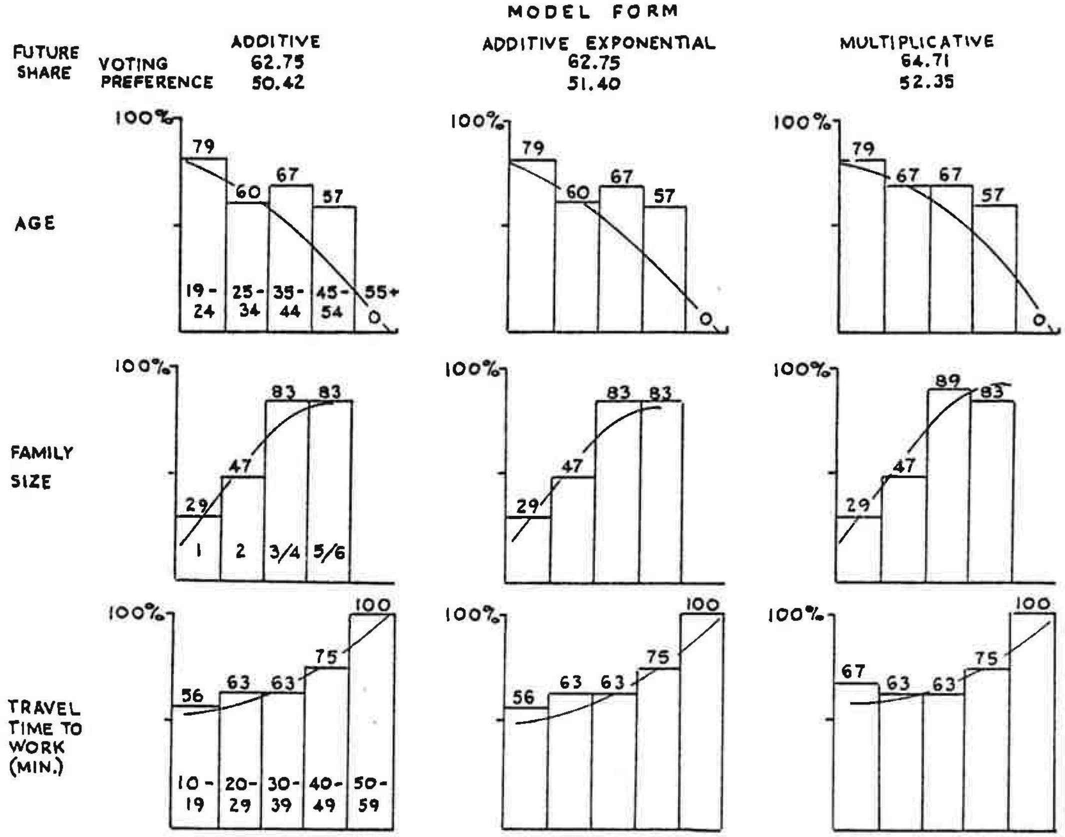



Figure 4. Preference predictions under different functional forms.

change in the overall voting share for the additive and logit models. In both cases, one person (out of 51) shifted his vote for the future when going from the additive-logit models to the multiplicative model. While we expected no shift at all, this shift does not upset the hypothesis, since the hypothesis is valid only for the case of a perfect fit of the utilities to the data and an infinite data set. Both of these conditions are not true in this (and any practical) application.

In the case of the preference share model, we find, as expected, a slight but insignificant shift. The logit model as applied to additive utilities stands between the additive and multiplicative model. From a demographic viewpoint, the three models are equivalent. If we were to select a model from a psychological point of view, we would tend to favor the additive model, which has been found to be an acceptable representation of many decision processes in other studies. However, from a measurement theoretical point of view, the additive model is to be preferred only for the purpose of a voting share prediction. It is not admissible for the calculation of a preference split unless it is subject to the exponential transformation, leading to the multinomial logit model. If we are interested in the preference split, this is the model with the best theoretical foundation.

On the individual level, the exponential additive model requires scales with known unit and one arbitrary origin per individual. On the aggregate level the unit of measurement does not have to be the same over individuals as long as we interpret the utilities as likelihoods. This represents a significant relaxation of the necessary conditions for the establishment of the utility structure compared to the additive and multiplicative models. Thus, the exponential additive model, which leads to the logit model when standardized, is generally the preferred utility integration rule for trade-off analysis.

## SUMMARY

To summarize, we found that the lengthy questionnaire needed for trade-off analysis is likely to lead to "position bias" in the data collected. Randomizing techniques would be needed in the composition of the questionnaires administered. While this may lead to larger variability in the data, it will reduce the effect of bias. We also found, however, that a 50 percent reduction in questionnaire length can be achieved without a significant loss in accuracy. A "circular" design is preferable. Last, for most practical purposes, all three utility models tested (additive, exponential additive, multiplicative) are equivalent. While the additive model is more commonly used, the exponential additive model is superior to the additive model on measurement-theoretical grounds.

The transportation policy planner should find trade-off analysis a helpful tool in the assessment of attitudes toward policy alternatives, and it has in fact been applied usefully in studies of staggered work hours, public transportation, and carpooling. The method is robust to reductions in questionnaire length and to various utility integration rules, and allows for an easy assessment of preferences broken down by demographics. It is hoped that such applications will be extended and further investigated.

## ACKNOWLEDGMENTS

We acknowledge John Fisk (State University of New York at Albany), Don Griesinger (Union College), and Dr. David T. Hartgen (NYSDOT) who facilitated the conduct of this research. Anis Tannir and Elene Donnelly (NYSDOT) provided data and computer assistance. The assistance of Wilma C. Marhafer and Barbara J. Blowers in preparing the manuscript is gratefully appreciated.

REFERENCES

1. R. M. Johnson. Trade-Off Analysis of Consumer Values. Journal of Marketing Research, Vol. 9, May 1974.
2. Peat, Marwick, Mitchell and Co. Carpooling Impact Study: Technical Memorandum III: Trade-Off Model and Policy Simulation. U.S. Department of Transportation, Feb. 1976.
3. E. P. Donnelly, S. M. Howe, and J. A. DesChamps. Trade-Off Analysis: Theory and Applications to Transportation Policy Planning. High Speed Ground Transportation Journal, Vol. 11, No. 1, Spring 1977, pp. 93-110.
4. E. P. Donnelly and others. Statewide Public Opinion Survey on Transit Operating Assistance: Technical Report. Planning Research Unit, New York State Department of Transportation, Preliminary Research Rept. No. 80, April 1975.
5. A. Tanir and D. T. Hartgen. Who Favors Alternative Work Schedules and Why? TRB, Transportation Research Record 677, 1978, in preparation.
6. E. P. Donnelly. Preference Elasticities of Transit Fare Increases and Decreases by Demographic Groups. TRB, Transportation Research Record 589, 1976, pp. 30-32.
7. P. M. Eberts. Trade-Off Versus Categorical Judgment: A Comparative Analysis of Two Attitude Scaling Methods for Transportation Planning. Planning Research Unit, New York State Department of Transportation, Preliminary Research Rept. No. 114, Feb. 1977.
8. K.-W. P. Koeppel. Functional Forms of the Utility Integration Rule in Trade-Off Analysis. Planning Research Unit, New York State Department of Transportation, Preliminary Research Rept. No. 123, June 1977.
9. W. S. Torgerson. Theory and Methods of Scaling. Wiley, New York, 1958.

*Publication of this paper sponsored by Committee on Passenger Travel Demand Forecasting and Committee on Traveler Behavior and Values.*

**This research was performed while the authors were student researchers at the Planning Research Unit, New York State Department of Transportation.*

# Air Passenger Distribution Model for a Multiterminal Airport System

Johannes G. Augustinus and Steve A. Demakopoulos, Port Authority of New York and New Jersey

**This paper reports on work aimed at calibrating the concepts of a theoretical air passenger airport distribution model with observations on actual passenger behavior as derived from inflight surveys. The original model, as developed for the U.S. Department of Transportation, has been modified to reflect more realistic passenger behavior patterns. Specifically, the simplistic assumption that passengers always select the most convenient airport regardless of the relative convenience (or inconvenience) of other available facilities or service has been replaced by a formulation that permits a more flexible distribution among facilities. The calibration of this modified distribution model with inflight survey data for the New York-New Jersey metropolitan area shows that model estimates that correspond closely with actual passenger distributions can be obtained, provided proper sensitivity coefficients are selected.**

In 1970, Peat, Marwick, Mitchell and Company, under contract to the U.S. Department of Transportation, developed a computerized intercity transportation effectiveness (ITE) model, of which a separate access-assignment (AAM) model deals with airport access problems, other factors related to airport choice such as congestion, and the potential role of sepecialized access systems such as off-airport satellite terminals (1).

The access-assignment model has two components: (a) a demand assignment model and (b) a cost benefit analysis model.

The following report discusses considerable expansions and modifications of the demand assignment model, developed by the Port Authority of New York and New Jersey under contract to the Tri-State Regional Planning Commission. Besides these technical expansions and modifications, the report also deals with the adaptation and application of the model to the Tri-State region. Finally, as its main focus, it discusses some results of the model's premises in terms of observations on actual air passenger behavior observed in Port Authority inflight surveys.

## GENERAL STRUCTURE OF THE ITE-AAM MODEL

This model attempts to simulate a transportation system in which passenger behavior and physical elements of the system interact. Such an interactive process is described by an iterative simulation in which one set of variables determines the level of another set in one phase (iteration), while the process is reversed in the next phase. For example, in the first iteration the passenger's airport choice is determined solely by convenience of access. Passenger volumes assigned on that basis then determine congestion levels at each of the airports in the system (aircraft, roadway, check-in delays) and frequency of flights at each facility. These convenience and inconvenience factors are then added to the access factors in redistributing passengers in the next iteration on the basis of total convenience, all expressed in monetary terms. The passengers for whom differences among facilities were marginal may change their choices from one iteration to another.

Total cost as conceived in the model includes all elements of cost incurred by the passenger from point of origination to aircraft take-off. These costs consist of out-of-pocket user costs as well as the cost of time involved in this process. Three such costs are centroid-oriented costs such as over-the-road access time and costs primarily physically (geographically) determined; nonvolume dependent costs such as parking fees, the fare of public transportation, and time lost in moving through the terminal, which do not depend on volume of (assigned) passengers, but are simply given at any point in time; and volume-dependent costs such as costs of congestion delay and schedule waiting time, which depend on passengers, vehicles, and flights assigned by the previous iteration.

The model as originally developed assigned passengers from each origin zone on a winner-take-all basis; i.e., all passengers from each zone were exclusively assigned to the one airport or satellite and airport combination that produced minimum cost to the passenger.

## ITE MODEL AND ACTUAL AIR PASSENGER ACCESS PATTERNS

Whereas the original model assigns passengers to facilities on the basis of a priori assumptions, the recurring inflight surveys conducted at the New York-New Jersey metropolitan airports by the Port Authority in cooperation with the airlines contain a wealth of completely empirical information on the passengers' airport choice in the real world, providing information on local origin, choice of airport, ground access mode, access travel time, destination of air trip, purpose of trip, etc. (2).

The obvious question presenting itself is whether the a priori asssumptions of the ITE model are confirmed by these empirical observations and, if not, how the model could be modified to incorporate the results of such observations in the real world.

The basic concept of the original model, that a passenger will always select the facility most convenient to him or her regardless of the relative degree of convenience as compared to alternate facilities (i.e., winner take all), is most likely an oversimplification of passenger behavior. The Port Authority inflight surveys show ample evidence that, when differences in convenience among alternate facilities are small, passengers distribute themselves among the available facilities rather than select exclusively the theoretically most convenient airport as determined by access congestion and schedule frequency. In the New York-New Jersey area this is particularly evident in the distribution of Manhattan passengers. As these account for more than a third of the region's traffic, this is obviously of major significance in any distribution model that is to have practical application in the Tri-State area.

## MODIFIED DISTRIBUTION FUNCTION

A more general model formulation that permits much more flexibility and presents many more opportunities for verification with and adaptation to empirical data is one similar to a model developed by the Rand Corporation in a 1967 study for the Port Authority (3). Adapted to the basic structure of the ITE model, this formulation says that, for each origin zone, the distribution among alternate airport facilities will follow the function

$$W_{ij} = (C'_{ij}/C_{ij})^{\alpha} \Bigg/ \left[\sum_{j=1}^{P} (C'_{ij}/C_{ij})^{\alpha}\right] \qquad (1)$$

where

$W_{ij}$ = fraction of passengers from centroid i who will select airport j,
i = area (centroid) = 1, . . . , C,
j = airport = 1, . . . , P,
$C_{ij}$ = cost for a passenger from centroid i to use airport j (roadway time, process time, waiting time, etc.),
$C'_{ij}$ = cost of the cheapest airport j for a passenger from centroid i, and
$\alpha$ = an index of passenger sensitivity with respect to cost differences among airports.

This model says in essence that the fraction W of total passengers originating in a particular centroid i who are to select a particular airport j is a function of the cost involved in using that particular airport versus the cost of using any of the other competing airports.

Although the particular functional relationship chosen is not necessarily the only one possible, it is clear that the relationship, as expressed in general terms, is logical; in a multiterminal situation, the passenger is assumed to be confronted with a choice among available airports and, in making a choice, will weigh the airports for relative convenience in a particular situation. The cost elements specified in the ITE model are obviously major components of this factor convenience.

Some more specific mathematical properties of this model are also appealing, as they further demonstrate the generality in the logic of the model as a mathematical description of passenger behavior.

In the first place, it satisfies the condition of conceptual logic that if passengers were infinitely sensitive to differences in access time, then they would always select the nearest airport. In this case, the coefficient $\alpha$ would approach infinity (or become infinitely large), and under that condition the value of $W_{ij}$ approaches zero for all airports except the nearest one, for which it approaches a value of one. This is the all-or-nothing or winner-take-all concept.

On the other extreme, if passengers were absolutely insensitive to differences in access times (if $\alpha$ were assumed to approach zero), the model would produce an equal distribution of passengers among the three airports, regardless of differences in access time.

The mathematical formulation of the model, finally, ensures that the sum of the individual shares of each airport for each particular centroid by definition always equals one, or, in other words, there never is an undistributed residual

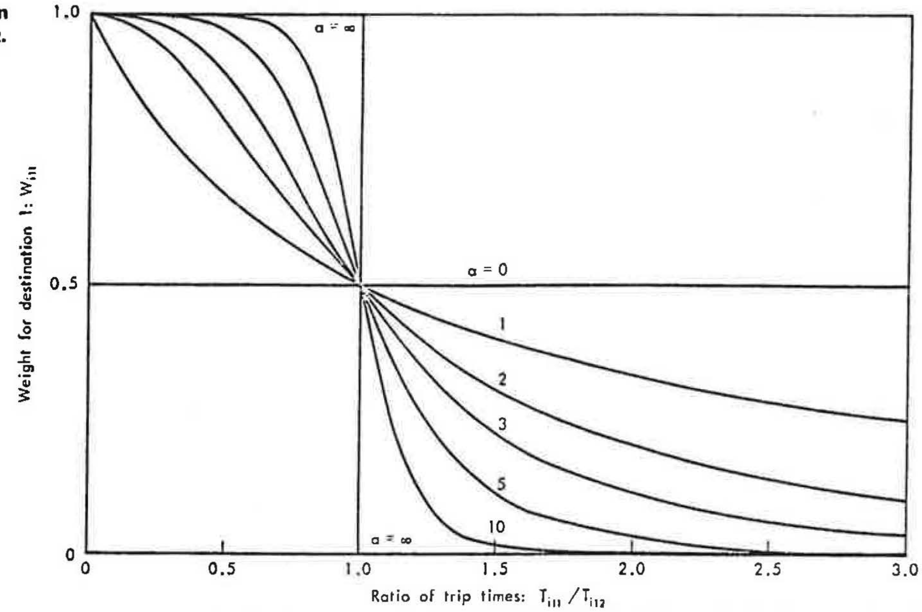
$$\sum_{j=1}^{P} W_{ij} = 1 \qquad (2)$$

Numerical examples illustrating these properties are given by Augustinus (4).

Although a passenger's sensitivity with respect to access time is probably high, it is not a priori infinitely high; therefore it is likely that in real life the coefficient will have a value somewhere between zero and infinity.

To provide an indication of the impact of changes in the value of $\alpha$ on passenger allocation estimates, Rand in their report produced a figure (Figure 1) showing the passenger distribution between two airports ($W_{ij}$) as a function of the ratio of access times to each airport ($T_{i1}/T_{i2}$). It is clear from this figure that, with increasing values of $\alpha$, except for the very low ones, the curve rapidly approaches the shape of the zero-one distribution that results from an assumed $\alpha = \infty$. Probably, this is also a fairly good reflection of real life because most of the passengers, in particular when access differences are significant, will select the nearest airport. If, however, major segments of the passenger market are located in centroids where the differences in convenience among available airports are

Figure 1. Behavior of the volume allocation weighting function for different values of $\alpha$.




marginal or at best small, as is the case in New York for many origin zones in the central business district, this assumption becomes clearly defective.

## MODEL CALIBRATION AND MEASUREMENTS

Having modified the model in such a way that it can reflect proportionate distributions of passengers among airports rather than only exclusive choices of one facility, a number of steps had to be taken to calibrate this modified model with actual survey data and to establish optimum values for the $\alpha$ coefficient.

### Aviation Planning Zones

For purposes of this study, the Tri-State region was broken down into 131 analysis zones (aviation planning zones). In addition, 11 zones bordering on but outside the region were included, as Port Authority inflight surveys indicate that these zones generate not insignificant numbers of passenger trips through the three metropolitan airports. The zone structure selected was primarily based on the availability of data (the smallest zone unit for passenger origin data was the zip-code designation), the trip-generating density, and the geography of the access network. With a total of 142 origin zones and three airports serving the region, the model had to deal with 142 $\times$ 3 = 426 origin-to-airport links.

### Trip Generation by Zone

The number of air passenger trips per aviation planning zone, rather than being computed from a theoretical trip generation model (as in the original ITE-AAM model), in this calibration study was determined from actual inflight survey data on passenger originations collected in 1972 (2).

As expected, the core area, and specifically Manhattan, is the largest traffic generator relative to its size, generating approximately one-third of the total locally generated trips. Moving out from the core area, trip generation density generally diminishes.

Besides passenger originations, the inflight surveys provide information on other items, such as trip destination, trip purpose, and residences of passengers. In the calibration study, the destination of the air trip (as represented by length-of-haul brackets) has been used to stratify the passenger market to determine differences in sensitivity to airport access convenience. Stratification with respect to other passenger characteristics that conceivably could reveal differences in access sensitivity, such as trip purpose (business versus personal air trips) or passenger residence (Tri-State region residents versus visitors to the region from elsewhere), have not yet been tested in this study.

### Zone Centroids

For each zone a geographic centroid was selected from which data on travel times and cost to airports and satellite terminals were measured. Centroids were selected on the basis of two factors: (a) traffic-generating density, defined by the areas of relatively high traffic generation within a zone, and (b) geographic location with respect to major highway intersections.

### Network Data: Access Times and Cost

For all practical purposes, airport access in 1972 was exclusively over highways. All travel times used in the calibration phase, therefore, reflect only highway times, by private car, taxi, or airport limousine. Times used are arithmetic averages of peak and off-peak travel times, as Port Authority airport statistics indicate that approximately one-half of the air passengers travel to and from the airports during highway peak hours and the other half during the off-peak hours.

The data do reflect today's congestion patterns in the region and thus do assume differing speeds over different highway segments.

As to the costs of access for 1972, the cost of using private automobile was computed on the basis of a cost of 4.0 cents/km (6.5 cents/mile), which included the cost of maintenance, tires, and gas, but no fixed cost such as depreciation, garage insurance, etc. This reflects the cost as presumably perceived by the passenger. Other (unrelated) Port Authority modal split studies produced the most realistic distributions between public and private modes when applying this concept for private automobile users. Taxi rates were computed on the basis of the then existing (1972) fare structure of 50 cents for the first 0.12 km (0.2 mile) plus 10 cents per additional 0.12 km, the structure in effect in New York City, but fairly representative for most other taxi fares in the metropolitan area. For Manhattan access cost, a taxi-private car mix of 70:30 was assumed, and a (reverse) 30:70 ratio for Brooklyn, Queens, the Bronx, and the nearby urban areas across the Hudson. Access costs to all other parts of the region were based exclusively on the use of private automobile. Highway, bridge, or tunnel tolls were added where applicable.

## Airline Schedules

Although airline schedules can be generated by the ITE model internally, in the calibration of the model with passenger survey data, airline schedules by distance range were fed into the model as they actually existed in 1972. This procedure should produce better measurements of passenger behavior, as actual schedules in the model simulate congestion and level of service conditions as actually experienced by passengers in making their selection of airport decisions during the survey period.

## Market Breakdown

In the calibration runs the total domestic passenger market was broken down in five markets by length of haul: under 300 km (250 miles), 300-800 km (250-499 miles), 800-1280 km (500-799 miles), 1280-2400 km (800-1499 miles), and over 2400 km (over 1500 miles). Such a breakdown was meaningful for these reasons.

First, from a theoretical point of view it is reasonable to postulate that short-haul passengers would be more sensitive with respect to access time and cost and more discriminating in their choice of airport than long-haul passengers. If confirmed, this should manifest itself in the values of the coefficients in the model that produce passenger distributions most corresponding to those observed in the passenger surveys.

Second, the level of service at each airport is not uniform in each market, partly for historical, partly for operational reasons (e.g., no transcontinental service at La Guardia). Thus, empirical measurements made separately for each market permit changes in service patterns in the future if indicated by expected technological developments or plans for airport expansion.

## Calibration Results

After feeding the input data as described into the ITE model, the distribution of the air passengers in each market among the three airports, as simulated by the model under varying values of $\alpha$, was calibrated against the actual distributions observed in the 1972 inflight survey. The results are shown in Table 1.

The model estimates appear to reflect actual distributions fairly well for any of the selected $\alpha$ values, which indicates a basic soundness of the logic of the model as a simulation of actual behavior.

It is also evident from the table that certain $\alpha$ values produce numerically better results than others. This supports the original premise of the study that, once a proper theoretical framework (model) has been developed, passengers' sensitivities to convenience differences can be estimated from actual survey observations.

It should be emphasized that there is no a priori connection between the model estimates and survey observations other than the common base of passenger originations (the origin numbers used as input in the model were taken from the survey). The airport distributions developed by the ITE model are generated through the model's internal logic. The survey data show which airports were actually selected by the passengers.

Although no overall measure of goodness of fit has been incorporated into the model at this time, the results strongly indicate that very high values of $\alpha$ (most closely corresponding with $\alpha = \infty$; i.e., the all-or-nothing hypothesis) generally produce less realistic results than $\alpha$ values in the general range of 5-15. Where deviations of some significance occur with respect to an individual airport, a more detailed analysis of the underlying market structure (business versus personal travelers, residents of the region versus visitors) might well explain some of such deviations and enable us to further refine the measurements and reduce the differences.

Although some irregularities occur, the data do further indicate that in the lower distance ranges the best-fitting results are produced by higher values of $\alpha$, while in the longer ranges lower values of $\alpha$ produce better results. This confirms the a priori expectation that access and convenience factors are more important to short-haul than to long-haul travelers, as they account for a relatively larger share of the time and cost of the total trip.

It may be mentioned that, in deriving this conclusion, less weight was given to the longest range, where service at Kennedy and Newark in 1972 definitely favored Kennedy. Equal service levels might well have reduced actual passenger levels at Kennedy to the benefit of Newark.

Table 1. ITE model estimates versus actual regional totals.

| | | | Market | | | | |
|---|---|---|---|---|---|---|---|
| Item | α-Value | Terminal Location | Under 300 km | 300-800 km | 800-1280 km | 1280-2400 km | Over 2400 km |
| Actual | | LGA | 189.0 | 167.3 | 211.2 | 232.5 | 27.4 |
| | | JFK | 20.6 | 27.4 | 18.7 | 205.8 | 182.7 |
| | | NWK | 90.3 | 81.7 | 95.1 | 120.4 | 45.4 |
| Model estimates | 25 | LGA | 200.1 | 205.0 | 221.7 | 246.7 | - |
| | | JKF | 18.7 | 10.0 | 23.6 | 200.1 | 210.3 |
| | | NWK | 81.0 | 61.9 | 80.0 | 104.5 | 45.0 |
| | 15 | LGA | 190.6 | 195.3 | 214.8 | 241.4 | - |
| | | JFK | 30.5 | 20.2 | 31.1 | 219.1 | 208.3 |
| | | NWK | 79.1 | 61.4 | 79.6 | 98.6 | 46.9 |
| | 10 | LGA | 176.4 | 181.0 | 202.6 | 233.8 | - |
| | | JFK | 45.0 | 33.8 | 43.2 | 229.5 | 203.3 |
| | | NWK | 78.2 | 61.8 | 79.3 | 95.9 | 52.1 |
| | 5 | LGA | 145.8 | 147.9 | 171.2 | 223.2 | - |
| | | JFK | 73.0 | 60.2 | 70.8 | 231.0 | 186.3 |
| | | NWK | 81.0 | 68.6 | 84.0 | 105.1 | 69.0 |

Note: 1 km = 0.62 mile.

Table 2. ITE model estimates versus actual Manhattan only.

| Item | α-Value | Terminal Location | Market: Under 300 km | 300-800 km | 800-1280 km | 1280-2400 km | Over 2400 km |
|---|---|---|---|---|---|---|---|
| Actual | | LGA | 150.4 | 118.7 | 123.6 | 100.8 | 13.5 |
| | | JFK | 16.4 | 19.2 | 10.7 | 60.1 | 65.6 |
| | | NWK | 15.1 | 21.2 | 17.3 | 12.8 | 2.8 |
| Model estimates | 25 | LGA | 175.7 | 156.2 | 148.C | 173.7 | - |
| | | JFK | 1.3 | 0.4 | 0.2 | 3.8 | 77.4 |
| | | NWK | 4.8 | 2.2 | 3.5 | 3.9 | 4.4 |
| | 15 | LGA | 165.6 | 150.0 | 143.1 | 159.1 | - |
| | | JFK | 8.4 | 3.9 | 2.4 | 15.1 | 73.8 |
| | | NWK | 7.9 | 4.8 | 6.1 | 7.1 | 8.0 |
| | 10 | LGA | 148.8 | 137.4 | 132.3 | 140.5 | - |
| | | JFK | 19.8 | 11.7 | 8.4 | 28.6 | 68.8 |
| | | NWK | 13.2 | 9.6 | 10.9 | 12.3 | 13.0 |
| | 5 | LGA | 113.8 | 105.5 | 102.4 | 107.7 | - |
| | | JFK | 40.1 | 29.6 | 24.9 | 47.1 | 58.1 |
| | | NWK | 28.0 | 23.5 | 24.4 | 26.5 | 23.1 |

Note: 1 km = 0.62 mile.

Table 2 shows the model results versus actual for Manhattan only, which accounts for more than one-third of total regional traffic generation in all distance ranges. Another reason why the model's performance in Manhattan has special significance is that for many zones the differences in accessibility to the airports are relatively small and thus the model estimates are more sensitive to the value of $\alpha$ to be selected. If differences in convenience among airports are large, differences in assumed values of $\alpha$ do not produce significant differences in model estimates.

The optimum values for $\alpha$ appear to be here somewhat lower than in the total regional numbers and generally fall in the 5-10 range. The declining trend as a function of length of haul is also here much in evidence. Recognizing that the actual observations as summarized are subject to sample fluctuations and, moreover, that actual behavior may reflect factors not accounted for in the model, it may be postulated that the $\alpha$ values basically could be represented by a linearly declining curve as a function of length of haul.

The Port Authority report to the Tri-State Regional Planning Commission also included some examples of how a model, as developed here, could be applied in estimating the traffic potential for a couple of off-airport satellite terminals.

The main objective of this paper, however, was to report on some results of the development of a passenger distribution model that could be calibrated against survey data on actual passenger choice patterns. We hope such attempts to merge theoretical model concepts with empirical data on actual passenger behavior will contribute toward the development of more realistic demand forecasting tools for use in transportation planning.

ACKNOWLEDGMENTS

This report is based upon work performed by the Port Authority of New York and New Jersey for the Tri-State Regional Planning Commission and was financed in part by a planning grant from the Federal Aviation Administration. The views expressed are those of the authors and do not necessarily reflect the official views or policies of the sponsoring organizations.

REFERENCES

1. Peat, Marwick, Mitchell and Co. Intercity Transportation Effectiveness Access/Assignment Model. U.S. Department of Transportation, Rept. No. DOT-OS-00027/16, Nov. 1971.
2. New York's Domestic Airpassengers, 1972. Aviation Economics Division, Port Authority of New York and New Jersey, 1972.
3. D. M. Landi and A. J. Rolfe. A Model and Computer Code for Studying Alternative Air Passenger Processing Strategies. Rand Memorandum RM-5818-PA, Aug. 1969.
4. J. G. Augustinus. An Air Passenger Distribution Model for the New York-New Jersey Area. In Airport Economic Planning (George P. Howard, ed.), Massachusetts Institute of Technology Press, Cambridge, 1974, pp. 193-209.

*Publication of this paper sponsored by Committee on Aviation Demand Forecasting.*

# System for Planning Local Air Service

Maximilian M. Etschmaier, Department of Industrial Engineering, Systems Management Engineering, and Operations Research, University of Pittsburgh

**Local air service is characterized by a strong sensitivity of traffic to a number of factors such as frequency, time of departure, trip time, and alternative transportation available by ground modes. Consequently, in planning local air service, the demand function and the scheduling constraints must be considered in more detail than is necessary with other types of air transportation. This paper presents a system developed at the University of Pittsburgh for planning local air service. The system was used in studying the potential of air service between the provincial capitals of Austria.**

The motivation for this study was an overall regional development plan drawn up by the province of Steiermark in Austria. Although Austria is a federal republic of nine provinces, each with its own government of considerable power, there is a strong trend toward centralization in Vienna. Ever since the collapse of the Austro-Hungarian monarchy, Vienna has been too large a city for a country the size of Austria. The population of Vienna is now more than a quarter of that of all Austria. This fact, together with the gravitational force of the seat of the federal government, creates a strong momentum toward centralization. Already most major corporations are headquartered in Vienna; the bulk of federally subsidized cultural activity is also located there; and Vienna has the only airport in Austria that ever had more than three scheduled daily departures.

The trend toward Vienna is aggravated for the province of Steiermark, for which holding its own would be enough. This province is faced with certain structural deficiencies. On the one hand there is sizable heavy industry with a long-standing tradition and high technological standards. This industry's existence is being threatened by today's economy. On the other hand, there are numerous farms too small to support a family at today's level of expectations. For these reasons the province has to attract settlement of new businesses and industry.

To this end, government is trying to make the province more attractive to the business community. Much has already been done in this direction. Numerous tax incentive plans and subsidies have been devised. Graz, the capital, has acquired a reputation as a center for avant-garde cultural activity. Excellent recreational facilities were developed throughout the province, which is host to a number of significant international sporting events. The availability of good transportation facilities is viewed as an important factor in its attractiveness.

The government therefore commissioned studies to examine the railway, highway, and air transportation systems. The intention was not so much to determine the long-range effects that improved transportation facilities would have on the economic development of the province, but to determine the short-range economic feasibility of improved transportation facilities. In the case of air transportation this meant determining whether air service attractive enough to justify subsidy could be designed. There were no firm limitations imposed on the amount of subsidy that could be expected, but clearly any subsidy would have to be reasonable in relation to the anticipated patronage.

Air service was expected to perform the following functions: (a) connect the province with the other provinces of Austria, (b) connect the province to large population centers outside Austria, and (c) tie the province into the international air transportation system. Special emphasis of the study was to be placed on the interaction between air and surface transportation, in particular, how major highway construction projects would affect the demand for air transportation. A new highway system is under construction and, although it will not be completed for quite some time, the development of an air transportation system could not be justified if it were to be threatened with extinction after completion of the highway system.

Our study was complicated by a number of factors, the most important ones being that, except for a brief episode, there was practically no history of local air service in the area. Furthermore, the stage length of most potential routes is quite short, which makes them highly vulnerable to competition from ground transportation. Also, the area is quite densely populated, so a number of different cities have to be considered as potential users of any one airport.

Of course, many of these factors are quite typical for the environment in which local air service operates anywhere. We have thus supplemented the funds provided by Steiermark with funds from the University of Pittsburgh and expanded the study to a far broader scope than would have been justified within the context of the Austrian situation alone. The result is a general system for planning of local air service (PLATO) that can provide a framework under which any local air service, operating over reasonably short routes, can be viewed.

Of course, as any demand function would, the demand function included in PLATO reflects the sociopolitical and cultural background of central Europe, the area for which it was developed. In any area in which the relations among factors differ it would be necessary to reformulate the demand function. Any demand function can only express the most important dependencies and incorporate only a small selection of variables. The dependencies not included are expected not to vary significantly within the framework of the study. As one moves to another region, one can expect changes in the importance of different dependencies and variables. Also, certain variables may simply not be available for other areas.

The PLATO system, nevertheless, can aid a study for another region in two ways. First, it can help in the development of a demand function by handling the massive socioeconomic and geographic data. Second, once a new demand function has been developed, that function can be entered into PLATO, and traffic estimates, revenues, and costs for air service can be obtained without difficulty.

## SETTING FOR LOCAL AIR SERVICE

The type of environment PLATO is primarily designed for is characterized by a large number of relatively small cities scattered over a comparatively small geographic area. The area is served by a small number of local airports. The potential demand for air transportation comes both from people moving between one city and any other city in the region and from people moving from one city to a point outside the region.

The airports are located in such a way that passengers from some cities might use two or more different airports, depending on their final destination. The region is served comparatively well by surface transportation on highways and railways. Surface transportation does provide an alternative to air transportation and, for many, can actually beat air transportation in terms of total travel time between cities.

In this kind of environment air transportation can only compete effectively if it can meet the true transportation needs of the traveling public. This means it has to provide service at the right times, with high enough frequency so that there is a big enough chance that some transportation need can be satisfied, and at a cost differential against surface transportation that would be commensurate with the time saved. As a consequence, the traffic that can be attracted by air transportation will be highly sensitive to schedules and air fares, and it is generally meaningless to develop plans for local air service that do not express this kind of sensitivity.

On the other hand, the cost function for providing local air service is considerably less favorable than it is for interregional air service. The reasons for this are the short stage lengths, the small size, and the relatively low operating efficiency of the aircraft. This leaves a narrow margin for local air service to operate in. And, although much of the local air service enjoys healthy public subsidies of one form or another, it is generally only profitable for small companies that can avoid the pressures of strong national unions and can operate under comparatively low overhead costs.

Planning local air service thus means finding a delicate balance between the preferences of the traveling public and the dictates of the cost function. The public refuses to fly at other than the most desirable departure times: the high aircraft utilization mandated by the cost function forces one to schedule flights throughout the day. The problem can of course be solved best if differences in the time-of-day preferences for travelers on different routes can be matched.

## AVAILABLE DATA

As the air service offered in Steiermark was only marginal, limited traffic data were available. A summary of the traffic originating in Graz in 1972 is shown in Figure 1. At that time, the flights offered in the morning were Graz-Salzburg-Zurich and Graz-Linz-Frankfurt; at midday were Graz-Vienna, Vienna-Graz and Zurich-Salzburg-Graz; and in the evening were Frankfurt-Linz-Graz. All flights were with DC-9 aircraft. Traffic data for the Graz-Vienna flight were not available.

Also in 1971 airport passenger surveys were conducted in all Austrian and German airports. For Graz this survey was conducted in 1972. In these surveys a sample of departing passengers were asked questions about, among other things, the starting points and final destinations of their journeys and how they got to the airports. Figure 2 shows an analysis of the passengers departing from Graz. The numbers in the figure are extrapolated average daily passengers. Since the survey did not include the routing of a passenger, the results cannot be compared directly with the traffic data of Figure 1. Exceptions would be the flights to Linz and Salzburg, for which the results are reasonably close.

It was expected that a considerable number of passengers went to Vienna by surface transportation and started their flight there. These passengers could be extracted from the Vienna airport passenger survey. It was found that an average of 38 passengers a day were traveling from Steiermark by surface to Vienna to board a flight there. This figure is only a lower bound, since it does not include passengers who spent a night in Vienna before continuing their trips.

Also available was a market survey done by an Austrian institute (Khoulhavi) in which businesses and households were asked how many trips a year they would expect to make to different destinations if air service were available. The questions that must be raised against this survey are the same for any survey of the demand for an essentially new mode of transportation. The prospective passengers cannot possibly be given a full impression of what it would actually mean to use the proposed mode. Consequently, they may use a completely unrealistic perception of the proposed mode to base their estimates on. Also, it is quite difficult for one to estimate how often he or she would travel to some city, since one cannot always distinguish between desire and reality. The results of this survey indicate quite clearly that not all of these problems could have been dealt with successfully. Therefore, the survey appeared to be of limited value within the context of this study and was not used.

Clearly the available data are not sufficient for developing a demand model. We therefore selected a cross-sectional approach with models developed for comparable environments and calibrated for the specific situation. The comparable situation was found in

Figure 1. Average passengers per day, 1972.

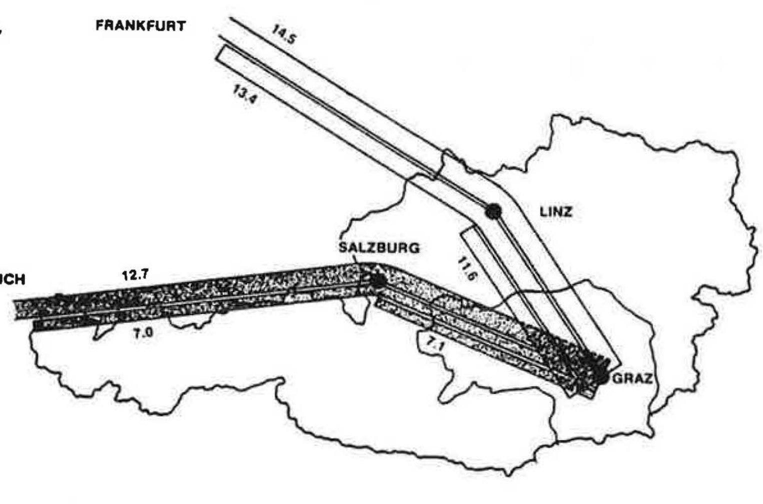
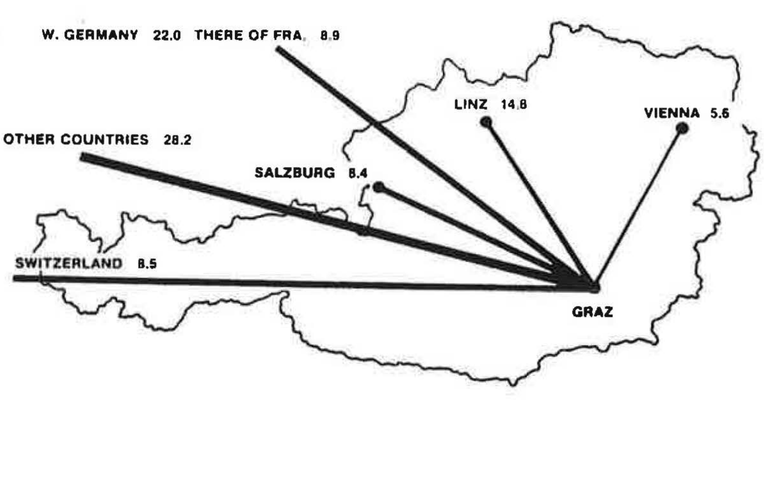



Figure 2. Passenger survey of 1972.



Germany, where considerable data as well as past models were available. The air traffic data were taken from the surveys for most German airports. Additionally, socioeconomic and geographic data about all German districts were gathered.

A model of the demand for local air service was developed by Intertraffic (1) in the study for North Rhine-Westphalia. This model was further refined by Nüsser (2), who applied it in a study of local air service for all Germany. We adopted much of the functional form of these models.

The need for change arose from the fact that some required socioeconomic variables were not available for Austria. Also, we went into considerably greater detail in the representation of the surface transportation alternatives. To accomplish this, the demand had to be determined from any district to any other district, with one airport possibly serving many districts. This led to considerable computer work and required the development of new procedures and program organizations. Finally, some modifications were made to the time-of-day variability of demand. Details of these modifications will be given in the discussion of PLATO.

## STRUCTURE OF PLATO

As mentioned before, the central problem in planning local air service is one of developing a schedule for which the expected revenue matches or exceeds the expected total cost. In solving this problem for one particular company, one might use any one of a number of objective functions such as return on investment and total profit one wants to maximize. If one is solving the problem for a local government, one would be interested in maximizing the air service—possibly favoring certain airports—that can be economically justified. Often this means that the subsidies required to operate some schedules have to be reasonably related to the amount of service offered by the schedule. Quite obviously the considerations entering the evaluation of a schedule are almost impossible to formulate into a mathematical objective function. The problem really is not so much one of optimization, but of providing the government with a number of reasonable choices.

Of course it would be desirable to have an algorithm that could automatically generate a number of "most reasonable" choices, but the developmental effort, computer run times to actually perform the optimization, as well as the information required to formulate the problem, might prove prohibitive. Even the optimization simply in terms of a single-valued company objective might prove to be such an enormous task that it could actually be carried out only for very small systems. For such systems, however, good schedules might be arrived at fairly easily by hand. What seems more important, therefore, than a closed form optimization method is a system that can help a human planner to set up a schedule and that can quickly evaluate the schedule in terms of revenue and cost.

Setting up a schedule involves arranging and rearranging flights in different strings, determining departure and arrival times for all flights, and just putting the whole schedule on paper in a readable form. The evaluation of the schedule in terms of cost and revenue involves large numbers of tedious calculations for each flight. Following the pattern of most airline planning systems (3), PLATO does not perform any optimization, but concentrates on schedule editing and evaluation.

An overview of PLATO is given in Figure 3. By far the greatest part of the system is taken up by the generation of the demand function. The demand for air transportation is determined by city pair on the basis of a wealth of socioeconomic and geographic data. In this process the competition from all surface transportation modes is considered explicitly. The city pair demand functions are then assigned to the relevant airport pairs to form airport-pair demand functions. The virtue of the approach of PLATO is that all geographic and socioeconomic data are processed in the modules for local and interregional route generation and demand generation. These modules require large core space and considerable computer time, but they are used only once for any given situation. For each airport pair they arrive at two functions in which the traffic can be calculated by inserting schedule variables. This minute number of functions is then used to evaluate as many schedules as desired with a minimum amount of computer time. A special feature in the system permits the evaluation of one schedule for two different years at the same time.

## ROUTE GENERATION MODULE

The route generation module preprocesses some of the input required for the local demand generation module. It takes as input the complete transportation network of the region. Highway, railway, and potential air route networks are entered separately in the forms in which they can be most easily obtained. The program converts them into the form required for the calculations.

Common to all three networks are nodes to represent the cities of the region. Other nodes are defined to represent junctions or changes in the travel speed. A number of auxiliary nodes may be defined to represent times spent to get through airport terminals or railway stations or to make connections. An example for the nodes and arcs that might be associated with one city is given in Figure 4. Figure 5 gives a schematic overview over the route generation module. The algorithm used to find the fastest routes is the one by Floyd (4).

## DEMAND GENERATION MODULE

PLATO differentiates between two different kinds of demand for air transportation, namely, purely local demand—the trip originates and terminates within the region—and interregional demand—the trip originates in the region and terminates outside or vice versa. Each kind of demand is determined in a separate module. The essence of these modules is that they take the large volume of socioeconomic and geographic data and reduce them to a very small size. This permits the succeeding schedule evaluation module to work with only modest core requirements. Overviews of the two modules are given in Figures 6 and 7.

The total interregional demand that originates in a district is obtained as a function of socioeconomic and geographic parameters of that district. The model was developed using traffic data for 130 German districts obtained from the airport passenger survey of 1971. Numerous model assumptions were tested. The model that yielded the best fit had the following form:

$$Y = a \times PD^{b} GP^{c} CP^{d} [(EP - EMP)/EP]^{e} + f \times AD^{g} B^{h} \times (1/P) \quad (1)$$

where

$\hat{Y}$ = estimated number of annual interregional passengers per 1000 inhabitants,
PD = population density (inhabitants per square kilometer),
GP = gross domestic product per 1000 inhabitants,
CP = number of businesses per 1000 inhabitants,
EP = number of employees per 1000 inhabitants,

Figure 3. PLATO overview.

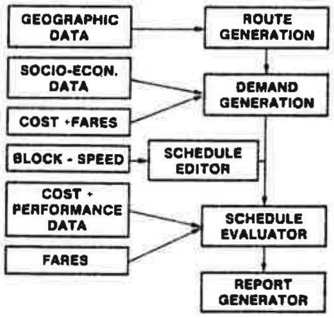



Figure 4. Nodes and arcs for Vienna.

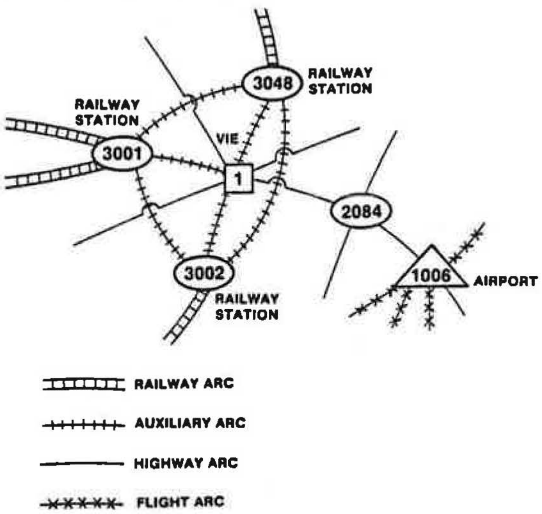
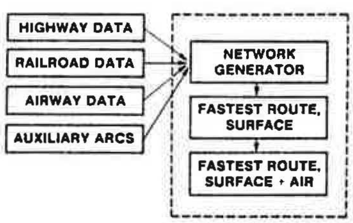



Figure 5. Route generation module.




EMP = number of employees in manufacturing per 1000 inhabitants,

AD = aircraft movements at the nearest large airport, divided by the square of the distance to that airport (a value of 10 km was used if the actual distance would have been less than that),

B = number of hotel beds, and

P = population in 1000s.

The rationale for this model formulation developed from two reasons for the transportation activity of a district. The first is the economic power of the region itself; this is represented by the first part of the equation. The second reason is the existence of transportation facilities. Although this reason cannot be separated completely from the indigenous traffic, the relationships, which govern this kind of traffic, are clearly different. The kind of traffic generated by the availability of transportation facilities would have not only the form of vacation tourism but probably, more importantly, the form of what one might call business tourism.

Business meetings of all kinds are held at places that are easily accessible. In the long run, the avail-

Figure 6. Demand generation module.

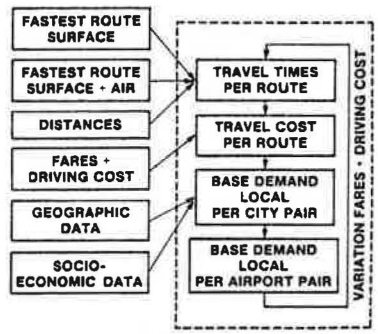
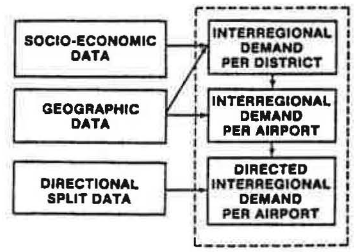



Figure 7. Interregional demand generation module.




ability of transportation facilities tends to change the structure of business located in a district, encouraging settlement of businesses that are either origins or destinations of travel. The second part of our equation gives a modest account of this dependence.

Although overall the model provided a good fit, there is one difficulty, which eventually led us to abandon it in the Austrian study. The specific number of interregional passengers per district varies widely, ranging from a high of 1890 passengers per 1000 inhabitants for Frankfurt to a low of 41 for Gelsenkirchen. All districts in Steiermark are definitely at the low end of this range, where the variance would introduce a considerable uncertainty into the estimate. Thus the confidence in the model is no greater than the confidence in simple extrapolations from past traffic data and estimates. Therefore, in the Austrian study, such estimates, scaled up according to growth projections for local demand, were used in place of the model.

The total interregional demand that originates in a district is initially determined in undirected form. This demand is then split up according to destinations of travel prevailing in the area. In the Austrian study a total of thirteen destination regions was used. Initially, all interregional demand originating in a district is assigned to the nearest airport. How much of that demand actually uses that airport and how much travels by surface directly to the nearest hub is a function of the schedule offered and is determined in the schedule evaluation module.

For the purely local demand a multistage model is used. The local traffic attracted by some flight segment is assumed to be dependent on some base demand for that airport pair, the departure time, and the frequency with which the airport pair is served. The functional relationship is assumed to have the form

$$\text{Demand} = XLD \times SERLEV \times [LD/F1(NF)] \qquad (2)$$

where

Demand = local traffic attracted by some flight,

XLD = the base demand,

SERLEV = service level of the route as a function of the frequency offered,

LD = value of the time-of-day preference, and
F1(NF) = maximum possible sum of the time-of-day preferences for the given frequency NF.

The base demand XLD between an airport pair is obtained as the sum of the base demands CLD between those city pairs that use the given airport pair. CLD for a given city pair i, j is given by the following function:

$$CLD = FI \times DT^{\alpha} \{ [f(B_i B_j) + f(D_i D_j)] / [f(DP) E_{ij}^{\beta} A_{ij}] \} \quad (3)$$

where

FI = calibration coefficient,
DT, DP = differences in travel time and price respectively for shortest route for surface transportation alone and shortest route including air transportation,
$B_i$ = population of city i,
$D_i$ = gross domestic product of city i,
$E_{ij}$ = straight line distance between i and j, and
$A_{ij}$ = mean of airport access and egress times for cities i and j.

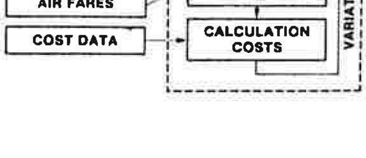
CLD is set to zero if the time gained by using air transportation is so small that it would not justify the increased cost. It is assumed that time has a certain value for each traveler and that he or she would be willing to pay only up to a certain amount for every hour saved. The cut-off value is a variable that can be set by the user.

As was mentioned previously, the basis for these models is work done by Intertraffic (1) for North Rhine-Westphalia and by Nüsser (2). The function for the base demand is taken almost directly from Intertraffic. The major modification was necessary because no detailed balance of economic accounts is available for Austrian districts. For this reason, the gross domestic product had to be substituted in place of the gross domestic product for services required by the Intertraffic model. The most adverse effect of this substitution was eliminated by scaling and by means of a calibration coefficient.

The function Demand is a device used to consider some criticism that was already leveled by Nüsser against the Intertraffic model. The time-of-day dependence does not appear pronounced enough for the very short distances of interest here. Nüsser derived a more appropriate function, but he correctly pointed out that incorporation in the Intertraffic model would lead to changed coefficients. We tried to get around this problem by separating the frequency dependence of demand from the time-of-day dependence.

Thus we determined for each frequency the maximum number of passengers who can be attracted under optimum departure times. This is given by XLD × SERLEV,

Figure 8. Schedule evaluation module.




with SERLEV derived directly from the time-of-day function of Intertraffic. Then the actual traffic on a flight is determined using the actual weights under Nüsser's function relative to the maximum values possible under the given frequency. Certainly this procedure would be open to criticism. Also, good judgment has to be used in its application. As we will show, however, the results obtained thus far have been quite convincing.

The demand generation module only calculates XLD for the local demand. The calculation of Demand is left to the schedule evaluation model. The demand generation module is set up in such a way that results for different functions for the cost of surface and air transportation can be used in one run. This permits the user to perform a certain amount of sensitivity analyses with a minimum effort.

## SCHEDULE EVALUATION MODULE

This module performs a highly efficient computation of the total cost and revenue associated with a given schedule. By producing results for a number of variations of some parameters in one run it facilitates a sensitivity analysis of the results.

The basic structure of the module is given in Figure 8. The module consists essentially of four more or less independent parts:

1. Processing interregional demand,
2. Processing local demand,
3. Considering aircraft capacity, and
4. Calculating cost and revenues.

All data entered in this module have been reduced to a form beyond which they could not be reduced without the knowledge of the flight schedule actually offered. The reason, of course, is that, while the proceding modules will be used only once for any given problem, one will want to evaluate many different schedules. Minimizing the work to be done within this module therefore means minimizing the total computational effort required for PLATO. It also means that this module can be operated from a terminal with fast run and response times.

The maximum interregional demand $XJD_{ij}$ from airport i to region j is obtained as output of the interregional demand generation module. If airport i is not itself a hub, then this demand may be assigned to flights from i to any hub, provided they permit reasonably good connections. To permit this check all flights from the possible hubs to the different regions must be entered into the program. This is done in the form of a list that gives the number of non-, one-, and multistop flights to and from each region for each hub for every hour. The user must also enter a maximum value CT1 of the connect time accepted in each hub. Then for each local airport i and for each destination region j the following procedure is followed.

For each possible hub k all connections i-k-j are identified that use a connect time of less than CT1. If one or more connections can be found, then $XJD_{ij}$ is split between them in proportion to their value 1/(number of stops × distance i-k-j). If the shortest connection available requires a connect time TC > CT1 and surface transportation is available to the nearest hub, requiring a time TS, then $XJD_{ij}$ {(TS/TC)/[(TS/TC) + (TC/TS)]} is assigned to the shortest connection. The rest are assumed to travel to the nearest hub by surface transportation. If no connection can be found, all demand $XJD_{ij}$ travels to the nearest hub by surface transportation.

A distinction is made between hubs inside the region and hubs outside of it. While the traffic that moves by surface to a hub outside the region is of no further interest, the traffic that moves to a hub inside the region can be loaded on flights departing from that hub.

The procedure used to determine the local traffic generated by a flight was essentially outlined in the discussion of the demand generation module and will not be repeated here. It requires a time-of-day preference curve and a service level function. The time-of-day preference curve permits one to read off directly a preference value LD for each departure time.

We feel that this approach is more reasonable for a local air service environment than one that integrates the demand over a region of attraction for each flight, especially if that region of attraction reaches to the zones of indifference of the preceding and the succeeding flight. The reason is that the approach chosen here permits a better representation of the elasticity of demand. Difficulties arise when two or more flights are scheduled too close to each other so that they are competing for the same passengers. This can always be ruled out if the schedule is properly set up. In the environment of interest here any such schedule does not appear reasonable. If it were desirable to consider such situations, however, it would not be difficult to enter special provisions into the system.

PLATO permits the use of different time-of-day preference curves for each airport pair. On some routes the demand might arise mostly from business travel with a strong preference for flights near the beginning and end of the business day. On other routes there might be a substantial share of tourists who tend to prefer flights closer to the middle of the day. Recognizing this fact might permit a high utilization of the fleet at a comfortable average load factor and thus lead to reasonably profitable operations.

The service level function (SERLEV) multiplied by the base demand gives the maximum traffic that can be generated with some frequency NF. This traffic can be realized if the flights are placed optimally throughout the day, in which case the sum of the time-of-day preference factors is given by F1(NF). The traffic actually generated is proportional to the actual values of the time-of-day preference factors LD for each flight or their sum for the entire route.

The function SERLEV is characterized by an s-shape. While the first frequency will attract a relatively small volume of traffic, additional frequencies will attract increasing volumes corresponding to the penetration of the market. From a certain frequency onward, the increase in traffic volume will start to decrease, indicating beginning saturation. Eventually the curve will become completely horizontal. The width of the s is strongly dependent on the length of the route and the shape of the time-of-day preference curve. Obviously, the longer the route, the steeper the ascent. For extremely long routes the first frequency can attract all the traffic that can be generated on the route, and additional frequencies can only be justified for capacity reasons. A similar observation can be made for a route with a very pronounced peak in the time-of-day preference curve.

The limited aircraft capacity is simply treated as a restriction on the traffic that can be carried on one flight. In order to take into account the random day-to-day variations of traffic, it is reasonable to reduce the true aircraft capacity by some maximum average load factor. Any traffic that cannot be accommodated is assumed to be lost. We do not believe that in the local air service environment, at least in operations below the point of beginning saturation, it would be reasonable to reassign this demand to flights at other departure times. Excess traffic is removed in increasing order of the per capita revenue.

The last part of this module, the calculation of cost and revenues, is just a lot of multiplication and addition. The only thing that deserves mentioning is that any conceivable cost function can be handled by it. Along with cost and revenue data, statistics on seat kilometers offered, revenue seat kilometers, and average load factors are compiled per aircraft type.

## SCHEDULE EDITOR AND REPORT GENERATOR

The schedule editor permits the user to develop a schedule on a terminal without having to perform any calculations in his or her head or on the side. Thus for any flight with a given departure time, it automatically calculates the arrival time and the earliest possible departure time for the following flight. The schedule is developed in the form of aircraft rotations, and the user can insert or delete flights from a string of flights without having to redefine the whole string. The schedule editor also performs a number of elementary feasibility checks. At the time of this writing, the schedule editor is still not in its final form, and we therefore prefer not to give any details.

The report generator produces a report of the economic results of each flight. For multistop flights, traffic figures are given for each origin-destination (O-D) pair as well as for the flight segment. Also shown are the passengers by O-D pair that could not be accommodated. For interregional flights it is possible to show the traffic originating or terminating in the region without concern for other traffic or cost and revenue figures.

In addition, a summary report is printed. All the intermediate results from the route generation and demand generation modules can be displayed optionally.

## RESULTS AND VALIDATION

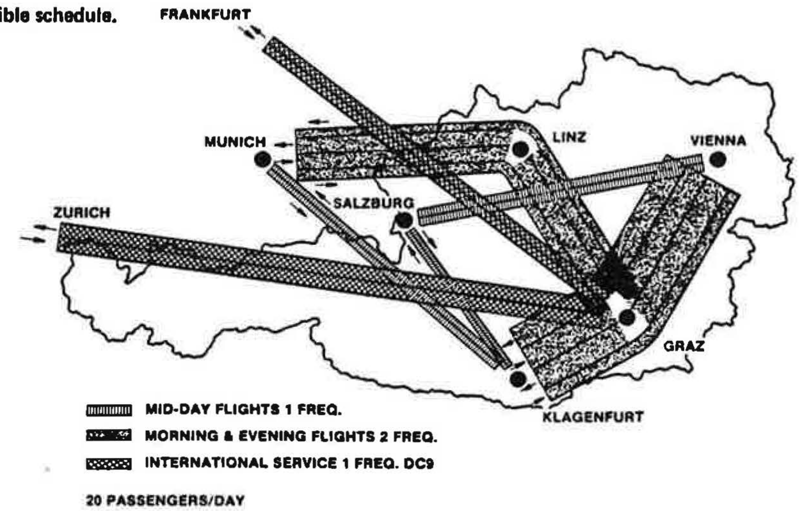
The objective of this study was to demonstrate if and how a viable local air service for Steiermark could be developed. The objective was not satisfied by producing some form of optimum schedule. Rather, a lot of judgment and restraint had to be used to develop a reasonable schedule that would be attractive to all parties concerned. Reasonable meant not only keeping any kind of financial risk within limits but also passing the minimum threshold that would make future growth possible.

Following this, a number of different schedules were produced. All these schedules offered a maximum of two daily frequencies on any route. The results for one of the most attractive schedules in the sample printouts using Fokker F27 equipment are shown graphically in Figure 9. The table below gives a summary of the economic results for that schedule, including some sensitivity analysis.

| | Year | | | | | |
|---|---|---|---|---|---|---|
| Variable | 1975 | | | 1985 | | |
| Fare | 1.00 | 1.25 | 1.25 | 1.00 | 1.25 | 1.25 |
| Driving cost | 1.00 | 1.00 | 1.25 | 1.00 | 1.00 | 1.25 |
| Load factor | 52 | 46 | 52 | 56 | 45 | 56 |
| Revenue: DOC | 1.03 | 1.12 | 1.27 | 1.12 | 1.08 | 1.38 |
| Including fuel tax | 0.81 | 0.89 | 1.00 | 0.87 | 0.87 | 1.09 |

It can be seen that revenue in all cases covers the direct operating cost so long as the domestic fuel tax is not levied. If the fuel tax has to be paid, it is almost impossible for the revenues to cover the direct operating cost. It is interesting to note that the highway projects, which are expected to be completed by 1985, have a stimulating effect on air traffic. Not only do they de-

Figure 9. Proposed possible schedule.




crease the airport access times, but they increase the attraction between districts, leading to a general increase in travel activity.

The results obtained appear quite reasonable and realistic. Since the proposed air service has not yet been realized, it is not possible to check the model results against reality. We were, however, fortunate that at the end of our study a completely independent proposal for local air service was developed by Austrian Airlines (AUA).

| Schedule | Load Factor AUA | Load Factor PLATO |
|---|---|---|
| I | 50 | 46 |
| II | 55 | 43 |
| III | 45 | 46 |

They were considering three schedule alternatives. Their traffic predictions were made without sophisticated models but using the best judgment of the airline. We evaluated all three schedules by means of PLATO. A comparison of results given in Table 2 shows a reasonably good agreement. We should add, however, that these three schedules were economically much less attractive than schedules we developed.

ACKNOWLEDGMENT

I wish to acknowledge the cooperation of Ms. Shan-Shan Wang in developing the PLATO system.

REFERENCES

1. Intertraffic. Prognose des Regionalluftverkehres, Bundesrepublik Deutschland. Düsseldorf, 1970.
2. H. G. Nüsser. Methodik zur Ermittlung der Wirtschaftlichkeit des Regionalluftverkehres. Verkehrswissenschaftliches Institut, Stuttgart, 1972.
3. M. M. Etschmaier. Projects and Implementations in the Schedule Development Process of an Airline. Department of Industrial Engineering, Univ. of Pittsburgh, Technical Rept. No. 17, 1973.
4. R. W. Floyd. Algorithm 97: Shortest Path. Communications of the Association for Computing Machinery, Vol. 5, 1962.

*Publication of this paper sponsored by Committee on Aviation Demand Forecasting.*

# Model to Estimate Commuter Airline Demand in Small Cities

Bruce A. Thorson,* Civil Engineer, City of Des Moines, Iowa
Kenneth A. Brewer, Department of Civil Engineering, Iowa State University, Ames

**This paper considers the factors indicating a community's potential demand for commuter air carrier service, compares these factors to a profile of commuter airline passenger characteristics, and reports the results of an extensive regression analysis to develop a model to estimate commuter airline demand in small cities. The final regression model was nonlinear in nature and incorporated community populations and measurements of isolation from the certificated air carrier transportation system. This model was the basis of a recommended program to integrate commuter air carriers into Iowa's total transportation system.**

Historically, commuter air carrier service has had a general public image of instability. This image resulted from operational failures and discontinuing service to cities along system routes. Primary reasons for such failures and service interruptions are associated with financing, operations, and marketing. Specifically, the aircraft being used have been too large for the markets being served; the operators have not been able to finance operating expenses in the initial period during which patronage is being sought; communities have been reluctant to give financial or marketing support for such operations; and travel time delays may have been excessive when the routes have been short or when numerous station stops have occurred between a community and a terminating hub airport.

An Iowa state-level interest in evaluating the role commuter air carriers and intercity express bus route service might play in long-range transportation development to enhance personal mobility generated a research effort to provide some means of estimating the potential of a given community to utilize commuter air carrier services (1, 2).

Commuter air carriers were defined by the Civil Aeronautics Board (CAB) in 1969 in amendments to the Economic Regulations Part 298: Classification and Exemption of Air Taxi Operators. Air taxi operators are defined as those who "perform, pursuant to published schedules, at least five round trips per week between two or more points, or carry mail."

As such, the roles of the commuter airlines in serving the nation and the public are at least threefold. The primary role is providing passenger service between communities. In addition, they furnish a much needed air cargo service for many businesses and industries. They also transport mail under contract with the U.S. Postal Service.

Commuter airlines serve three different sectors of the air transportation market. First, service between hub airports is provided; second, operations connect CAB certificated points with noncertificated points (usually connecting small, isolated, or rural communities to hub airports); finally, service is offered as a replacement for local service airlines' certificates when authorized by CAB.

Generally, commuter air carriers specialize in serving small communities. Of 414 communities receiving scheduled passenger service in 1976, 108 (26.1 percent) had populations less than 50 000 people (3). Furthermore, during fiscal year 1974, 77 percent of the markets enplaned fewer than 10 passengers per day and 41 percent of the markets served were less than 160 km (100 miles) apart (4).

Calendar year 1972 statistics have shown the average passenger trip length to be 160 km (5). These figures reinforce the primary role of commuter air carriers: service to small communities with short haul lengths and low-density air demand. Such figures also reinforce the need to predict the ultimate market demand for commuter air service in any given community if a state-level transportation planning agency is going to be able to develop plans that can enhance stability and productivity in the contribution of commuter air carriers to the total transportation system.

## IDENTIFICATION OF COMMUTER AIRLINE DEMAND INDICATORS

A comprehensive literature review enumerated a wide range of variables used in estimating passenger demand (1, 4, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32). Among the most frequent variables were population of community served, passenger's household income, community retail sales, and employment in the community (classified by occupational categories). Others of less frequency included community wholesale sales, kilometers to the nearest Federal Aviation Administration (FAA) designated hub airport, passenger education, and passenger age.

A check on the appropriateness of variables reported in the literature related to passenger characteristics was obtained through an on-board commuter air carrier survey at Iowa stations. The 1976 commuter air carrier routes in Iowa and the populations of cities are shown in Figure 1.

Each airline serving these routes was sampled for three consecutive days in the summer of 1976; questionnaires were completed once by all passengers. Pertinent results of the survey are shown in Table 1. This additional input to the variable selection process appeared to reinforce the potential variable list named

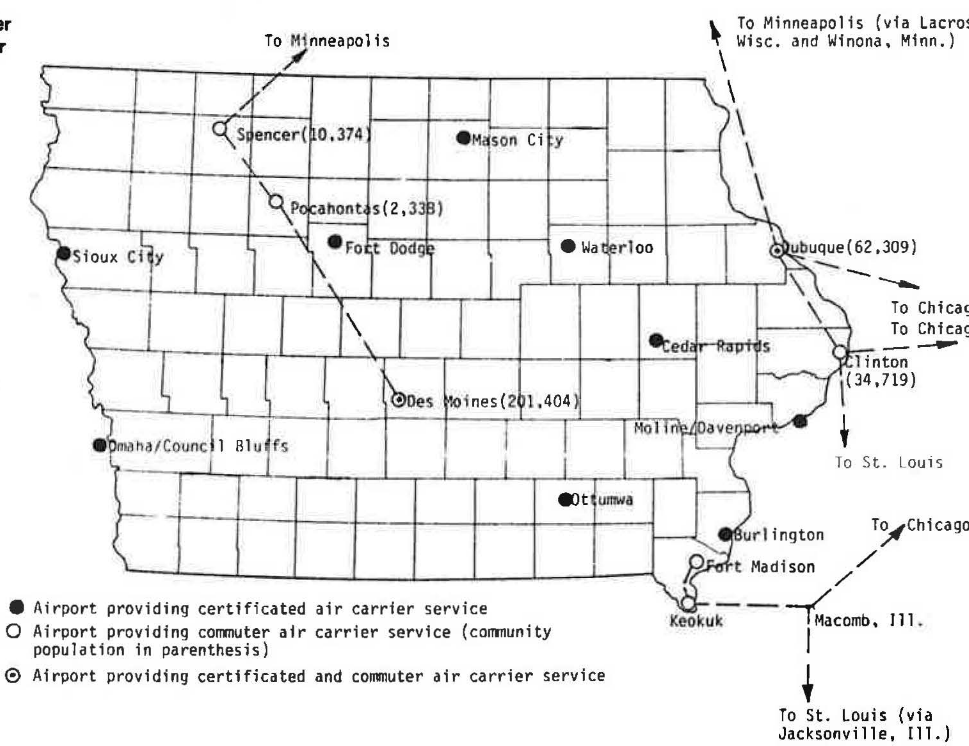



**Figure 1. Iowa commuter air carrier passenger routes and certificated air carrier stations for 1976.**

Table 1. Iowa commuter air carrier passenger survey results for 1976.

| Questionnaire Variable | Response Category | Response[a] Number | Response[a] Percentage |
|---|---|---|---|
| Total trip length, km | 0-158.4 | 0 | 0.0 |
| | 160-238.4 | 32 | 14.7 |
| | 240-318.4 | 62 | 28.4 |
| | 320-478.4 | 49 | 22.5 |
| | 480-638.4 | 8 | 3.7 |
| | 640-798.4 | 7 | 3.2 |
| | 800+ | 60 | 27.5 |
| | | 218 | 100.0 |
| Trip purpose | Business | 164 | 74.2 |
| | Personal or family affairs or shopping | 15 | 6.8 |
| | Medical | 1 | 0.5 |
| | Social or recreational | 39 | 17.6 |
| | Other | 2 | 0.9 |
| | | 221 | 100.0 |
| Number of times previously flown on commuter airlines in past year | 0 | 119 | 54.6 |
| | 1 or 2 | 27 | 12.4 |
| | 3 or 4 | 21 | 9.6 |
| | 5 or 6 | 19 | 8.7 |
| | 7-12 | 11 | 5.0 |
| | 13-24 | 11 | 5.0 |
| | 25-36 | 7 | 3.2 |
| | 36 | 3 | 1.4 |
| | | 218 | 99.9[b] |
| Reason for traveling by commuter airline | Travel time saving | 72 | 34.8 |
| | Travel cost saving | 1 | 0.5 |
| | Convenience or scheduling | 57 | 27.5 |
| | Comfort | 1 | 0.5 |
| | Owned no car or one not available | 1 | 0.5 |
| | Travel time plus other factors | 23 | 11.1 |
| | Only airline available | 41 | 19.8 |
| | Other | 11 | 5.3 |
| | | 207 | 100.0 |
| Traveler occupation | Professional, technical, or managerial | 164 | 81.6 |
| | Farm owner or manager | 3 | 1.5 |
| | Clerical or sales | 12 | 6.0 |
| | Craftworker, equipment operator, or laborer | 5 | 2.5 |
| | Unemployed | 2 | 1.0 |
| | Retired | 4 | 2.0 |
| | Other | 11 | 5.5 |
| | | 195 | 99.9[b] |
| Traveler annual household income, $ | <5 000 | 1 | 0.5 |
| | 5 000-9 999 | 13 | 6.7 |
| | 10 000-14 999 | 19 | 9.7 |
| | 15 000-24 999 | 66 | 33.8 |
| | 25 000-49 999 | 74 | 37.9 |
| | 50 000+ | 22 | 11.3 |
| | | 212 | 100.0 |
| Traveler age | <18 | 7 | 0.9 |
| | 18-24 | 19 | 9.0 |
| | 25-39 | 78 | 36.8 |
| | 40-64 | 102 | 48.1 |
| | >64 | 6 | 2.8 |
| | | 212 | 100.0 |
| Traveler education level | Grade school | 2 | 0.9 |
| | Attended high school | 7 | 3.3 |
| | High school | 27 | 12.6 |
| | Technical school | 5 | 2.3 |
| | Attended college | 47 | 22.0 |
| | College graduate | 81 | 37.9 |
| | Postgraduate | 45 | 21.0 |
| | | 214 | 100.0 |

Note: 1 km = 0.62 mile.

[a] Total sample of 226 passengers; no response to selected items created varying survey response.
[b] Does not equal 100 percent due to rounding.

above, including occupations, annual household income, age, and education level. The survey results indicated that among commuter airline passengers 81.6 percent of the heads of households were from professional, technical, or managerial occupations; 83.0 percent of the households had annual incomes of at least $15 000; over 85 percent of the passengers were 25 years of age or older; and about 60 percent of the passengers were college graduates.

An equation form was desired for the community commuter air carrier demand predictor and, therefore, multiple regression was selected as the technique to generate the predictor. Selection of the variables to be included in the regression analysis was based on three considerations. First, the variables had to be reasonable measures of Iowa community characteristics; second, they should have been substantiated by previous research; third, data on them had to be available and accessible.

Five variables were chosen for a linear multiple regression analysis after carefully evaluating previous studies, examining the results of the on-board passenger survey, and considering the nature of the existing problem. These variables were

1. POPL = 1970 community population,
2. INCOME = percentage of families in the community with annual incomes of at least $15 000,
3. OCCUP = percentage of persons in the community employed in professional, technical, or managerial occupations,
4. EDUC = percentage of persons in the community over 25 years of age with 4 or more years of college education, and
5. ISOLATE = kilometers to the nearest FAA hub airport.

The dependent variable was the average daily passenger enplanements (ADPE) at 58 cities having commuter air carrier service during 1974 (4). Cities from a six-state area with social and economic environments approximating Iowa cities and having commuter air carrier service were selected. Table 2 indicates the variable values for the regression analysis data base.

REGRESSION ANALYSIS

A stepwise variable inclusion multiple regression procedure available in the statistical package for social sciences (SPSS) was utilized to provide linear regression models of the variable data set. The first step included all five independent variables and the data from the 58 communities listed in Table 2. The resulting equation was

$$\begin{aligned} ADPE = & -35.613\,81 + 1.394\,21(POPL) + 1.964\,47(ISOLATE) \\ & + 1.450\,53(OCCUP) - 2.084\,72(INCOME) \\ & + 0.409\,50(EDUC) \end{aligned} \quad (1)$$

**Table 2. Average daily passenger enplanements and independent variable data used in the regression analysis.**

| | Variable | | | | | |
|---|---|---|---|---|---|---|
| | | Independent | | | | |
| Community | Dependent ADPE | POPL (000s) | INCOME | OCCUP | EDUC | ISOLATE (km) |
| Clinton, IA | 13 | 35 | 18.9 | 21.7 | 9.7 | 64 |
| Dubuque, IA | 93 | 91 | 20.0 | 21.7 | 10.1 | 113 |
| Ft. Dodge, IA | 18 | 31 | 17.2 | 24.2 | 10.6 | 145 |
| Ft. Madison, IA | 3 | 14 | 15.3 | 22.8 | 7.6 | 145 |
| Keokuk, IA | 3 | 15 | 14.5 | 24.3 | 8.3 | 161 |
| Mason City, IA | 44 | 30 | 15.9 | 24.2 | 11.0 | 129 |
| Ottumwa, IA | 19 | 30 | 12.0 | 22.8 | 6.6 | 129 |
| Spencer, IA | 5 | 10 | 18.5 | 25.8 | 10.8 | 145 |
| Dodge City, KS | 12 | 14 | 14.5 | 25.8 | 11.4 | 225 |
| Garden City, KS | 25 | 15 | 17.6 | 23.6 | 12.1 | 306 |
| Goodland, KS | 8 | 6 | 12.1 | 19.9 | 7.3 | 290 |
| Great Bend, KS | 12 | 16 | 15.7 | 27.8 | 10.0 | 145 |
| Hays, KS | 27 | 15 | 14.3 | 30.4 | 20.5 | 209 |
| Hutchinson, KS | 5 | 37 | 12.3 | 25.5 | 11.5 | 64 |
| Independence, KS | 6 | 38 | 11.0 | 26.5 | 9.3 | 145 |
| Lawrence, KS | 15 | 46 | 18.9 | 34.5 | 30.0 | 64 |
| Liberal, KS | 35 | 21 | 16.0 | 20.6 | 12.8 | 242 |
| Manhattan, KS | 182 | 59 | 18.6 | 31.5 | 34.5 | 177 |
| Olathe, KS | 7 | 18 | 18.4 | 24.1 | 11.8 | 32 |
| Salina, KS | 41 | 38 | 14.8 | 26.2 | 12.3 | 129 |
| Carbondale, IL | 39 | 23 | 22.3 | 39.2 | 35.4 | 129 |
| Danville, IL | 30 | 43 | 19.3 | 21.9 | 8.5 | 129 |
| Galesburg, IL | 17 | 36 | 17.1 | 21.8 | 9.5 | 64 |
| Jacksonville, IL | 2 | 21 | 17.0 | 24.0 | 11.3 | 113 |
| Macomb, IL | 5 | 20 | 22.7 | 30.1 | 24.0 | 113 |
| Marion, IL | 37 | 21 | 13.2 | 27.2 | 9.1 | 129 |
| Mattoon, IL | 17 | 36 | 15.4 | 19.3 | 7.9 | 161 |
| Mount Vernon, IL | 20 | 16 | 15.1 | 22.5 | 8.7 | 129 |
| Quincy, IL | 62 | 64 | 13.7 | 21.0 | 8.0 | 177 |
| Sterling/Rock Falls, IL | 14 | 26 | 18.0 | 16.7 | 6.7 | 81 |
| Bemidji, MN | 37 | 11 | 14.7 | 30.6 | 15.7 | 161 |
| Brainerd, MN | 30 | 12 | 13.1 | 25.8 | 10.2 | 177 |
| Chisholm/Hibbing, MN | 50 | 22 | 12.5 | 23.7 | 8.9 | 97 |
| Eveleth, MN | 2 | 5 | 9.3 | 21.7 | 7.9 | 81 |
| Fairmont, MN | 11 | 11 | 13.3 | 22.9 | 8.3 | 161 |
| Grand Rapids, MN | 7 | 7 | 15.4 | 27.6 | 11.5 | 129 |
| International Falls, MN | 39 | 6 | 18.8 | 23.8 | 8.1 | 225 |
| Mankato, MN | 9 | 31 | 19.6 | 25.6 | 18.8 | 113 |
| New Ulm, MN | 6 | 13 | 11.6 | 22.4 | 8.0 | 129 |
| Thief River Falls, MN | 25 | 9 | 13.5 | 27.1 | 9.3 | 145 |
| Winona, MN | 7 | 26 | 14.2 | 24.8 | 12.9 | 64 |
| Worthington, MN | 9 | 10 | 16.4 | 29.8 | 10.1 | 97 |
| Cape Girardeau, MO | 36 | 46 | 15.2 | 27.6 | 12.8 | 177 |
| Jefferson City, MO | 10 | 32 | 23.6 | 30.6 | 15.8 | 161 |
| Joplin, MO | 134 | 39 | 12.0 | 25.2 | 8.5 | 113 |
| Kirksville, MO | 9 | 16 | 14.0 | 27.0 | 17.7 | 209 |
| Rolla, MO | 2 | 13 | 20.6 | 37.9 | 27.7 | 161 |
| Alliance, NE | 5 | 7 | 10.8 | 29.0 | 8.6 | 225 |
| Chadron, NE | 6 | 6 | 11.7 | 29.5 | 12.4 | 145 |
| Columbus, NE | 7 | 15 | 13.9 | 21.9 | 8.7 | 113 |
| Grand Island, NE | 76 | 31 | 12.9 | 23.8 | 8.0 | 145 |
| Hastings, NE | 14 | 24 | 15.0 | 24.8 | 9.6 | 101 |
| Kearney, NE | 13 | 19 | 12.9 | 24.7 | 15.2 | 209 |
| McCook, NE | 9 | 8 | 11.8 | 27.4 | 7.3 | 322 |
| Norfolk, NE | 15 | 17 | 14.3 | 22.8 | 7.9 | 97 |
| North Platte, NE | 52 | 19 | 12.4 | 22.7 | 8.3 | 322 |
| Scottsbluff, NE | 61 | 15 | 14.8 | 28.5 | 12.1 | 258 |
| Sidney, NE | 6 | 6 | 9.2 | 22.3 | 6.4 | 242 |

Note: 1 km = 0.62 mile.
Source: 1970 Census (33).

The regression coefficients of this equation are shown in Table 3. The number in parenthesis following the regression coefficient indicates the order in which that variable was introduced into the regression equation by the SPSS program. Below each regression are indicated the number of data points used in the analysis (for example, 58 for Equation 1), the simple correlation of each independent variable with ADPE (r), the cumulative multiple $r^2$, and the absolute value of the t-statistic for the regression coefficient. Examination of these results in Table 3 indicates some undesirable effects in Equation 1. First, the large negative constant is undesirable. Second, the negative sign on INCOME is illogical. Third, EDUC is more highly correlated with INCOME than any of the independent variables are with ADPE. Fourth, only the regression coefficients of POPL and ISOLATE have associated t-values that are significant at the 0.05 level or better. Finally, the last three variables did not add much to the predictive capability of the estimating equation as measured by the cumulative multiple $r^2$.

Examination of the data revealed that Manhattan, Kansas, and Joplin, Missouri, had excessively high levels of actual average daily passenger enplanements when compared to cities of comparable size. The proximity of Manhattan to a military base and Joplin to the Ozark region recreation areas apparently made these two cities atypical in relation to Iowa cities, so they were deleted from the data set for further analysis.

The next analysis step consisted of using POPL, ISOLATE, and OCCUP to estimate demand. Equation 2 in Table 3 resulted. A third step was the same analysis with INCOME replacing OCCUP (Equation 3). Comparing the results of Equations 2 and 3 in Table 3, it is evident that the variables OCCUP and INCOME did not enhance the reliability of either equation to predict average daily passenger enplanements. Analysis of the residuals between predicted values and actual values for the data set indicated that some stratification of the data might improve the predictability of the estimating equation.

Since there was some evidence that a level of 10-15 average daily passenger enplanements was a demand level assuring that commuter air carriers could operate successfully (4), all five independent variables were included in an analysis with all data points having ADPE less than or equal to 15 (Equation 4 results in Table 3) and all data with ADPE greater than or equal to 10 (Equation 5 results in Table 3). Note that EDUC did not explain enough of the residual variance to be introduced into Equation 4, that Equation 4 has very low $r^2$, and that the only POPL has a regression coefficient significant at the 0.05 level. Equation 5 has a higher $r^2$, has illogical negative signs on INCOME and EDUC, and has significant coefficients only on POPL and ISOLATE. In general, neither of these equations is satisfactory. Equations 6 and 7 resulted from the same data stratification and included only POPL and ISOLATE as independent variables. Examination of residuals showed no significant difference between Equations 4 and 6 and between Equations 5 and 7. Furthermore, the regression coefficient on ISOLATE in Equation 6 is not significant at the 0.05 level.

The sensitivity of the regression parameters to the stratification level was examined by selecting a subsample consisting of all data with ADPE greater than 15. Equation 8 resulted from this analysis. The regression coefficient on ISOLATE is no longer significant at the 0.05 level as compared to Equation 7 in which it was significant.

Based on the analysis described above, a regression of POPL and ISOLATE with ADPE for the total sample of 56 data points was considered to be the best general linear model (Equation 9). Stratification by community population as an independent variable was preferred to stratification by ADPE, because the mean value of ISOLATE was substantially different for those cities over 20 000 people as compared to cities under 20 000 people.

Two regressions were performed on the 30 cities with populations less than 20 000 (Equation 10) and on all cities having populations 20 000 or more (Equation 11). Note that in Equation 10 the variable ISOLATE entered the solution first and that, while the regression coefficient POPL is not significant at the 0.05 level, it is at the 0.10 level. An inverse situation existed for Equation 11 as compared to Equation 10 (reinforcing the logical consequence of city size and degree of isolation as intervening factors). Even though Equations 10 and 11 based on stratified data explain less of the sample variance as measured by $r^2$, comparison of residuals indicated that Equations 10 and 11 were preferred to Equation 9. Of the 30 communities used as a data base in Equation 10, 16 had smaller residuals than resulted from Equation 9 with an average decrease of 1.24; for the 14 communities producing an increased residual value with Equation 10, the average increase was 0.93. Similarly, 14 of the 26 communities in the Equation 11 data base had lower residuals with Equation 11 than with Equation 9, with an average decrease of 2.06. The 12 communities having increased residuals with Equation 11 as compared to Equation 9 residuals increased an average of 0.84. Thus, Equations 10 and 11 were considered to be the best linear multiple regression model.

Since a stratified sample produced the best linear model and the relatively low percentage of the sample variance was never explained (40 plus percent), an extensive multiple regression analysis was conducted to seek a nonlinear model with higher $r^2$. The following list shows the various independent variable transformations utilized in this analysis [the dependent variables are ADPE, ADPE/1000, ln (ADPE)]:

$(\mathrm{ISOLATE})^{-1}$
$\ln(\mathrm{ISOLATE})$
$\ln(\mathrm{POPL})$
$(\mathrm{EDUC}) \times (\mathrm{INCOME})$
$\ln(\mathrm{EDUC})$
$\ln(\mathrm{INCOME})$
$(\mathrm{ISOLATE}) \times (\mathrm{ISOLATE})$
$[(\mathrm{ISOLATE}) \times (\mathrm{ISOLATE})]^{-1}$
$\ln(\mathrm{OCCUP})$
$(\mathrm{ISOLATE}) \times (\mathrm{INCOME})$
$(\mathrm{OCCUP}) \times (\mathrm{INCOME})$
$(\mathrm{OCCUP}) \times (\mathrm{EDUC})$
$(\mathrm{INCOME}) \times (\mathrm{INCOME})$
$(\mathrm{POPL})^{1/2}$
$(\mathrm{ISOLATE})^{1/2}$
$(\mathrm{POPL})^{-1}$
$(\mathrm{EDUC})^{-1}$
$(\mathrm{INCOME})^{-1}$
$(\mathrm{OCCUP})^{-1}$
$(\mathrm{OCCUP})^{1/2}$
$(\mathrm{EDUC})^{1/2}$
$(\mathrm{INCOME})^{1/2}$
$(\mathrm{POPL}) \times (\mathrm{POPL})$
$(\mathrm{OCCUP}) \times (\mathrm{OCCUP})$
$(\mathrm{EDUC}) \times (\mathrm{EDUC})$
$[(\mathrm{POPL}) \times (\mathrm{POPL})]^{-1}$
$[(\mathrm{EDUC}) \times (\mathrm{EDUC})]^{-1}$
$[(\mathrm{INCOME}) \times (\mathrm{INCOME})]^{-1}$
$[(\mathrm{OCCUP}) \times (\mathrm{OCCUP})]^{-1}$
$(\mathrm{POPL})^{3}$
$(\mathrm{EDUC})^{3}$

Table 3. Results of linear stepwise inclusion multiple regression analysis.

| Equation No. | Constant | POPL | ISOLATE | OCCUP | INCOME | EDUC |
|---|---|---|---|---|---|---|
| 1 | -35.613 81 | 1.394 21 (1) | 1.966 47 (2) | 1.450 53 (3) | -2.084 72 (4) | +0.409 50 (5) |
| | 58 (r, $r^2$, t) | (0.57, 0.32, 6.03) | (0.09, 0.39, 2.26) | (0.18, 0.43, 1.00) | (0.95, 0.46, 1.58) | (0.24, 0.46, 0.37) |
| 2 | -16.315 49 | 0.828 19 (1) | 1.759 12 (2) | 0.078 58 (3) | N.A. | N.A. |
| | 56 (r, $r^2$, t) | (0.53, 0.28, 5.64) | (0.16, 0.40, 3.19) | (-0.08, 0.40, 0.15) | - | - |
| 3 | -9.982 40 | 0.838 98 (1) | 1.714 00 (2) | -0.27102 (3) | N.A. | N.A. |
| | 56 (r, $r^2$, t) | (0.53, 0.28, 5.61) | (0.16, 0.40, 3.06) | (0.07, 0.40, 0.38) | - | - |
| 4 | 5.778 77 | 0.1720 (1) | 0.273 38 (2) | -0.235 30 (3) | 0.179 75 (4) | - |
| | 32 (r, $r^2$, t) | (0.36, 0.13, 2.29) | (-0.02, 0.17, 1.41) | (-0.11, 0.20, 1.32) | (0.16, 0.22, 0.79) | - |
| 5 | -22.576 31 | 0.867 99 (1) | 1.859 95 (2) | 1.916 00 (4) | -1.523 04 (3) | -1.026 10 (5) |
| | 34 (r, $r^2$, t) | (0.46, 0.21, 4.37) | (0.16, 0.34, 2.33) | (-0.03, 0.40, 1.44) | (-0.09, 0.38, 1.24) | (-0.06, 0.42, 1.02) |
| 6 | 3.041 53 | 0.173 12 (1) | 0.206 00 (2) | N.A. | N.A. | N.A. |
| | 32 (r, $r^2$, t) | (0.36, 0.13, 2.43) | (-0.02, 0.17, 1.12) | - | - | - |
| 7 | -7.132 43 | 0.701 71 (1) | 1.955 91 (2) | N.A. | N.A. | N.A. |
| | 34 (r, $r^2$, t) | (0.46, 0.21, 3.78) | (0.16, 0.34, 2.42) | - | - | - |
| 8 | 6.968 70 | 0.618 63 (1) | 1.365 86 (2) | N.A. | N.A. | N.A. |
| | 24 (r, $r^2$, t) | (0.51, 0.26, 3.20) | (0.06, 0.33, 1.46) | - | - | - |
| 9 | -14.194 85 | 0.824 58 (1) | 1.754 61 (2) | N.A. | N.A. | N.A. |
| | 56 (r, $r^2$, t) | (0.53, 0.28, 5.74) | (0.16, 0.40, 3.22) | - | - | - |
| 10 | -12.292 17 | 0.946 23 (2) | 1.484 59 (1) | N.A. | N.A. | N.A. |
| | 30 (r, $r^2$, t) | 0.25, 0.26, 1.63) | (0.43, 0.18, 2.67) | - | - | - |
| 11 | -21.099 84 | 0.854 18 (1) | 2.461 14 (2) | N.A. | N.A. | N.A. |
| | 26 (r, $r^2$, t) | (0.56, 0.32, 3. 46) | (0.28, 0.40, 1.74) | - | - | - |

Table 4. Results of nonlinear stepwise inclusion multiple-regression analysis number 5.

| Equation (N = 56) | Constant | (ISOLATE) × (POPL) | $(POPL)^3$ | $(POPL)^2$ | $(POPL)^{1/2}$ | $r^2$ | Parameter |
|---|---|---|---|---|---|---|---|
| 12 | 2.816 94 | 0.093 72 | | | | 0.401 | Coefficients |
| | | (6.01) | | | | | (t) |
| 13 | 6.226 03 | 0.067 55 | 0.000 05 | | | 0.439 | Coefficients |
| | | (3.29) | (1.89) | | | | (t) |
| 14 | 6.565 90 | 0.090 48 | 0.000 17 | -0.011 89 | | 0.460 | Coefficients |
| | | (3.49) | (1.95) | (1.43) | | | (t) |
| 15 | -8.134 46 | 0.088 99 | 0.000 30 | -0.027 43 | +4.875 62 | 0.469 | Coefficients |
| | | (3.42) | (1.76) | (1.43 | (0.90) | | (t) |

Table 5. Simple correlation matrix for analysis number 5.

| Variable | ADPE | (ISOLATE) × (POPL) | $(POPL)^3$ | $(POPL)^2$ | $(POPL)^{1/2}$ |
|---|---|---|---|---|---|
| ADPE | 1.00 | 0.633 12 | 0.569 72 | 0.575 19 | 0.484 02 |
| (ISOLATE) × (POPL) | | 1.00 | 0.671 84 | 0.764 61 | 0.782 25 |
| $(POPL)^3$ | | | 1.00 | 0.968 53 | 0.685 83 |
| $(POPL)^2$ | | | | 1.00 | 0.833 96 |
| $(POPL)^{1/2}$ | | | | | 1.00 |

$(ISOLATE)^3$
$(INCOME)^3$
$(OCCUP)^3$
$(POPL)^{-3}$
$(EDUC)^{-3}$
$(ISOLATE)^{-3}$
$(INCOME)^{-3}$
$(OCCUP)^{-3}$
(POPL) × (INCOME)
(ISOLATE) × (OCCUP)
(ISOLATE) × (EDUC)
(ISOLATE) × (POPL)
(EDUC) × (POPL)
(POPL) × (OCCUP)

Six different stepwise analyses were conducted resulting in 26 different nonlinear regression models.

Table 4, where ADPE is the dependent variable, indicates the results of the nonlinear regression analysis yielding the best $r^2$ and most significant regression coefficients. Equations 12-15 have equal or better ability to explain variation in the 56-community sample data than do the best linear equations (Equations 9, 10, and 11). However, only the coefficients on the variable (ISOLATE) × (POPL) are significant at the 0.05 level. As Table 5 indicates, other independent variables are more highly intercorrelated than they are correlated with the dependent variable, ADPE. Thus, the best nonlinear regression expression is Equation 12.

## CONCLUSIONS

Comparing linear to nonlinear regression results indicates that the expression ADPE = 2.81694 + 0.9372 (ISOLATE) × (POPL), which is Equation 12, is the best estimator of average daily passenger enplanements at a small community. The constraint term is very small and yields little conceptual error when the community is very close to an FAA hub airport (ISOLATE approaches zero) or when the community population is very small (POPL approaches zero). The variable interaction of isolation times population is analogous to the concept of gravity model-travel resistance formulas long accepted in trip distribution analyses. Finally, it is a simple expression.

Recognizing that Equation 12 only explains 40 percent of the variance in the sample data, we still considered it a sufficiently adequate estimator of ultimate demand for commuter air carrier service to be one of the prime bases of the recommended program for integrating commuter air carriers into Iowa's total transportation system (34).

Demand estimates were calculated for 42 cities in Iowa considered to be candidates for expanded commuter air carrier service. These demand estimates along with public attitudes and highway trip diversion estimates resulted in a recommended planning program.

ACKNOWLEDGMENTS

We wish to acknowledge the support of the Engineering Research Institute and the Iowa Department of Transportation for whom this research was conducted. Opinions and conclusions drawn from these analyses are, however, solely our own.

REFERENCES

1. Intercity Passenger Carrier Improvement Study. Engineering Research Institute, Iowa State Univ., Ames, 1977.
2. B. A. Thorson. Commuter Airline Demand and Estimates for Selected Iowa Communities. Iowa State Univ., Ames, MS thesis, 1977.
3. Time for Commuters. Commuter Airlines Association of America, Washington, DC, Rept. 4, Oct. 1976.
4. P. A. Wynns. Air Service to Small Communities. U.S. Department of Transportation, Office of Transportation Regulatory Policy, March 1976.
5. Commuter Air Carriers—An Overview. Federal Aviation Administration, Office of Aviation Policy, Policy Development Division, Aug. 1974.
6. Oregon Commuter Air Service Project. The Aerospace Corporation, El Segundo, California, and the Oregon Department of Transportation, Salem, Summary Rept. and Technical Rept., 1975.
7. W. E. Guenzler. A Study of Trip Making From Iowa's Commercial Airports. Iowa State Univ., Ames, MS thesis, 1973.
8. Integrated Analysis of Small Cities' Intercity Transportation to Facilitate the Achievement of Regional Urban Goals. Engineering Research Institute, Iowa State Univ., Ames; U.S. Department of Transportation, Final Rept. DOT-TST-75-13, June 1974.
9. J. E. Habanek. Commuter Airline Demand Analysis for Mississippi Valley Airlines, Inc. Graduate School of Business Administration, Univ. of Minnesota, Minneapolis, thesis, June 1975.
10. Iowa State University System Plan. Engineering Research Institute, Iowa State Univ., Ames, Vol. 2, Technical Supplement, Nov. 1972.
11. Iowa State Airport System Plan Update. Engineering Research Institute, Iowa State Univ., Ames, Final Rept., May 1976.
12. G. M. Clark and W. Bates. Ohio Third-Level Aviation Program Analysis. Systems Research Group, Ohio State Univ., Columbus, Technical Rept. EES 5511-3, Sept. 30, 1975.
13. The Aerospace Corporation. Pacific Northwest Region Air Service Project. El Segundo, CA, Final Rept., Feb. 1975.
14. J. E. Greiner Company, Inc. Air Transportation Market Demand in the Coastal Plains, Tampa, 1971.
15. Ralph H. Burke, Inc. Air Transportation Requirements Study for the State of Nebraska. Chicago, Aug. 1973.
16. R. A. Calderone. A Digest and Assessment of Air Travel Forecasting Techniques. Institute of Transportation and Traffic Engineering, Univ. of California, Berkeley, Graduate Rept., Aug. 1967.
17. Vogt, Ivers and Associates. Social and Economic Factors Affecting Intercity Travel. NCHRP, Rept. 70, 1969.
18. Air Travel Forecasting. Aviation Department, Forecast and Analysis Division, Port of New York Authority; Eno Foundation for Highway Traffic Control, Saugatuck, CT, Jan. 1957.
19. N. D. Baxter and E. P. Howrey. The Determinants of General Aviation Activity: A Cross-Sectional Analysis. Transportation Research, Vol. 2, No. 1, March 1968, pp. 73-81.
20. D. F. Wood. Analysis of General Aviation Activity. Transportation Research, Vol. 4, No. 4, Dec. 1970, pp. 349-357.
21. B. T. Ratchford. A Model for Estimating the Demand for General Aviation. Transportation Research, Vol. 8, No. 3, Aug. 1974, pp. 193-203.
22. A. S. DeVany and E. H. Garges. A Forecast of Air Travel and Airport and Airway Use in 1980. Transportation Research, Vol. 6, No. 1, March 1972, pp. 1-18.
23. R. L. Schultz. Studies of Airline Passenger Demand: A Review. Transportation Journal, Vol. 2, No. 4, Summer 1972, pp. 48-62.
24. P. M. Pearson. Demand for Travel on the Canadian Airway System. HRB, Highway Research Record 369, 1971, pp. 47-64.
25. N. L. Johnson. Forecasting Airport Traffic. Proc., ASCE, Vol. 90, No. AT2, Oct. 1964, pp. 181-195.
26. F. P. D. Navin and R. P. Wolsfeld, Jr. Analysis of Air Passenger Travel in the Twin Cities Metropolitan Area. HRB, Highway Research Record 369, 1971, pp. 26-38.
27. Preliminary Results—Mesaba Airline Passenger Survey. Office of Transportation Research, Iowa Department of Transportation, Ames, 1976.
28. J. F. Brown. Airport Accessibility Affects Passenger Development. Proc., ASCE, Vol. 91, No. AT1, April 1965, pp. 47-58.
29. J. L. Scheiner. The Effect of the Interstate System on Short-Haul Passenger Demand. Transportation Science, Vol. 1, No. 4, Nov. 1967, pp. 286-294.
30. R. H. Ellis, P. R. Rassam, and J. C. Bennett. Consideration of Intermodal Competition in the Forecasting of National Intercity Travel. HRB, Highway Research Record 369, 1971, pp. 253-261.
31. J. W. Billheimer. Segmented, Multimodal, Intercity Passenger Demand Model. HRB, Highway Research Record 392, 1972, pp. 47-57.
32. J. C. Goodknight. Model for Estimating Regional Air Passenger Travel Demands. HRB, Highway Research Record 472, 1973, pp. 82-91.
33. 1970 Census of Population: Characteristics of the Population. U.S. Department of Commerce, Bureau of the Census, 1973.
34. Intercity Passenger Carrier Improvement. Engineering Research Institute, Iowa State Univ., Ames, Aug. 1977.

*Publication of this paper sponsored by Committee on Aviation Demand Forecasting.*

**Mr. Thorson was at the Engineering Research Institute, Iowa State University, when this research was done.*

# Distributing Air Cargo in the Baltimore-Washington Region

Mark E. Tomassoni, Department of Geography and Environmental Engineering, Johns Hopkins University, Baltimore
David Rubin, COMSIS Corporation, Wheaton, Maryland

**The planning of air cargo terminal systems requires accurate forecasts of demand, particularly of the impact of points of cargo origin and destination on demand. A methodology for this problem should meet three criteria if it is to be useful for airport planning. It must use low-cost and generally available data; it must be based on cargo shipper and receiver behavior; and it must allow for the transferability of results required for general forecasting. This paper describes a methodology based on econometric analysis of data from a number of small geographic areas within the Baltimore-Washington region. Tests of the method were performed by the Maryland Department of Transportation as part of a Maryland Aviation System Plan designed to address the current status of all aviation facilities in the state and to prepare recommendations for any required expansion or development. This case study suggests that the method is a useful tool for forecasting the arrival and departure of air cargo within the region but that it is not necessarily adequate for testing whether policy changes affect demand.**

The forecast of intraurban commodity distribution is central to an understanding of the locations of future cargo generation and consumption. (In the following, "cargo" refers to all freight, mail, and express shipments by air.) Yet, recent goods-movement studies emphasize the significant absence of data bases and rudimentary record keeping necessary for forecasting growth and for understanding the implications of the forecasts for fixed facility development.

According to one federally sponsored project (1, p. 9), "little is known at this time about the intra-urban flow of goods. Neither adequate theory, nor models, nor hypotheses detailing activity linkages within urban areas exist. Thus there is no base from which the design of empirical studies of goods movement can be systematically derived." Unless proper record keeping is established for ongoing intraurban commodity distribution, the potential increases for preparing inaccurate projections and developing imprecise assessments of transportation externalities such as land use, industrial location, regional taxation, and environmental disamenities.

The purpose of the present paper is to summarize the results of a two-day survey methodology that forecast air cargo generation for 78 geographic areas in the Baltimore-Washington region. Although air cargo transportation is often recognized as an interurban commodity transportation phenomenon, the intraurban surface haul of air cargo before and after actual air transportation is central to successful and timely movement. So important is the truck haul of air cargo that it was the subject of considerable study, particularly in the middle and late 1960s.

Brewer was often found using the phrase "chaos on the ground" when describing the surface transportation of air cargo (2, 3). It was his contention that the surface transportation and ground handling of air cargo could account for as much as 40-80 percent of total in-transit time.

Commodities transported by air are typically of high value per unit weight. Examples include electronic equipment, perishable foods, nonmechanical equipment, replacement parts for operating machinery, and printed matter. One would have expected the airlines to devote extensive research efforts toward the analysis of the intraurban and surface haul of air cargo. But available research is scarce.

In addition to Brewer's work, there are at least two other studies that address the air cargo issue. Reeher and Dwyer examined the air-truck network serving the Baltimore-Washington region (4). They sought to reveal the relationships among airport air cargo market areas, air carrier operating schedules, motor carrier operations, rate-setting behavior, and intraurban commodity flows. While no previous study had presented in such a comprehensive fashion the numerous variables affecting the surface movement of air cargo, the Reeher and Dwyer study suffered from the absence of an adequate cargo movement data base.

Tomassoni and Weissbrod (5) extended the work of Reeher and Dwyer by suggesting that, within multiairport environments, market area competition is easily determinable by plotting surface haul cartage rates set by local trucking agencies to and from the airports. Accordingly, they identified three types of market areas—exempt, indifferent, and preferred—within the same Baltimore-Washington regional study area. Exempt market areas were defined as locations within a 40-km (25-mile) radius of the airport or the city limits of the air carrier's certificated route point (6). Indifferent areas were classified as areas with no preference of airport use. Preferred market areas were described as areas with a clear preference for one airport over another when measured in terms of motor carrier freight rates.

Despite the insights provided by these two research teams, neither was able to examine in geographic depth the places of air cargo origination and destination due to the absence of an adequate cargo movement data base. It was evident that the next major study of air cargo in the region should begin to develop the information background needed to describe the existing transportation of air cargo and to provide the framework for projections of forecast demand.

## MARYLAND AVIATION SYSTEM PLAN

Mindful of the need for more sophistocated commodity movement data, the staff of the Maryland Department of Transportation included, as part of their statewide aviation system plan, an extensive treatment of air cargo flow. Efforts were initiated to "identify and examine the structure, operations, and quantity" of domestic air cargo originating from the three airports Baltimore-Washington International (BWI), Dulles International (IAD), and Washington National (DCA) (7).

### The Airports

BWI, IAD, and DCA form an important multiairport air cargo system. The market area served by these three airports extends roughly north to Harrisburg, northwest to Cumberland, southwest to Roanoke, south to Elizabeth City, east to the Delmarva peninsula, and northeast to Aberdeen. Handling more than 11.5 Gg (12 700 tons) of enplaning cargo in 1976, these regional

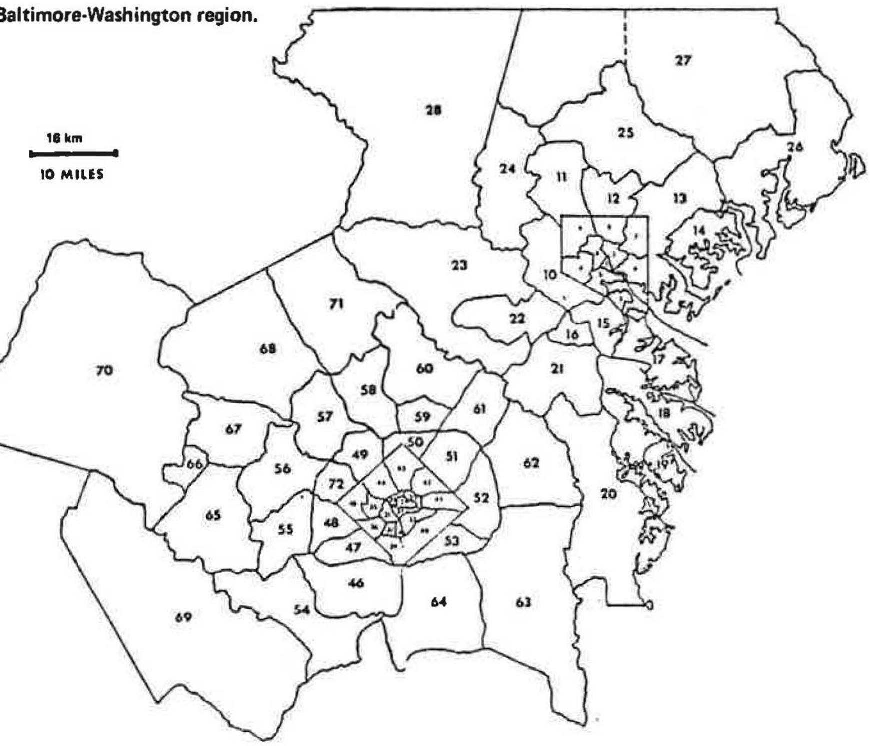


Figure 1. Zones within the Baltimore-Washington region.

airports together accounted for 2.28 percent of all domestically enplaned cargo (8).

The historical development and current position of the airports can be described as follows. The first airport of recognized significance to be constructed in the region was Washington National following a decree set forth by President Roosevelt in the late 1930s. Federal funds were allocated and DCA was completed in 1941 at a distance of only 6.5 km (4 miles) from the Washington, D.C., central business district. Though dominating the transportation of cargo until only recently, the increased demand for air cargo (9) and the associated growth in surface vehicle movement has led to serious on-airport vehicle congestion problems. This, combined with its obsolete 1393-$m^2$ (15 000-$ft^2$) air cargo building, has forced cargo-handling activity to relocate in dispersed hangar areas and has decreased incentive to ship air cargo via DCA.

The second airport to be constructed in the region was the Baltimore-Washington International Airport, then known as Friendship International. Christened in 1951 by President Truman, BWI is located 16 km (10 miles) south of Baltimore's central business district and only 32.3 km (20 miles) north of the District's mall area. BWI is currently experiencing an unprecedented rate of modernization and expansion of terminal facilities, cargo handling areas, airline schedules, and marketing efforts. The state-owned and state-operated 10 219 $m^2$ (110 000 $ft^2$) of air cargo facility and the airline-owned rampside cargo structures accounting for an additional 5946 $m^2$ (64 000 $ft^2$) allow BWI to claim the greatest regional freight-handling ability.

Indeed, by late 1976, BWI began to handle more air cargo than DCA for the first time in the region's history. Moreover, since motor carriers are not subject to the continued on-airport congestion that they would experience at Washington National, terminal delays are reduced, and truck and driver operating times can be more effectively utilized.

These observations, together with the capability to increase cargo warehousing space to more than 27 871 $m^2$ (300 000 $ft^2$), make it apparent that BWI has a major advantage in air cargo and has the incentive to obtain as large as possible a percentage share of the region's total air cargo market.

Completing the facilities is Dulles International Airport, dedicated in 1962 by John F. Kennedy on an expansive semirural Virginia plot some 38.6 km (24 miles) from downtown Washington, D.C. Constructed in large part in response to the then growing congestion at DCA, IAD has attracted long-haul aircraft schedules that are restricted from use out of DCA. Obstacles in the form of lengthy highway access times from the District, access road closed to truck traffic because of structural design, and limited aircraft schedules have limited air cargo use at Dulles.

## The Survey

To assess the amount and types of air cargo originating in the region, a two-day, 100 percent sample of surface haul airway bills was obtained from the four major air cargo forwarders (10). These forwarders were Air Cargo International (ACI), Emery Air Freight, Airborne Freight Corporation, and Railway Express Association (REA) Air Express (now defunct). A total of 2178 shipments were surveyed.

From the airway bills the following information was coded for each shipment: name of forwarder, place of cargo origination, airport in the region used to ship the cargo, destination airport, number of pieces in shipment, total weight of shipment, and date. The region was subdivided into 78 analysis zones (Figure 1), of which 6 were used to describe nonregional points of origination.

The total cost of the air cargo survey including survey distribution, computer coding, and expansion to annual figures came to approximately \$5000. With a total of 2178 survey shipments recorded, the average cost of each survey item came to about \$2.30.

Table 1. Air cargo forwarder survey results.

| Forwarder | Shipments to Airports BWI | DCA-IAD | Total | Percentage of Total |
|---|---|---|---|---|
| REA Air Express | 283 | 598 | 881 | 40 |
| Airborne Air Freight | 183 | 161 | 344 | 16 |
| Air Cargo, Inc. | 181 | 79 | 260 | 12 |
| Emery Air Freight | 277 | 416 | 693 | 32 |
| Total | 924 | 1254 | 2178 | 100 |

Table 2. Air cargo forwarder survey weights.

| Airport | Forwarder | Number of Samples | Total Sample Weight (kg) | Average Weight per Sample (kg) |
|---|---|---|---|---|
| BWI | REA Air Express | 283 | 4 283 | 15.1 |
| | Airborne Air Freight | 183 | 8 355 | 45.7 |
| | Air Cargo, Inc. | 181 | 20 957 | 115.8 |
| | Emery Air Freight | 277 | 14 482 | 53.3 |
| | Subtotal | 924 | 48 077 | 57.5 |
| DCA-IAD | REA Air Express | 598 | 9 695 | 16.2 |
| | Airborne Air Freight | 161 | 7 257 | 45.1 |
| | Air Cargo, Inc. | 79 | 8 508 | 107.3 |
| | Emery Air Freight | 416 | 12 303 | 29.6 |
| | Subtotal | 1254 | 37 763 | 50.0 |
| | Total | 2178 | 85 840 | 39.4 |

Note: 1 kg = 2.2 lb.

### Results

Of the total sample, REA accounted for 40 percent of the total, Emery was represented by 32 percent, Airborne Freight Corporation by 16 percent, and ACI by 12 percent (Table 1). The samples accounted for a total weight of 85 840 kg (189 252 lb) or an average weight of 39.4 kg (86.9 lb) per shipment (Table 2). ACI handled the largest average weight per shipment in deliveries to BWI as well as DCA and IAD.

### Factoring and Forecasts

The total cargo originating in each zone during the two-day period was then expanded proportionately to annual totals during the survey year. Two sources of annual air cargo volume data were used to find the proper expansion factors: Maryland Department of Transportation's State Aviation Administration Comparative Summary of Activity (June 1973-June 1974) and the Federal Aviation Administration's National Capital Airports Activity Reports for Dulles International and Washington National Airports (July 1973-June 1974). Both of these data sources break down the air cargo activity into enplaning and deplaning cargo by type, i.e., mail, freight (domestic-international), and express. The expansion factor was calculated by simply dividing the annual kilograms by the survey kilograms.

Forecasts of 1985 and 1995 cargo origination were then developed based on demographic forecasts supplied by the Baltimore Regional Planning Council and the Washington Council of Governments. The forecasts were controlled by a macroforecast for the region as a share of the national forecast.

### Distribution of Cargo to Air Destinations

The forecast for originating cargo was then distributed to leading domestic destinations by using the Simat, Hellisen, and Eichner air passenger distribution model. Aircraft belly capacity was determined for various aircraft types. A 65 percent load factor was assumed, and a certain percentage of cargo capacity was eliminated for use as luggage space.

The originating cargo demand was calculated for each of the three regional airports and from each of the 78 zones based on accessibility measures from the air passenger portion of the state system plan (9). Cargo was distributed to the closest airport, but DCA was limited to shorter trips. Final output displayed total cargo from each aviation analysis zone to each of the three airports.

## CONCLUSIONS

This paper has attempted to summarize the results of a two-day survey that forecast air cargo generation for the Baltimore-Washington region. Some questions that must be raised and carefully considered, before accurate evaluations of air cargo origination are possible, are

1. Which points generate most air cargo?
2. Which commodities are produced at generating points?
3. Are local airports providing airlines with adequate cargo storage capacity at or near the airport? If not, is this reducing the amount and/or type of cargo shipped through the airport?
4. When a survey is used, is consideration given to sampling cargo destined for the local airport(s)?
5. Can an ongoing record-keeping system of air cargo flows be established?
6. What are the identifiable market areas surrounding a particular airport?

This list could be greatly expanded. What is at issue is the relative importance of gathering this information, the cost of such a study, and the final use to which the information could be profitably put. It seems at least plausible to conclude from the results of this study that, placed in the hands of the proper marketing experts, the type of study conducted by the Maryland Department of Transportation could prove useful in expanding cargo movement by air through local air carrier airports.

This paper has not in any way attempted to describe in comprehensive detail the procedures and final forecasts resulting from the study. Those interested in obtaining more background information are asked to contact the Maryland Department of Transportation.

## ACKNOWLEDGMENT

Acknowledgment is made to R. Dixon Speas and Associates and Alan M. Voorhees Associates, the Baltimore Regional Planning Council, the Washington Council of Governments, and Simat, Hellisen, and Eichner, Inc.

## REFERENCES

1. An Overview of Urban Goods Movement: Project and Data Sources. Office of Systems Analysis and Information, U.S. Department of Transportation, 1973.
2. S. Brewer. Air Cargo Comes of Age. Graduate School of Business Administration, Univ. of Washington, Seattle, 1966.
3. R. G. O'Lone. Cargo Slump Laid to Chaos on Ground. Aviation Week and Space Technology, June 17, 1968.
4. D. H. Reeher and J. W. Dwyer. The Washington-Baltimore Regional Air Freight Transportation

System. Analytic Services, Inc., Falls Church, VA, 1970.
5. M. E. Tomassoni and R. S. P. Weissbrod. Motor Carrier Freight Rates and Airport Market Areas in the Washington-Baltimore Bi-Region. Proc., 4th Intersociety Conference on Transportation, Los Angeles, July 1976.
6. Economic Regulations. Civil Aeronautics Board, Chap. 2, Subchap. A, Pt. 222: Air Cargo Pickup and Delivery Zones.
7. Maryland Aviation System Plan. Maryland Department of Transportation, Technical Rept., March 1976, pp. 1-13.
8. Federal Aviation Administration. Airport Activity Statistics. 12 months ending June 30, 1976.
9. L. M. Schneider. The Future of the U.S. Domestic Air Freight Industry. Graduate School of Business Administration, Harvard Univ., Cambridge, MA, 1973.
10. Maryland Aviation System Plan. Maryland Department of Transportation, Working Paper 5, Oct. 1975.

*Publication of this paper sponsored by Committee on Aviation Demand Forecasting.*

*Abridgment*

# Commuter Rail Diversion Model

Gary A. Gordon, City Engineer, Gloucester, Massachusetts
Thomas E. Mulinazzi, Department of Civil Engineering, University of Maryland

Over the past two decades commuter railroads, in general, have been experiencing declines in ridership that have led to a rather stable but low level of ridership. Alternative action therefore must be taken to recoup financial losses. These losses can be offset by reducing or eliminating commuter service, by obtaining outside funding, or by increasing commutation rates or the entire fare structure.

Railroads requesting a commuter fare increase either make no effort to determine the number of people diverted to other modes of transportation as a result of the fare increase or come up with unrealistic figures. Outside of a few general models for transit diversion in relation to fare increases (1, 2), little has been done to determine the relationship between fare increases and ridership for rail transit, not to mention commuter railroads.

The problem of not having a model that measures passenger diversion attributable to a given fare increase for commuter railroads is significant. Not only is the diversion figure important in the economic analysis of a fare increase, it is a key factor in determining environmental and transportation impacts of commuter service. Government policy in recent years has been advocating mass transportation as a mitigating measure to traffic congestion and hazardous levels of air pollution and more recently to energy consumption. Economically, an accurate diversion figure is needed to project passenger revenues and to determine whether, in fact, the fare increase is the solution.

The general objective of this paper is to present a model developed to determine the diversion of commuter rail users as a direct result of a fare increase and to compare this with other transit diversion models (3). Also, the model should be easy to understand because many of its users will be nontechnical people, many of whom represent railroads or transit agencies in a legal capacity. Therefore, the simplistic model resulting from the limited scope of this research will be beneficial. On the other hand, the model must be accurate enough to provide the user with a reasonable estimate of the diversion so that the resulting impacts on transportation, the environment, and revenues can be determined.

## CHICAGO MODEL

A model for Chicago, developed by the Interstate Commerce Commission (4), determined the functional relationship between the diversion of commuter rail passenger traffic to other modes of transportation and commuter fare increases by analyzing data provided by the six Chicago area commuter railroads. These data included monthly revenue passenger volumes for a period of at least 52 months that dated as far back as 1969 and rate increases that occurred during the same period. A computerized forecasting program (with seasonal variation capabilities) was used to determine the historic trend of each railroad with respect to growth or decline.

The linear function that resulted from the regression analysis is

$$Y = 0.52(X) - 3.68 \qquad (1)$$

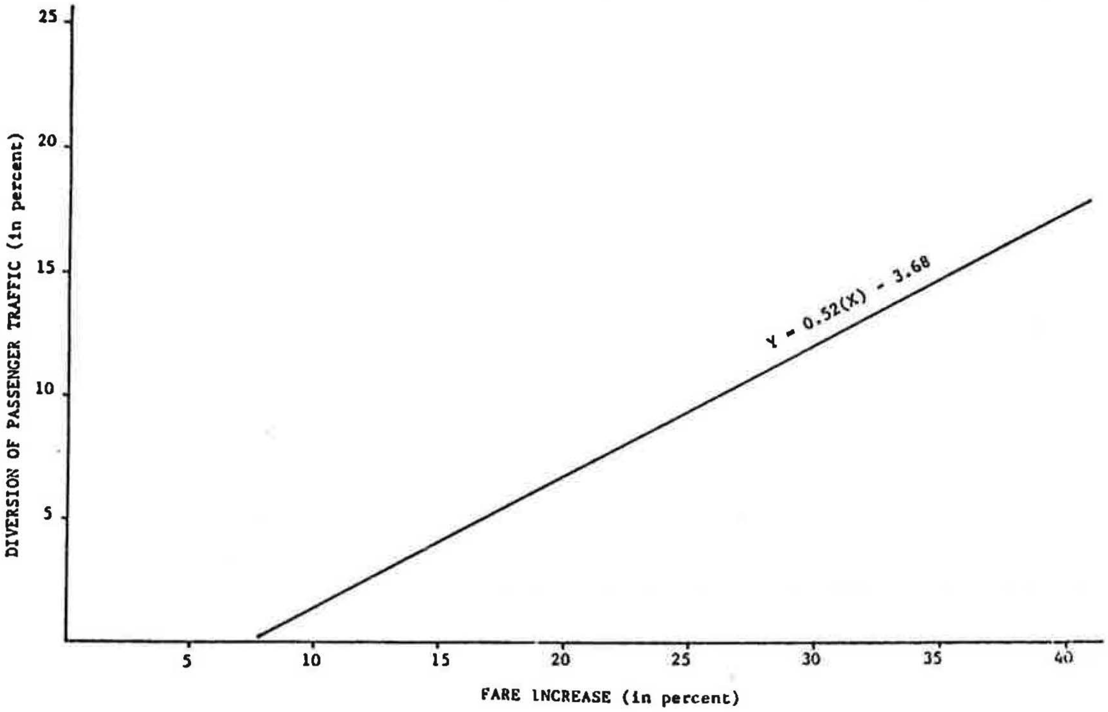
where X is greater than 7 percent and

X = fare increase in percent and
Y = diversion of passenger traffic in percent.

The relationship is graphically represented in Figure 1.

The analysis also revealed that there is no significant diversion for an increase in fares averaging 7 percent or less. For fare increases of greater than 7 percent, there is an expected 0.5 percent diversion for each additional 1 percent increase in fares in excess of 7 percent. For example, if a railroad increased its fares an average of 13 percent, it can expect a diversion of passenger traffic to other modes of transportation of approximately 3 percent.

Figure 1. Loss of ridership from fare increases on Chicago commuter rail lines.




## DATA COLLECTION

As an extension of the Chicago model, additional data from a sample of the commuter railroad systems in the United States were analyzed.

The group consisted of six major urban areas with commuter railroads and was intended to be representative of commuter rail service in this country. The six major urban areas surveyed were Boston, Chicago, New York City, Philadelphia, San Francisco, and Washington, D.C.

The survey encompassed 16 railroads that provide commuter rail service in these urban areas. The data analyzed included rate increases, monthly passenger counts, and any unusual circumstances such as strikes that might affect ridership. The study period was five years, beginning January 1, 1970.

## DETERMINATION OF DIVERSION

New passenger losses or gains were determined by forecasting ridership after a fare increase, using the relationship derived for the period prior to the rate increase. Ridership forecasts, in addition to fares, considered such variables as automobile operating costs, travel times, levels of service, the cost of alternate modes of transportation, etc., some of which were not used in the final analysis because of insufficient data and statistical insignificance.

A forecast of ridership for the six-month period following the fare increase was compared to the actual ridership for that period, and the difference yielded the net passenger loss or gain for that period. Diversion was then determined by factoring out that portion of the net loss or gain attributable to the historic trends developed previously. Fare increases were found to have a tendency to accelerate the rate of ridership losses and to reverse a positive trend. Diversion is, therefore, made up of those passengers directly affected by the fare increases who, after the six-month period, have not returned to the trains and are not explained in the historic growth rate of the railroad.

If, after this procedure, ridership increased during the six-month period, it was assumed that the fare increase was accompanied by service improvements or an external condition that could increase ridership and offset the fare increase (e.g., sharp rise in automobile operating costs). Because of time constraints and data limitations, further analysis along this line was not performed. However, the effect of service improvements and/or increased costs of other modes of transportation on fare increases warrants further study.

## DIVERSION CURVE

Eighteen usable data points resulted from the analysis of the data. These data points show the relationship between a specific fare increase and the resulting diversion. The next task was to manipulate the data pairs in such a manner that a general relationship between fare increases and diversion resulted. Since regression analysis is a proved method of curve fitting, it was also used in this portion of the analysis.

The two sets of data analyzed separately by linear regression resulted in the following relationships: for time factor included (period between fare increases considered)

$$\mathrm{DIV} = 0.3(\mathrm{FARE\ INC}) + 0.5 \tag{2}$$

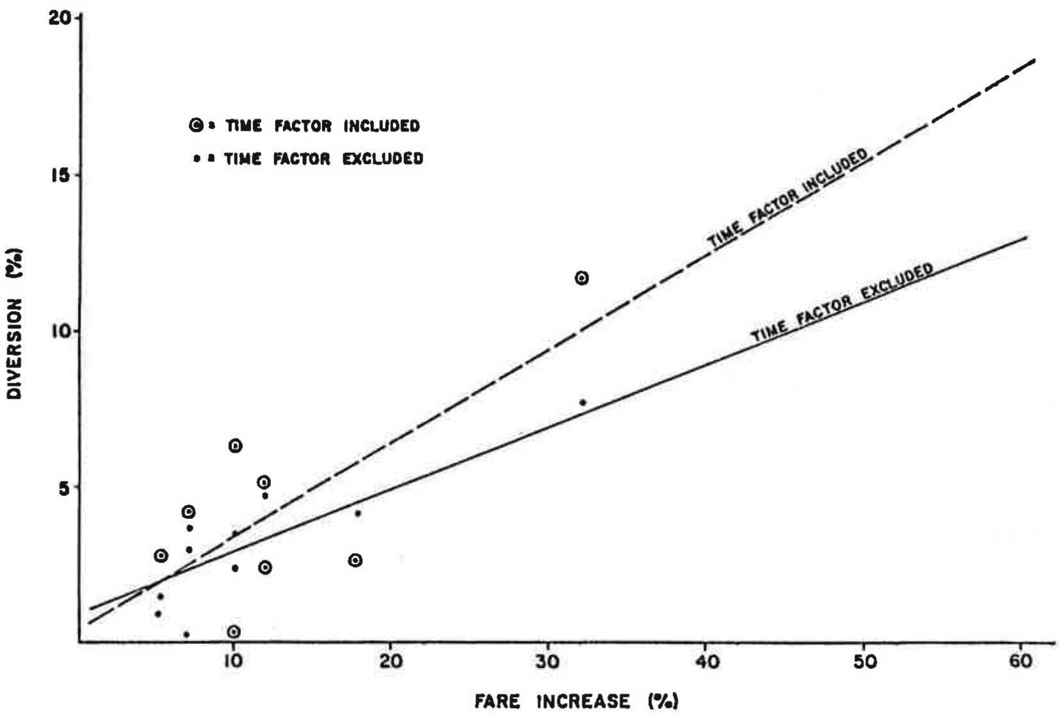
and for time factor excluded

$$\mathrm{DIV} = 0.2(\mathrm{FARE\ INC}) + 1.0 \tag{3}$$

where DIV is diversion in percent, and FARE INC is fare increase in percent. A graphical representation of these relationships is shown in Figure 2. Once again it should be noted that other methods of curve fitting were considered, but the linear regression provided the best fit.

Rather than using the separate relationships, we derived a single relationship representative of both. It is

Figure 2. Diversion of commuter rail ridership from fare increases.




simply an equation averaging the coefficients of both; it defines a line midway between the two. The resulting equation is to be used as a general, nationwide relationship relating diversion (in percent) to fare increases (in percent). The diversion equation is

$$DIV = 0.25(FARE\ INC) + 0.75 \qquad (4)$$

The derived relationship shows that for every 10 percent increase in fares a loss in ridership of 3.25 percent will occur.

## CONCLUSIONS

The diversion model developed in this paper presents a general technique for evaluating the impact of a fare increase on commuter rail ridership. Although the diversion rate may differ from railroad to railroad and from urban area to urban area, the model can be used to determine a general diversion value applicable anywhere in the country. Furthermore, the methodology provides a framework in which railroads and transit properties, or any other agency, can develop their own relationships or assess the impact of a specific rate increase.

As stated previously, the methodology was constrained by the availability of data, time, and the capability of the computer system used. Nonetheless, a thorough investigation has been made using the available resources. The result is, as stated previously, a methodology and simplistic model that can be used nationwide to delineate the relationship between fare increases and concomitant ridership losses. Based on the analyses performed, it can be concluded that

1. A ridership loss of approximately 3 percent can be expected for every 10 percent increase in commuter rail fares,
2. Although commuter rail riders and other transit riders display different user characteristics, they appear to be affected similarly by fare increases,
3. An increase in commuter fares, when accompanied by an increase in service, does not necessarily result in ridership losses, and
4. User and system characteristics differ throughout the country.

It should be noted that the assumption that passenger gains occurred as a result of service improvements was made because of the limited scope of this research and in an effort to simplify the analysis.

It is important that the results of this research be meaningful and consistent with existing concepts. The relationship developed found ridership to be fare inelastic. The fare elasticity was calculated to be -0.325, which approximates that of the generally accepted fare elasticity of -0.3. The similarity of the fare elasticities can be attributed to the model's being developed by using proved analysis procedures for forecasting. Also, transit users are not that unlike, regardless of transit mode. Furthermore, the approach to the problem and types of data selected for analysis are not any different from those used for other fare studies on other transit modes.

Most of the research related to fare elasticities was done for either bus or rail rapid transit. Consequently, and for lack of a better alternative, these relationships were used in the past for commuter rail. By developing this diversion model for commuter rail, nothing new has been proved that would revolutionize fare studies, but it did prove that the generally accepted relationship can be used for commuter rail with confidence.

## REFERENCES

1. J. F. Curtin. Effect of Fares on Transit Riding. HRB, Highway Research Record 213, 1968, pp. 8-18.
2. E. A. Harvey. Impact of Fare Change on Railroad Commuter Ridership. HRB, Highway Research

Record 213, 1968, pp. 35-41.
3. G. A. Gordon. The Impact of Fare Increases on Commuter Rail Ridership: A Diversion Model. Univ. of Maryland, master's thesis, 1977.
4. G. A. Gordon. Final Environmental Impact Statement—Illinois Central Gulf Railroad Company—Electric Commuter Train Fares, Docket No. 35889. Interstate Commerce Commission, 1975.

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

*Abridgment*

# Impact of the Relative Transit and Highway Service Levels on Trip Distribution

W. Thomas Walker, Delaware Valley Regional Planning Commission

The purpose of this investigation is to measure the impact of the public transit service level on the destination choices of trip makers. It is often hypothesized that trip makers will have a tendency, first, to make more trips to areas with a relatively high level of public transit, particularly if the service there is superior to that provided by the auto and, second, to make fewer trips to areas with poorer transit accessibility. This propensity is measured by comparing the error in travel volumes predicted by a standard Bureau of Public Roads highway time gravity model with the relative transit and highway service levels, as measured by the disutility difference measure used in most utilitarian modal split models. A well-defined and logical bias in gravity model output was discovered with respect to the relative transit and highway service levels. The impact of this bias on simulated person trips is evaluated by correcting the gravity model output and comparing the corrected and uncorrected trip tables.

## GRAVITY MODEL TRIP DISTRIBUTION

Doubly constrained gravity models were calibrated on the basis of highway travel times for three trip purposes: home-based work, home-based nonwork, and non-home-based nonwork (1). The formulation of the models was the standard Bureau of Public Roads (BPR) gravity model (2). For the most part, the procedures outlined in that report were followed during the calibration process.

The formulation of the BPR gravity model is

$$T_{ij} = P_i A_j F_{ij} \Big/ \sum_{J=1}^{n} A_j F_{ij} \qquad (1)$$

where

$T_{ij}$ = number of trip interchanges from zone i to zone j,
$P_i$ = number of trip productions in zone i,
$A_j$ = number of attractions in zone j, and
$F_{ij}$ = empirically derived highway travel time factor, which expresses the overall areawide effect of spatial separation on trip interchanges between zones $t_i$ minutes apart. This factor approximates $1/t^n$.

Specific zone-to-zone adjustment factors, or K-factors, were not used in this application of the BPR gravity model.

The trip data used to calibrate the gravity model were obtained from 1960 Penn-Jersey travel survey data, which were reformatted into standard production-attraction format and trip tables built for each purpose on the basis of an 832-zone area system.

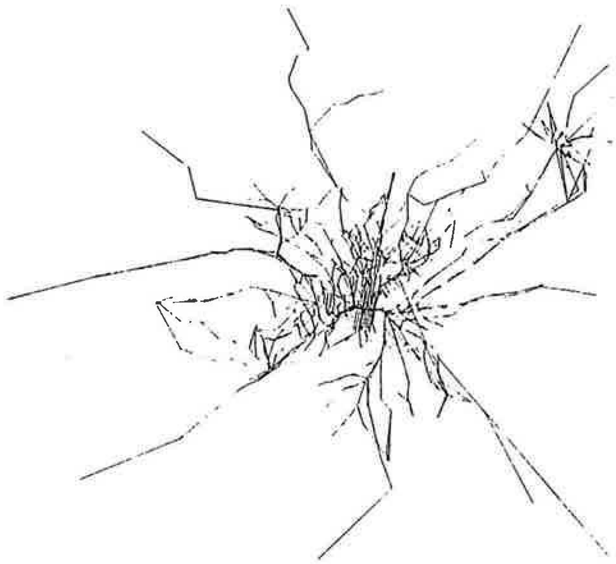
Highway travel times were obtained from a 1960 street and highway network, which was coded to the same zone system. Highway speeds were inserted into the network from a look-up table on the basis of functional class and area type. The highway travel times were then updated with terminal and interzonal times and with bridge penalties across the Schuylkill and Delaware rivers. These parameters were calibrated by using recommended BPR procedures. The updated set of highway travel times was used for all three trip purpose models.

## BIAS WITH RESPECT TO RELATIVE TRANSIT AND HIGHWAY SERVICE LEVEL

### The Test

If the public transit service level has a measurable impact on the distribution of person trips, then a distribution solely on the basis of highway travel time should lead to an underestimation of person trips for interchanges with good transit service and poor highway service, and should lead to overestimation where transit service is poor relative to highway service. This hypothesis was tested by comparing the relative transit and highway service levels (as measured by the disutility or impedance difference measure shown in Equation 2) with the ratio of the 1960 gravity model synthetic trips to 1960 survey person trips. This comparison was done for each of the three trip purposes and for total trips.

Figure 1. Coded transit network.



$$ID = K_1(TE - HE) + K_2(TR - HR) + K_3(TF - HOP - PKG) + K_4(TRFR + 1) + 200 \quad (2)$$

where

- ID = disutility or impedance difference,
- HE = highway out-of-vehicle time,
- TE = transit excess or out-of-vehicle time,
- HR = highway in-vehicle time,
- TR = transit in-vehicle time,
- TF = transit fare (cents, in 1960 dollars),
- HOP = highway operating cost (in 1960 dollars),
- PKG = auto parking cost,
- TRFR = number of transit transfers,
- $K_1$ = 2.50,
- $K_2$ = 1.67,
- $K_3$ = 1.0, and
- $K_4$ = 16.0.

This impedance difference is similar to the standard disutility measure used in most modal split models (3, 4).

The 1960 transit travel times and costs were obtained from a morning peak-hour transit network with travel times and headways taken from operating schedules. The network as it existed in 1960 is shown in Figure 1. Coded to the same 832-zone area system as the highway network, it contained all significant commuter rail, subway-elevated, and bus facilities within the 1960 Penn-Jersey cordon line. The same morning peak-hour network was used for all three trip purposes.

### Results

The ratio of 1960 synthetic to 1960 survey trip interchanges was plotted versus the transit-highway disutility difference for home-based work, home-based nonwork, non-home-based, and total trips. The curves clearly showed a systematic bias in the magnitude of synthetic trip interchanges with respect to the relative transit and highway service levels. This bias exists for all three trip purposes; its magnitude is greatest for home-based work trips and least for non-home-based trips.

## INCLUDING THE TRANSIT SERVICE LEVEL

### Correction Procedure

Several approaches are available for attempting to correct the bias with respect to the relative highway and transit service levels. The most common approach is to construct a combined interzone time or impedance measure and then to calibrate the gravity models on this basis. An appealing way to accomplish this is to construct a weighted average of the highway and transit service levels, using some function of the percentage of transit as a weighting factor. However, this approach is difficult to calibrate, and most studies simply assume an arbitrary weighting scheme.

Rather than adopt an arbitrarily calibrated formulation that would use an estimated modal split to weight the highway and transit travel times, the inverse of the bias curve was used to calculate an adjustment factor that would then be translated into a revised highway travel time through the inverse of the gravity model friction curve. As was shown in the previous section, the difference in bias curves for each trip purpose was only marginal. Therefore only the total-purpose curve was used to adjust the travel times; this resulted in one revised travel time matrix for all three trip purposes. The inverse total person trip bias curve shown in Equation 3 was fitted by least squares.

$$Z^{-1} = 1.299 - 0.000\,87(ID) \quad (3)$$

The coefficient of determination of the above equation was 0.64. In estimating bias corrections, $Z^{-1}$ was constrained to vary between 1.2 and 0.8.

It should be noted that this process is similar to the more usual practice of weighting the travel times with respect to the percentage of transit, since the bias is measured with respect to a disutility or impedance difference similar to the relative service measure used in most post-distribution modal split models.

However, it is more appropriate for two reasons. First, it is based on an explicit measurement of the bias with respect to the transit service level and hence was calibrated with base-year data. Second, it does not require recalibration of the existing highway time-based trip distribution model, which was performing reasonably well.

### Impact of the Combined Skim Adjustment on Person Trips

The combined skim adjustment was applied to the estimation of 1977 person trips for the nine-county Delaware Valley region, and the results were compared with the output of the gravity models by using a set of highway interzone travel times. When the resulting differences were aggregated to superdistricts, the average change was approximately 16 percent of the mean trip interchange volume. Spatially, the combined skim adjustment reduced circumferential movements, which had a poor level of transit service relative to highway; it also increased radial movements, which had relatively good transit service.

The combined skim adjustment tended to increase the average trip length in high-speed rail corridors because these facilities provide generally good service relative to the automobile for longer movements but poor service for trips involving short interstation movements.

## CONCLUSIONS

After examining the model results, I have drawn the following conclusions.

1. A highway-based gravity trip distribution model has a measurable bias in the Delaware Valley region with respect to the relative public transit and highway service levels.
2. This bias is well defined, rational, and statistically significant for home-based work, home-based nonwork, and non-home-based trips.
3. The bias varies only marginally by trip purpose; only the non-home-based trips are significantly different from home-based work trips and total trips.
4. The highway time-based gravity model has a significant tendency to underestimate person-trip interchanges even when the transit and highway service levels are equal.
5. The correction of the bias results in significant changes in the synthetic person-trip tables. This change is primarily a shift of person trips from circumferential corridors with poor transit to radial corridors with relatively good public transportation service.

The above results were obtained for the Delaware Valley region, which has an extensive public transportation system—some 2900 route kilometers (1800 route miles) of surface transit service and some 1100 route kilometers (700 route miles) of high-type rail facilities. However, the basic conclusions can probably be generalized to other regions that now have or are considering some form of high-speed public transit service, because the total amount of transit service may not be as significant as the relative quality of transit service in individual corridors.

## ACKNOWLEDGMENTS

This paper was financed in part by the Federal Highway Administration and the Urban Mass Transportation Administration of the U.S. Department of Transportation and by the Pennsylvania and New Jersey departments of transportation. The contents reflect my views, which do not necessarily reflect the official views or policy of the funding agencies.

## REFERENCES

1. Delaware Valley Regional Planning Commission. Calibration of the Person Trip Distribution Models. Philadelphia, Technical Rept. No. 4, 1975.
2. Calibrating and Testing a Gravity Model for Any Size Urban Area. Bureau of Public Roads, Oct. 1965.
3. G. A. Shunk and R. J. Bouchard. An Application of Marginal Utility to Travel Mode Choice. Paper presented at the 49th Annual Meeting, HRB, Jan. 1970.
4. Delaware Valley Regional Planning Commission. Calibration of the Mode Choice Models. Philadelphia, Technical Rept. No. 10, 1975.

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

*Abridgment*

# Automobile Availability per Worker: A Transportation-System-Sensitive Socioeconomic Variable

R. Ian Kingham, National Cooperative Highway Research Program

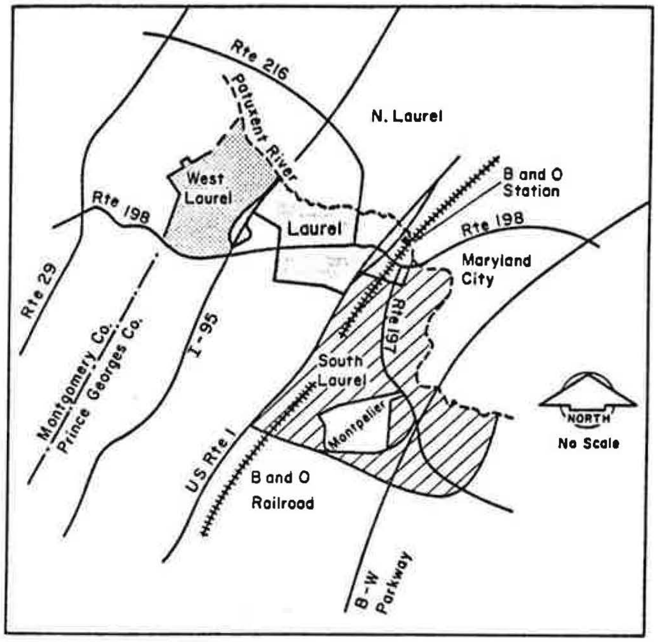
Environmental and energy impacts of transportation are related to vehicle-kilometers of travel. To reduce vehicle-kilometers traveled, strategies are needed to attack its two components: the number of vehicles and the distance the vehicles move. Transit has been suggested as an alternative to driving the automobile to work (thereby, presumably, leaving the automobile parked at home) and as an alternative to owning a second or third car. The research reported in this paper was an exploration of possible relationships between transit and automobile ownership and a determination of causality if such relationships were found (1).

## RESEARCH OBJECTIVES AND APPROACH

The general objective of the research reported in this paper was to investigate the impact of a viable transit alternative on household decisions to have automobiles for use in making home-based trips. The specific objectives were, first, to determine differences in automobiles available per worker (APERW) between households in areas served by transit and similar areas not served by transit and to determine causality, and, second, to recommend socioeconomic variables that appear to have high correlation and possible causal effects on APERW for consideration in travel demand models.

Automobile availability per worker was used in this research rather than car ownership or car availability because of findings from other completed or ongoing research. In recent years there has been general agreement among travel demand forecasters that car ownership should be replaced by car availability in mode-choice models (2). It is argued that mode choice and, in fact, travel behavior in general are influenced more by the cars available to a household than by the cars owned by the household. Company cars and rental cars

Figure 1. Greater Laurel study area.




are included in the more general term "car availability." Charles River Associates (3) further modified car availability by dividing it by the number of employed workers in the household and noted that APERW could reflect competition for the automobile within households. Therefore, it was hypothesized that APERW would correlate more strongly with mode choice and would be more strongly affected by the presence of transit.

## STUDY SETTING

The study was carried out in Laurel, Maryland, which has a population of 48 000 and an area of 23 $km^2$ (9 miles$^2$) and is located approximately midway between Baltimore and Washington, D.C. Figure 1 shows the principal subdivisions of "Greater Laurel." Approximately one-half of the population resides in South Laurel. It is the West Laurel, Laurel, and South Laurel areas that are important to this research because of socioeconomic similarities but vastly different transit service. Within South Laurel, varying density permits an investigation of any transit service and household-type interaction on APERW.

The transit service hypothesized to influence APERW is an express commuter service operated by Greyhound, using intercity coaches, from Laurel to Washington, D.C. Buses depart from the city and proceed along US-197 at 10-min intervals from 6:30 to 8:00 a.m. Similar service is offered during the afternoon peak period. Principal office areas in Washington are served directly. Running time per trip is 35-45 min (minor preferential treatment); cost is $0.825. It is estimated that the service carries 525 passengers in each direction per day.

Other transit service is provided by the B&O Railroad, which supplies three trains per day in each direction during peak periods to employment centers in the vicinity of Capitol Hill. Running times are less than for competing bus or car, but the remoteness of the Washington station from the office complexes forces many to transfer to other transportation modes to reach employment destinations. Trip fares by rail vary from $0.89 to $1.60, depending on whether a commitment is made to a monthly or a weekly pass.

Automobile competition with transit is substantial. One limited-access highway (Baltimore-Washington Parkway) and a primary highway (US-1) enter the city of Washington.

## DATA BASE

Ninety-two percent of all households in the study area were given a questionnaire to complete and return to convenient pick-up areas. In addition to socioeconomic and sociodemographic questions, respondents were asked to identify all work trips and the modes taken for all family members.

The response rate, considering all households in the study area, was 4.2 percent. For this small response, it could be expected that biases would exist in the data set because of under-reporting or over-reporting of various population segments. Thus, a study was made of the households that responded in each of the study areas. Population by dwelling type and the distribution of households by dwelling type were compared with the most recent estimates prepared by the Maryland National Capital Park and Planning Commission (MNCPPC). Also, distributions of population ages and incomes were compared with updated census data. Last, the response rate of bus and train households was determined for comparison with observed ridership.

The comparisons showed that biases existed as follows:

1. Single-family households responded at a greater rate than did either apartment or townhouse households;
2. Higher-income households responded at a greater rate than did lower-income households;
3. Households with adults in the 30-50 age group responded at a greater rate than did other households; and
4. Households with bus and train work trips responded at a greater rate than households having only automobile work trips.

Income, dwelling type, and choice of mode are variables thought to be most important to automobile availability. Bias from income and dwelling type were considered by segregating the data on those bases. The bias arising from mode split required the following weighting factors to be used for all households: (a) households with auto work trips only, 1.41; (b) households with 1 or more bus work trips, 0.34; and (c) households with 1 or more train work trips, 0.31.

## FINDINGS

The finding of others, that the variable automobiles-available-per-worker is strongly related to mode choice, was corroborated.

For work trips to Washington, D.C., the percentages carpooling or using transit greatly decreased with increasing automobiles available per worker. Figure 2 shows the mode choice of trip makers at various levels of APERW. Its effect was larger than any other socioeconomic variable investigated in this study. Variables investigated in addition to APERW were dwelling type (apartment, single-family home), income, life cycle (number of adults, existence of children), and sex of traveler. The influence of all these variables was small or negligible once the influence of APERW was accounted for.

A significant but small reduction in APERW appears to be the result of establishment of express bus service

Figure 2. Mode choice for work trips by automobiles per worker.

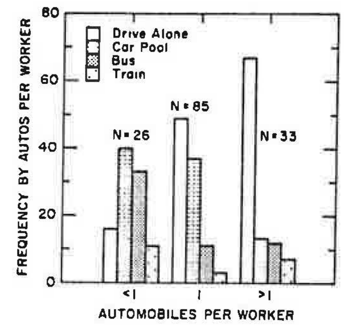
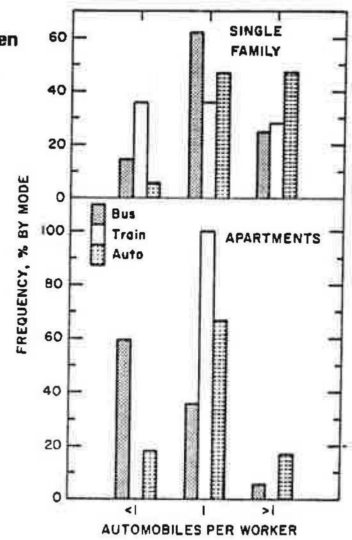



Figure 3. Differences in automobiles per worker between transit and automobile-only households.




for households having (a) Washington, D.C., work-trip destinations and (b) a free choice.

Households having a free choice were defined as those having 1975 household incomes in excess of $15 000 and having two or more adults. The income criterion reflects the ability of a household to own more than one car. For incomes above $15 000, income was found to have negligible effect on APERW. The number of adults criterion identifies a market having two or more automobiles and, hence, the flexibility to give up an automobile and still remain with at least one. The one-adult household usually had one automobile and was apparently reluctant to give it up for a transit service that served only the peak-period work trip.

Comparisons of APERW for bus, train, and automobile households are made separately for single-family households and apartment households in the discussion that follows. This was the only variable, other than income and number of adults, observed to have an effect on the APERW-transit relationship. Figure 3 shows the distribution of APERW for bus, train, and auto households. An auto household was defined as one having no bus or train work trips. The difference between bus and automobile-only households was statistically significant. The probability of chance variation for single-family homes was 7.5 percent and for apartments 1.4 percent.

A study of the data revealed that a high percentage of bus trips were from households with easy access to the bus, i.e., in Laurel and South Laurel. Therefore, it was hypothesized that a comparison of households having D.C.-bound work trips would reveal a lower APERW for Laurel and South Laurel households. Again, for the free choice market a statistically significant difference was observed for both single-family homes and apartments. The probabilities of chance variation were 20 percent and 23 percent respectively, indicating a weak relationship. A control group, however, showed a chance variation of 56 percent. The mode split for D.C.-bound trips was 13 percent by bus, 4 percent by train, and 83 percent by automobile. The low transit split tends to explain the 20 and 23 percent chance variations.

The findings to this point assume that no other effects could be identified that would explain the differences in the distribution of APERW between the two areas. There was a possibility however, that certain urban structure characteristics (such as the closeness of households to employment areas) could be responsible for the differences. Where walk trips are feasible for work and shopping trips, lower APERW could be anticipated. These factors were investigated and could be accounted for. Removing the city of Laurel data, where walk trips did occur, did not change the differences observed.

Inasmuch as the household survey was conducted at only one point in time and the survey instrument did not inquire of bus riders whether they sold their second or third automobiles because of the existence of bus service, it is pure conjecture at this point that transit was responsible for the reduction in car availability. People could have moved into the area because of the transit service and their desire to own fewer automobiles. Ferrari and Shindler (4) and Dunphy (5) also have found that automobile ownership rates varied with the relative level of service provided by the transit and highway systems. However, neither study could determine causality.

We explored the causality question by examining the survey forms for bus riders to determine how the respondents answered the question on transit experience. Specifically, the form asked Have you ever used public transportation? and then Where? This allowed each household to be rated on previous transit experience. Ratings were high, medium, low, and none. It could be anticipated that if the majority of bus households fell into the high experience category, people had moved into the area because of the bus service. If, however, the majority had no previous transit experience, it can be inferred that but for the existing Greyhound service these households would have the same automobile per worker characteristics as households making work trips to other destinations. High transit experience was attributed to households claiming experience from cities having rapid transit, medium experience to households claiming experience with large bus systems, and low experience to those who professed experience with small-town bus systems. No experience was attributed if only the Greyhound service between Laurel and D.C. was noted. The results of this assessment are given below.

| | Households Having Transit Experience | | | |
|---|---|---|---|---|
| | Medium to High | | None to Low | |
| Locale | No. | Percent | No. | Percent |
| City | 4 | 27 | 11 | 73 |
| West Laurel | 4 | 40 | 6 | 60 |
| South Laurel | 15 | 39 | 23 | 61 |
| All | 23 | | 40 | |

These results suggest that the majority of bus riders have had little experience and, therefore, could be expected to have sold their second and/or third cars when they became regular bus riders.

To this point, variables affecting free choice households have been described. For those households of lower income there was a predominant trend (74 percent) to one automobile per worker. Households having less than one APERW were principally apartments and townhouses (82 percent); households having more than one APERW were principally single-family homes (78 percent). A trend of increasing APERW with greater numbers of adults and larger households was evident. An influence of transit could not be identified, possibly because very few lower-income households had members working in D.C. and were hence served by transit. Therefore, an APERW model for this subgroup should include the variability of household size in addition to dwelling type and, possibly, transit availability.

## APPLICATION OF FINDINGS

In mode-choice models for work trips, APERW is a variable preferred over auto ownership or auto availability because (a) it is intuitively better, in that it reflects competition within a household for the automobile, and (b) the research findings show its correlation with mode choice to be statistically equal to or better than that of car availability. Group quarters or extremely large households can be accommodated in the data base without statistical analyses.

APERW is influenced by (a) the presence of a transit alternative, (b) the household composition (specifically the number of adults), (c) income, (d) dwelling type, and (e) household size (in the case of low-income households). For small satellite-type urban areas and suburban areas it is recommended that APERW models be estimated separately for three subgroups of the population, as follows:

1. Households with two or more adults and (1975) incomes in excess of $15 000,
2. Households with two or more adults and (1975) incomes less than $15 000, and
3. Households with one adult.

The first group has a free choice. Households in the group are able to afford more than one car and, depending on household composition, have need for more than one car unless travel needs can be satisfied by public transportation. The potential exists for APERW to be estimated using only dwelling type and transit level-of-service variables. The second group can ill afford more than one car. The study has shown that APERW models for the second group should consider household size, dwelling type, and perhaps transit service. The third group has a need for only one automobile. Thus, in most instances, it can be assumed that a one-adult household will have one automobile.

Provision of a high level of bus service will have some impact on reducing APERW for subgroups 1 and 2, but the impact is small even if transit is available to all employment destinations. Therefore, transit service alone cannot be an effective strategy in reducing automobile availability per worker.

## ACKNOWLEDGMENTS

Acknowledgment is given to Mr. R. Pratt of R. H. Pratt and Associates, who guided the research; to Mr. D. McGrath of George Washington University, who critiqued manuscripts; to Drs. J. Margolin and M. Miche also of George Washington University, who assisted in the design of the questionnaire; and to Mr. M. Petrenko of the city of Laurel, Maryland, for assistance in the printing and distribution of the questionnaire.

## REFERENCES

1. R. I. Kingham. Automobile Availability Per Worker: A Transportation-System-Sensitive, Socioeconomic Variable Related to Mode Choice. Department of Urban and Regional Planning, George Washington Univ., master's thesis, 1976.
2. P. R. Stopher and A. H. Meyburg. Urban Transportation Modeling and Planning. Heath, Lexington, MA, 1975.
3. Charles River Associates. Disaggregate Travel Demand Models. NCHRP, Agency Phase I Rept., Project 8-13, 2 Vols., 1976.
4. M. G. Ferrari and R. Shindler. Auto Ownership as Affected by Transportation System Alternatives. Traffic Engineering, Oct. 1967, pp. 24-28.
5. R. T. Dunphy. Transit Accessibility as a Determinant of Automobile Ownership. HRB, Highway Research Record 472, 1973, pp. 63-71.

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

*Abridgment*

# Trucks in the Traffic Assignment Process

John R. Hamburg, John R. Hamburg and Associates, Inc., Bethesda, Maryland

In order to deal with truck impacts, separate truck assignments are necessary so that the truck portion of the traffic stream on a highway link may be identified.

In some cities there are highways, parkways for example, on which trucks are specifically excluded. Then there are structures (e.g., overpasses and underpasses) whose very physical characteristics bar vehicles over a given height or weight. Conversely, there are routes whose signing encourages through truck travel. In dense urban areas, a separate assignment network should be built for trucks to reflect the policies and prohibitions on trucks. Moreover, if no such policies exist, separate network capacity should still be included within the assignment because such capability is necessary to test and evaluate the utility of such policies.

There are distinct differences among trucks in terms of size, weight, acceleration characteristics, number of axles, number of wheels, etc. Two classes of trucks are recommended. One is "auto-like" trucks, and these should be lumped in with autos. The second class would be heavy trucks and would be assigned separately. The criterion to be used is that auto-like trucks have four tires and heavy trucks have six or more tires.

A zone-to-zone trip table of truck trips is required to pursue separate truck assignments. Standardized procedures have been developed by FHWA to synthesize a table if an origin-destination survey is not available.

In addition to a truck trip table, the user should have some ground counts of truck volumes so that he or she can verify the simulation of routing and also generation and distribution, if this is required. Ideally, the truck link counts should be extensive enough to give a truck vehicle-kilometers of travel (VKMT) estimate against which simulated truck VKMT can be compared as a control.

Most existing assignment programs permit only one class of vehicles—autos, trucks, autos and trucks, or autos and truck "auto equivalents"—to be assigned. Assigning autos first and then trucks gives autos the preferred paths and tends to route trucks over longer paths if congestion is present and a capacity restraint mechanism is used. Concurrent assignment of autos and trucks with a separate account of truck volumes on links avoids the problem.

## IMPACT OF TRUCKS ON HIGHWAY CAPACITY

### Simplified View of Capacity

A simplified approach to the capacity of signalized intersections can be taken by finding the time required for the nth vehicle in a queue to clear the intersection and equating this to the green time. This approach is the basis for capacity determination in the microassignment process, in which all vehicles in the queue are assumed to be equally spaced and to accelerate uniformly to a specified average velocity approximately equal to the speed limit.

There are two (mutually exclusive and exhaustive) cases to be considered:

1. No vehicle in the queue reaches the specified velocity by the time it clears the intersection.
2. One or more vehicles in the queue reach the specified velocity before clearing the intersection.

In the following formulations, which were written in customary units, let

$t(n)$ = time for nth vehicle to reach the intersection,
$C_1$ = reaction time of the first vehicle (seconds),
$C_2$ = reaction time of other queued vehicles (seconds),
$D$ = spacing of vehicle (feet),
$V$ = free flow velocity (feet per second),
$a$ = acceleration rate,
$N_2$ = the number of the queue position from which a vehicle can accelerate to free flow velocity by the intersection, presuming the green time allows it ($N_2$ - $V^2/2Da$),
$t$ = time spent accelerating (seconds), and
$t_v$ = time spent traveling at terminal velocity (seconds).

Since the nth vehicle in the queue cannot start moving until the vehicle immediately ahead of it moves, the time for the nth vehicle to clear the intersection is the sum of the reaction time of the first vehicle, the reaction time of the following n - 1 vehicles, and the time for the nth vehicle to travel to the intersection.

For the first case this is

$$t(n) = C_1 + (n - 1) C_2 + t \tag{1}$$

But, since the average velocity of the nth vehicle in this case is $at/2 = nD/t$, from which $t = \sqrt{2nD/a}$, then

$$t(n) = C_1 + (n - 1) C_2 + \sqrt{2nD/a},\ n < N_2 \tag{2}$$

For the second case, where one or more vehicles reach average velocity before clearing an intersection ($n > N_2$), we must add a term to Equation 1 that accounts for the time spent traveling at terminal velocity by the nth vehicle. Therefore, for case 2

$$t(n) = C_1 + (n - 1) C_2 + t_v + V/a \tag{3}$$

But

$V/a$ = the time to reach V and
$V/2$ = the average velocity during t.

Therefore $(V/a)(V/2) = V^2/2a$ = the distance covered in t, so that $t_v = [nD - (V^2/2a)]/V$. Finally, substituting for $t_v + (V/a)$ in Equation 3, we get

$$t(n) = C_1 + (n - 1) C_2 + (nD/V) + (V/2a);\ N > N_2 \tag{4}$$

The first vehicle in the queue to reach terminal velocity at the intersection ($N_2$), providing that there is sufficient green time, can be found by dividing the distance required to achieve terminal velocity by the average spacing of vehicles in the queue: $N_2$ = distance/spacing = $(V/2)(V/a)(1/D) = V^2/2aD$. Therefore, the time required for this vehicle ($N_2$) to clear the intersection is

$$t_{(N_2)} = C_1 + [(V^2C_2)/2aD] - C_2 + (V/a) \tag{5}$$

As an example, if we let $C_1 = 1.9$, $C_2 = 1.4$, $V = 44$ ft/s, $a = 4.4$ ft/s$^{-1}$, and $D = 25$, then substituting these values in Equation 5 gives $t(N_2) = 22.82$ s. Thus, for green times in excess of 23 s, given the above constants, Equation 4 would be utilized. However, for green times in the range of 15-23 s, Equation 4 gives essentially the same results as Equation 2, and therefore the following discussion will be in terms of Equation 4, which carries the assumption that the traffic stream will reach the terminal velocity during the green phase.

Equation 4 can be expressed as a linear function of the form $t(n) = K + bn$, by setting $b = C_2 = D/V$ and $k = C_1 - C_2 + V/2a$. Then, by substituting the values given above, we get $t(n) = 2n + 5.5$. This can also be thought of as an expression for capacity. For instance, if an hour is divided into cycles of duration C (seconds) and green time of G (seconds), the hourly capacity of a lane of through traffic is

$$\text{Capacity} = (G - k)/b \times 3600/C = 1800\,[(G - 5.5)/C] \tag{6}$$

Note that, if the green time is 100 percent of cycle time and cycle time is one hour (approximately free flow conditions), the lane capacity is about 1800 vehicles per hour, which seems about right.

Using Equation 6, the hypothetical hourly capacities for different signal splits and cycle lengths would be as shown in the table below.

| Green Time (s) | Cycle Length (s) 60 | 80 | 100 |
|---|---|---|---|
| 30 | 735 | 551 | 441 |
| 40 | 1035 | 776 | 621 |
| 50 | 1335 | 1001 | 801 |

Turning to commercial vehicles, we can estimate their capacities given a homogeneous stream. In the next table, hypothetical acceleration rates to a terminal velocity of 30 mph are given for automobiles, single-unit trucks, semitrailers, and buses. Also shown are recommended maximum lengths (AASHO) and an estimate of spacing for each of several vehicle types.

| Vehicle Type | Acceleration Rate (mph/s) | Maximum Length (ft) | Estimated Spacing (ft) |
|---|---|---|---|
| Auto | 3.0 | N/A | 25 |
| Single-unit truck | 1.5 | 40 | 45 |
| Semitrailer | 1.0 | 55 | 60 |
| Other combination | 1.0 | 65 | 70 |
| Bus | 2.5 | 40 | 45 |

By using the table above, we can calculate the capacity coefficients for trucks and buses in the same form as automobiles. The next table gives the capacity coefficients by vehicle type.

| Vehicle Type | Terminal Velocity - 30 K | b |
|---|---|---|
| Auto | 5.50 | 1.97 |
| Single-unit truck | 10.50 | 2.42 |
| Semitrailer | 20.95 | 2.76 |
| Other combination | 20.95 | 2.99 |
| Bus | 10.50 | 2.42 |

By assuming that the ratio of green time to cycle length equals 0.5 and that cycle time is 60 s, the hourly capacities shown below are obtained.

| | Hourly Capacity | | Ratio of Autos to Trucks | |
|---|---|---|---|---|
| | Cycle Time | | Cycle Time | |
| Vehicle Type | 60 s | 90 s | 60 s | 90 s |
| Auto | 746 | 802 | 1.00 | 1.00 |
| Single-unit truck | 483 | 570 | 1.54 | 1.41 |
| Semitrailer | 197 | 349 | 3.78 | 2.30 |
| Other combination | 182 | 322 | 4.10 | 2.49 |
| Bus | 483 | 520 | 1.54 | 1.41 |

The above exercise points up the sensitivity of truck capacity to the assumed acceleration rate, which, of course, varies by truck size, load, and grade. Also, the ratio of auto to truck capacity gives a measure of equivalence ranging from 1.4 to 4.1, depending on truck size, acceleration assumed, and signal length.

## Estimating the Capacity for Mixed Autos and Trucks

### Weighted Capacities

In the last table above, hourly capacities for each vehicle type and two signal splits are shown. One way to estimate the capacity of a mixed stream would be to weight auto capacity by the proportion of automobiles in the stream and truck capacity by the proportion of trucks in the stream. For example, assuming truck capacity to be six per cycle and 90 percent of the traffic stream, weighted capacity would be 11.4 ($0.1 \times 6 + 0.9 \times 12 = 11.4$).

### Converting to Auto Equivalents

Alternatively, since in the example above trucks have only one-half the capacity of autos, we could convert to auto equivalents by multiplying each truck by two. If we continue to assume twelve autos per signal cycle, we can calculate the vehicle capacity for different mixes. For example, a fifty-fifty split of autos and trucks would give eight vehicle equivalents for the four trucks and the four autos, or a vehicle capacity per cycle of eight.

### Analysis of Queue Composition

By assuming that trucks are distributed in the traffic stream, the number of autos per truck can be obtained by the ratio of the proportion of autos to the proportion of trucks. Thus, any given proportion of trucks in the traffic stream can be thought of as a queue of vehicles containing one truck and $P_A/P_T$ autos. To assess the impact of the presence of the truck on capacity, we need simply calculate the capacity for that queue for each of the different positions of the truck in the queue; the mean of those capacities is the average capacity for that mix.

For example, assume that the percentage of trucks is 10 percent. Then for every nine autos there will be one truck. If we calculate the capacity of the signal for each of the different positions of the truck in the queue and average these capacities, we will have the average capacity for a traffic stream with 10 percent trucks.

In Table 1 we have calculated the time in seconds for the queue to clear the intersection within 30 s under varying assumptions of the number of trucks in the queue. Note that, with a queue of six vehicles, all will clear even if all six are trucks. For seven vehicles, seven will clear if three or fewer are trucks. For eight vehicles, eight will clear when one vehicle is a truck or all vehicles are autos. For nine or more vehicles up to twelve in the queue, only all-auto queues will clear.

Table 1 has been converted to a capacity chart as shown in Table 2. The probability for any of the conditions in Table 2 can be calculated if the proportion of trucks in the traffic stream is known. For instance, assuming that 10 percent of all vehicles are trucks, the probability of four trucks in the queue of seven vehicles is $P_4 = (7!/4!3!)(0.1^4 \times 0.9^3) = 0.002\,551\,5$.

The probability for each of the outcomes in Table 2 has been calculated assuming that 10 percent of the traffic stream is trucks. These probabilities, when multiplied by the capacity (vehicles clearing the green phase) associated with each probability, yield the weighted average capacity. For Table 2 this is 9.14. Thus the presence of 10 percent trucks in the stream reduces capacity from 12 to 9.14, a reduction of some 24 percent.

### Comparison of Methods of Estimating Capacity

The following table reveals that the two shortcut methods (methods 1 and 2 above) seriously understate the impact of trucks on capacity, especially for the higher travel percentages.

| | Method of Adjustment | | |
|---|---|---|---|
| Proportion of Trucks | Vehicle Equivalents | Weighted Capacities ($P_TC_T + P_AC_A$) | Presence of Truck in Queue |
| 0 | 12.00 | 12.00 | 12.00 |
| 0.01 | 11.88 | 11.94 | 11.60 |
| 0.05 | 11.43 | 11.70 | 10.28 |

| Proportion of Trucks | Method of Adjustment: Vehicle Equivalents | Weighted Capacities ($P_T C_T + P_A C_A$) | Presence of Truck in Queue |
|---|---|---|---|
| 0.10 | 10.91 | 11.40 | 9.14 |
| 0.25 | 9.60 | 10.50 | 7.54 |
| 0.50 | 8.00 | 9.00 | 6.53 |
| 1.00 | 6.00 | 6.00 | 6.00 |

While it is true that, overall, traffic rarely contains 10 percent or more of those vehicles having sizes and acceleration characteristics that match our assumptions, truck routes commonly carry high percentages of heavy trucks. Even when only 5 percent of the traffic is trucks, a reduction of 14 percent in capacity would be expected. Yet, the two alternative techniques show only a 3-5 percent reduction in capacity. Therefore, for any careful analysis of truck impact on capacity, the queue analysis for capacity impact analysis would be recommended.

It would be possible, however, to use an equivalency table to approximate this capacity impact. The number of vehicles in a mixed queue that can clear a given amount of green time can be approximated by knowing the proportion of trucks in the stream, the capacity of autos, and the capacity of trucks. In our example above we can calculate such an equivalency table as mixed capacity = capacity of autos/proportion of autos + proportion of trucks.

Auto Capacity of 12 and Truck Capacity of 6

| Proportion of Trucks | Value of K |
|---|---|
| 0.01 | 4.448 |
| 0.05 | 4.346 |
| 0.10 | 4.129 |
| 0.25 | 3.366 |
| 0.50 | 2.675 |
| 1.00 | 2.000 |

From the table above for 10 percent trucks, we have $9.14 = 12/(0.9 + 0.1K)$; $K = 4.129$. We have calculated the K values for the truck proportions used in the previous table.

It would be possible to generate a series of tables that give the equivalency as a function of signal green time, auto and truck acceleration and spacing assumptions, and terminal speed. The user would then be able to use the same capacity restraint mechanism he or she now uses but with trucks properly weighted.

## CONCLUSIONS

These calculations have led to the following conclusions:

1. Separate but concurrent assignment of the truck origin-destination matrix over the highway network and retention of the link truck volumes is desirable for many subregional and neighborhood planning problems;
2. Provision for a truck network designation is desirable so that truck routes or truck prohibitions or both can be considered in the assignment process;
3. As a minimum, trucks should be divided into light or auto-like trucks and heavy trucks;
4. Even a small proportion of trucks in the traffic stream can result in substantial reduction in street capacity (this capacity reduction, however, is not a simple linear reduction proportional to the percentage of trucks in the vehicle stream); and
5. The reduction in capacity from trucks in the traffic stream can be represented by a straightforward algorithm that weights trucks differentially according to their proportion in the traffic stream.

Table 1. Time for nth vehicle to clear intersection.

| Number of Vehicles in Queue | Number of Trucks in Queue: 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| 6 | 17.8 | 24.6 | 25.8 | 27.0 | 28.1 | 29.1 | 30.0 |
| 7 | 20.0 | 27.0 | 28.2 | 29.3 | 30.4 | | |
| 8 | 22.0 | 29.4 | 30.5 | | | | |
| 9 | 24.0 | 31.75 | | | | | |
| 10 | 26.0 | | | | | | |
| 11 | 28.0 | | | | | | |
| 12 | 30.0 | | | | | | |
| 13 | 31.9 | | | | | | |

Table 2. Intersection capacity for 30 s with different queue lengths and proportions of trucks.

| Number in Queue | Number of Trucks in Queue | Number of Vehicles Able to Clear in 30 s | Probability Assuming 10 Percent Trucks in Traffic |
|---|---|---|---|
| 7 | 4 or more | 6 | 0.003 |
| 8 | 2, 3, or 4 | 7 | 0.184 |
| 8 | 1 | 8 | 0.383 |
| 9 | 9th vehicle 1st truck | 8 | 0.043 |
| 10 | 10th vehicle 1st truck | 9 | 0.039 |
| 11 | 11th vehicle 1st truck | 10 | 0.035 |
| 12 | 12th vehicle 1st truck | 11 | 0.031 |
| 12 | 0 | 12 | 0.282 |

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

*Abridgment*

# Modified Simulation Technique for Mode-Choice Analysis

Charles D. Dougherty, Delaware Valley Regional Planning Commission

The Delaware Valley Regional Planning Commission (DVRPC) is the agency responsible for the development and maintenance of regional travel demand forecasting models for the bi-state Philadelphia metropolitan area. The nine-county region, with a 1970 population of 5.2 million people living in an area of 9926 $km^2$ (3833 $miles^2$), is the fourth largest metropolitan area in the nation.

The regional transportation system is a highly developed and completely interrelated network of transit and highway facilities whose present configuration is the result of nearly three centuries of growth. It is estimated that in 1970 this system was required to serve the travel needs of nearly 12 million person trips on an average day.

Travel demand estimation for large, complex metropolitan areas is a very resource-consuming effort. The computer time cost and work hours of effort to apply the standard DVRPC simulation process make such application in some studies unreasonable. Yet, the level of detail in the standard process may be quite appropriate for that study. It was the intention of DVRPC to develop a technique that would utilize as much of the standard simulation modeling process as possible with the greatest reduction of cost. One such technique, referred to as the modified simulation technique (MST), is discussed here. To properly set the context of MST, the standard DVRPC simulation process will first be described.

## STANDARD DVRPC SIMULATION PROCESS

The DVRPC travel demand modeling concept (1) follows the traditional four-step process: trip generation, trip distribution, mode choice, and trip assignment (Figure 1). The trip generation (step 1) uses a disaggregate trip rate model. This step requires extensive knowledge of the magnitude and location of regional activities such as land use, employment, and the demographic characteristics of the resident population. Trip distribution (step 2) uses a typical gravity formulation stratified by trip purpose. The separation variable in the DVRPC model is a composite highway and transit travel time.

The third step of the travel-forecasting process, mode choice, estimates the proportion of the trips between two zones that will use the transit system and the proportion that will use the highway system. Figure 2 is a more detailed diagram of step 3. The diagram shows that there are three tasks (A, B, and C) that must be performed for both transit and highway prior to the estimation of mode trips. Task A, code/edit networks, involves describing in numerical terms every link in the two networks [over 40 000 links on more than 13 300 km (8300 miles) of facilities]. In addition, the levels, types, and interrelationships of the service provided by the transit network must also be numerically described.

Task B within the mode-choice estimation (step 3) is the determination of the "best" path (a successive combination of links) between every pair of zones in the region (1342 zones, roughly 1.8 million pairings for both the highway network and the transit network). In the DVRPC process, the best path is defined as the one that costs the least to take. Cost is actually the sum of the dollar value of perceived time and out-of-pocket cash. This type of cost is felt to describe the network's resistance to free flowing, instantaneous travel and is referred to as the impedance to travel. In task C, the component parts of this impedance are skimmed off the network and cataloged.

Task D, the estimation of transit mode and highway mode trips from the person trips of step 2, requires knowledge of the available transportation system (supply), knowledge of the flows to be handled by that system (demand), and knowledge of the environment in which the demand will seek satisfaction from the supplied system. This environment includes the social, economic, and policy aspects of urban activity and their spatial orientations. The mode-choice estimating model relates this knowledge to predicting the percentage of trips likely to use the transit mode and the percentage likely to use the highway mode. The multiplication of these percentages by the person trips from step 2 yields the person trips on each mode. The DVRPC mode-choice estimating model is a post-distribution, stratified diversion curve formulation that primarily relates differences in highway and transit travel time and cost to mode percentages. The stratifications are by trip purpose, principal transit submode of best transit path, and auto availability.

The final step of the standard (DVRPC) simulation process (step 4) is the assignment of the various types of trips to the transportation system networks.

## MODIFIED SIMULATION TECHNIQUE

The modified simulation technique (MST) is applicable to the study of changes in mode choice resulting from changes in service level, skip-stop or express service, station spacing, and shifts in route alignment. The underlying assumption is not that the primary transit submode route does change, but that the specific links of the path might change. The primary impact of these service changes will be in the level of transit mode choice for a given trip interchange.

The MST procedure is as follows: first, the standard regional simulation process is run with the transit facility fully coded into the network as a base case; second, the impedance catalog is modified to reflect the service changes embodied in each alternative to be studied; and, third, the mode-choice model (step 3, task D) is rerun for each alternative.

The heart of the MST lies in how the impedance catalog is modified. Figure 3 shows the flow of effort in the MST.

Subtask 1 (determine station geographic market area) uses the data file on best transit paths to identify station market areas. Using the UTPS program USTOS and a special DVRPC program STATMKTS, a zone-to-station correspondence table is constructed. This table identifies the transit line station used by each zone in the base case.

Subtask 2 (redefine station market area) requires the analyst to examine the description of the alternative

Figure 1. DVRPC standard simulation process.

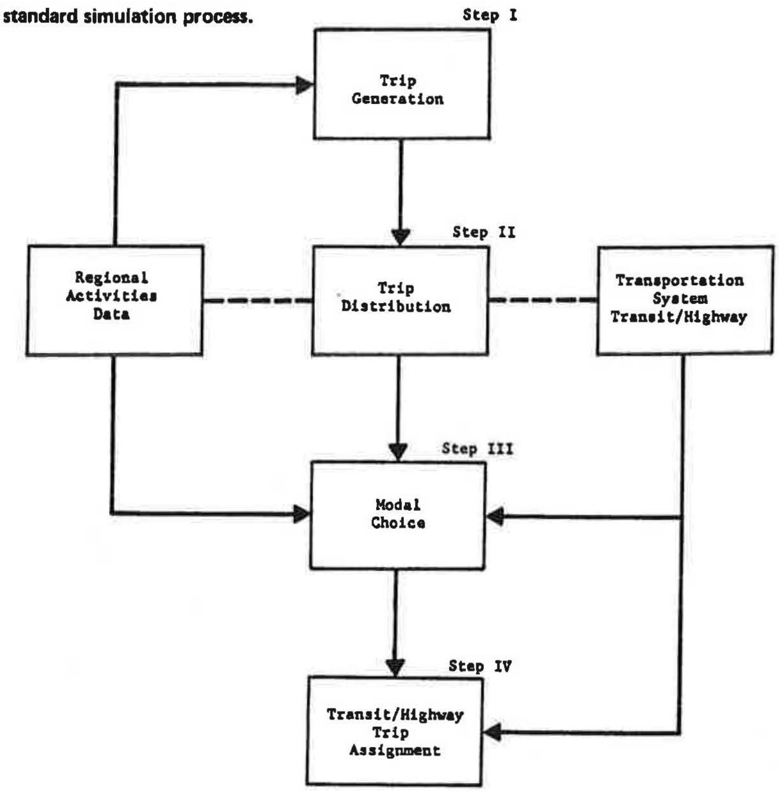



Figure 2. Standard simulation mode-choice process.

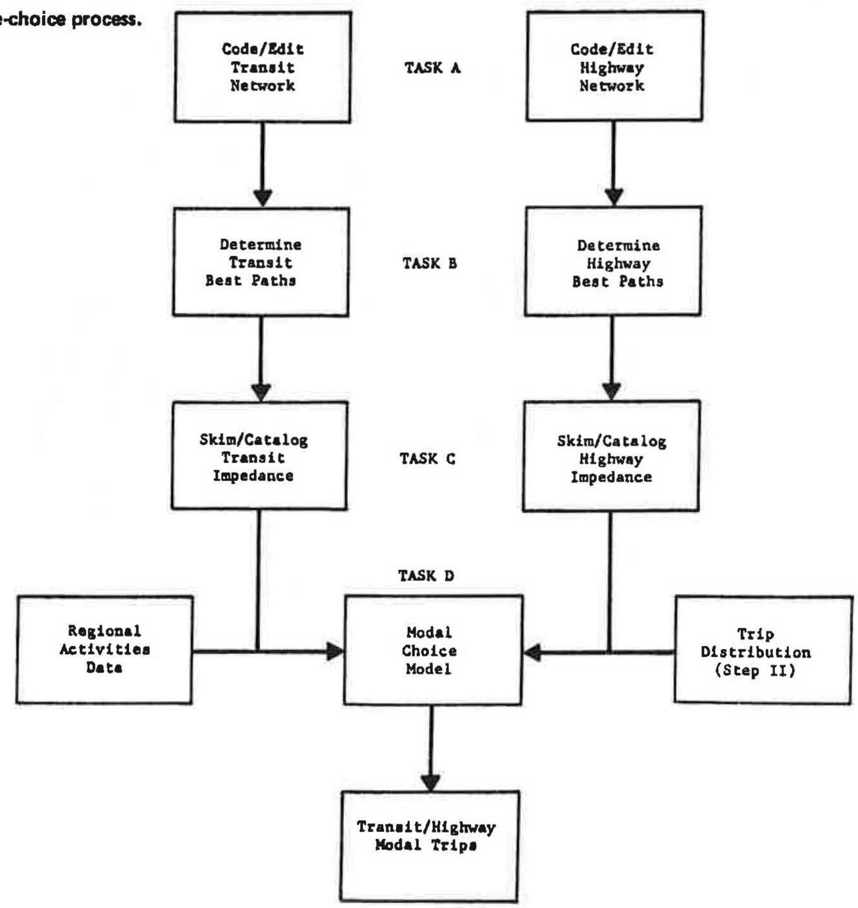
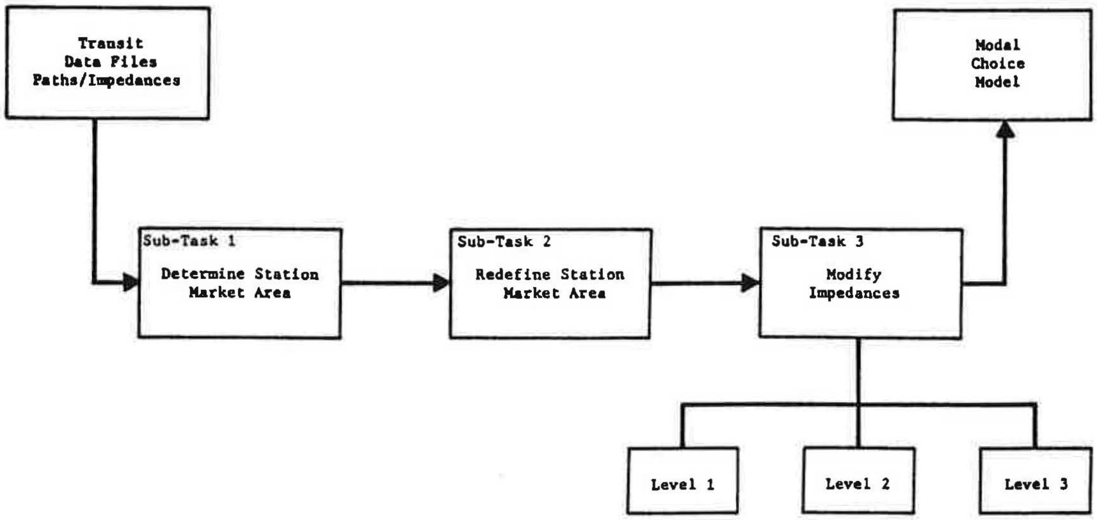


Figure 3. Modified simulation technique.




and the zone-to-station correspondence table and decide for each zone the station of choice under the conditions of the alternative. This task relies on the experience of the analyst and his or her familiarity with the service areas and travel behavior.

Subtask 3 (modify impedances) is divided into three potential levels of impedance modification. Level 1 accounts for the change in impedance resulting from any redefinition of the station market areas. Typically this level of modification occurs when a station is removed from the base case line. Trips that formerly used the removed station must use stations closer to or farther from their destinations, resulting in longer or shorter trips. The level 1 impedance change will be plus or minus the difference in running time between the old and new station of choice. Level 2 accounts for the changes in trip impedance resulting from changes in the line-haul operation that change interstation running times or station waiting times. Level 3 accounts for changes in impedance resulting from longer or shorter approach trips to the transit station. This type of change occurs when stations are removed from consideration or when the horizontal alignment of the line changes. Some changes may also require a change in approach mode and additional fare payments.

DVRPC has written a computer program, MODTIMPD, which reads in the zone-to-station correspondence table, the interstation impedance change matrix, the station reassignment table, and the trip-end access and egress impedance change table along with the skimmed impedance catalog. The output of the program is a modified impedance catalog. This modified impedance catalog is supplied as input to the mode-choice model for each alternative. The zone-to-station correspondence table can be used to reformat the new transit trip tables to get an interstation volume matrix.

## CONCLUSION

The modified simulation technique does provide a useful and cost-effective means around the expense of the standard simulation process. Of course, its applicability is limited to those studies where the underlying assumption of "no significant change in primary transit submode route" is considered a reasonable approximation. The procedure has been used by DVRPC for the screening of preliminary alternatives in the city of Philadelphia study of replacement alternatives for the Frankford Elevated rapid transit line (2).

## ACKNOWLEDGMENTS

I wish to acknowledge the assistance of the DVRPC, the city of Philadelphia and its consultants (Dalton, Dalton, Little, Newport), and the Urban Mass Transportation Administration.

## REFERENCES

1. Delaware Valley Regional Planning Commission. The Simulation of 1977 Travel on the Current (1977) Transportation Systems. Philadelphia, June 1977.
2. Dalton, Dalton, Little, Newport. Frankford Elevated Replacement Alternatives Analysis: Alternatives Analysis Report and Draft Environmental Impact Statement. Philadelphia, Mar. 1978.

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