An Application of Marginal Utility to Travel Mode Choice

GORDON A. SHUNK and RICHARD J. BOUCHARD, Alan M. Voorhees and Associates, Inc.

The increased interest in planning for transit as a portion of the solution to urban transportation problems has generated concern for the effectiveness of traditional modal-choice prediction procedures. The concern is centered on the ability of models calibrated on data reflecting base-year transportation service to predict the results of marked changes in service. A model that is basically more behavioristic in nature rather than simulative conceivably could be an approach to a solution of the problem. A modal-choice relationship was developed that utilized as its independent decision variable a composite of several, more traditional factors. It has been theorized that components of this variable represent the disutilities of travel by competing modes as perceived by the traveler. The differences in disutility represent the marginal disutility of a given mode. Marginal disutility was the decision variable for traveler choice between auto and transit. The final decision variable combined out-of-pocket cost of transit and highway travel, family income, and parking cost, as well as travel time for the trip. Modal choice was examined for 1958 data from St. Paul and Minneapolis, using marginal disutility as the independent variable. The results reproduced base-year transit travel patterns very well without using traditional curve-fitting calibration procedures. The distribution of the results appeared to approximate quite closely the normal. This seems to indicate that the variable used may well approach the actual, though unperceived, variable on which modal choices are based. The technique has the benefits of conceivably alleviating the need for calibration and offering a more basic, behavioristic formulation that may transcend constraints of transit service levels.

•THE INCREASING AWARENESS of the role of mass transit in the solution of the problems of urban transportation has stimulated new concern about conventional methods for forecasting transit travel patterns. A principal element of this concern has been associated with the need to reflect properly the effect on patronage of markedly improved levels of transit service. Such a need is particularly important when rapid transit service is to be introduced in an area previously served only by conventional bus operations. The planners for rapid transit in most major U. S. metropolitan areas will have to face such a situation. Furthermore, proper reflection of the effects of service is paramount to analyzing the impact on travel of new concepts in urban mass transportation. The essence of this need is to estimate properly the reaction to service levels not previously available in a particular urban area, if indeed they had been available anywhere. The implication in this is the need for a universal or basic behavioral representation, the theory of which transcends all physical systems and service levels.

In 1968, Alan M. Voorhees and Associates, Inc., undertook for the Twin Cities Area Metropolitan Transit Commission a study of the efficacy of several alternative rapid transit systems for serving the twin cities of Minneapolis and St. Paul, Minnesota (1).

Paper sponsored by Committee on Origin and Destination and presented at the 49th Annual Meeting.

A portion of that study required estimation of the ridership-generating potential of the candidate systems. Among the systems examined in detail were a conventional bus system operated to take maximum advantage of express highways; a modern rapid rail transit; and a small-vehicle, high-speed, "new concept" system. To estimate patronage for systems having such a broad range of service levels, it was necessary to develop a procedure that would properly reflect subtle service differentials among systems and between transit and highway alternatives presented the traveler. An additional factor required that the procedure have especially broad applicability.

The only detailed information available on travel in the Twin Cities was from a 1958 origin-destination survey conducted by the Twin Cities Area Transportation Study (TCATS). The bus and highway systems and travel of 10 years previous were therefore the only means available to calibrate the procedure.

The cornerstone of any such patronage-estimating procedure is a modal-split model. Such a model determines the split or allocation of all person trips among the available travel modes. Modal-split relationships usually use characteristics of the trip, the trip-maker, and the elements of the transportation system as independent variables. They also may employ any of several measures of transit service. Combining measures of these effects yields a reliable method for predicting transit patronage. Using data from an origin-destination survey and an inventory of the transportation system, one may observe the proportions of total trips made on transit by people of various socioeconomic groups and with varying conditions of transit service. These data are used to develop geometric or mathematical relations between important factors and travel mode choice. The relationships are generalized and used to estimate transit ridership in the base year. The estimate is compared with observed travel volumes, and the predictive procedure is adjusted to compensate for discrepancies. The process is repeated until the prediction is within tolerable limits of accuracy. Information on future transit service and socioeconomic and density factors is used as input to the model to predict patronage on proposed transit systems. The proposed systems are adjusted as necessary to achieve the best ridership possible within defined economic tolerances.

APPROACH

Modal-split relationships are of two generic types, predistribution or postdistribution. These types refer to whether the modal allocation is made prior to distributing trips between origins and destinations. Predistribution allocations generally tend to make primary use of the characteristics of the trip-makers and their residence zones. They cannot use the most effective measures of transit system service because these measures are related to zonal interchanges; i.e., they are effective on the trip itself. Predistribution allocations, therefore, must rely on weaker service measures. Such procedures are less difficult and less expensive than postdistribution types, and they are strongly wedded to existing and historic transit service levels.

Because the Twin Cities program was oriented to show the effects of greatly improved transit service, it was decided to use a postdistribution procedure. Postdistribution techniques allocate trips among the various modes on an interchange basis; i.e., the distribution of total trips from all origins among all destinations is assumed completed. Then, based on transit service levels for each zonal interchange (origin-destination pair) as well as other factors, the allocation of total travel is made among the available modes. Postdistribution procedures permit the best possible reflection of the effect of transit service differentials that exist between different trip interchanges. This is very important because not all trips from a zone will use transit at the same rate because all trips are not destined to places where transit provides good service.

The term "service level" has broad implications. It includes at least implicitly each of the following factors: first walk time to transit, first wait time for transit, transit speed, stop frequency of vehicles, fares, number of transfers required, wait time at transfer, and walk time from transit.

Service level usually is not examined on its own merits. Modal allocation of trips implies competition among modes and, therefore, examination of relative service levels. Previous modal-split models have used access or travel time ratios of transit to highway to represent relative transit service. The Twin Cities work used time differences between highway and transit travel. Time differences now are thought by some to better reflect differential service than ratios, although the theoretical basis for ratios is strong (2).

It has been known for some time that access and other nonmoving time components of total door-to-door time are perceived by the traveler in a different manner from time spent in motion. This is probably attributable to psychological factors that come into play as a person sees he is making no progress toward his destination as he waits, but progress is obvious when the vehicle moves. Nonrunning time, including walking time, occasionally has been given more influence on the computation of total door-to-door time than the running time. This influence has been effected by weighting the nonrunning or "excess" time by an empirical factor, usually evaluated at about 2.5 (2, p. 15). This procedure was employed in the Twin Cities work.

It follows that reducing running time differential by providing high-speed transit vehicles would have less effect on patronage than would reducing excess time. Selection of new transit systems should be directed at reducing those elements of transit travel time that are most effective in enhancing the relative attractiveness of transit. Alternatively, new transit systems may offer markedly greater speeds and commensurately reduced running times to offset uncorrectable excess times. These two concepts in improved service are the essential elements of new transit systems currently being developed.

Modal choice also is influenced by characteristics of the trip-maker, his trip, and the locations at which the trip starts and ends. Trip-maker characteristics that most influence modal choice are sex, age, and income. These factors occasionally are represented by a surrogate population density of the origin zone. Income is highly correlated with car ownership, which has a direct effect on modal choice. Parking cost has a very significant effect on travel mode selection; it may be represented by the surrogate employment density. The Twin Cities modal-split relationship utilized zonal, median family income to represent the effect of trip-maker and residence zone characteristics. Parking cost was used to represent destination zone characteristic effects.

Trip purpose has a great effect on modal choice. Work trips are usually the most oriented to transit. This may be attributed to such things as regularity, occurrence during hours of dense travel, and most employment historically being located in the central business district (CBD), the area best served by transit. School travel is also transit-oriented, whether the transit is public or school bus. People are less inclined to rely on transit for trips with other purposes, probably because the destinations for these trips are usually in less densely developed areas that, consequently, are not well served by transit. These trips are also less regular, and the timing of them is less important, so they can be made when the family automobile is available. Transit trips for these purposes are dominated heavily by transit captives, persons who have no automobile available permanently or temporarily. The modal-split relationships discussed here were for work trips only.

Most modal-split or allocation procedures employ and are referred to as "models." This terminology is consistent because some mathematical formulation is used. The Twin Cities relationships were not developed as mathematical relationships but as graphic plots of nonlinear relationships. To develop these, it was necessary to employ manual plotting of stratified observations. This approach was chosen because current, nonlinear, regression techniques are constrained by limits on order and requirements of consistency. Data describing the trip-maker, trip, transit service, and highway service were cross-stratified in several ways by a specially tailored computer routine. The transit percent of total ridership observed in each cell of the cross-stratification was plotted against the several strata levels. The curves or surfaces describing variation in transit ridership thus were defined. Manual plotting permitted use of multiple-dimension, curvilinear relationships. It did not require assumptions or constraint of data to force linearity. It provided the additional advantage of allowing extrapolation of the curve in regions of interest for which data were not available in a manner dictated by experience. Because the objective of this study was to analyze the effects of service that represents a major improvement over that currently available, the extrapolation permitted developing more complete and consistent curves.

THE CONCEPT

Development of the Twin Cities modal-split relationship involved examination of a universal utility measure. Many people have thought for some time that modal choice is much akin to traffic diversion on highways, albeit between different means of travel rather than different routes (3). People either choose or are diverted to alternative travel modes by their perception of the relative attractiveness of each. The important element of this theory is a hypothetical factor to which all others can and must be reduced. This is the element on which people at least implicitly may base modal choices. This factor, termed "modal-choice utility" in this study, may be represented properly by a combination of several, more basic factors of influence such as time and dollar cost. The Twin Cities work undertook examination of the essence and applicability of this hypothesis. Such an approach could lead to much more easily applied modal-choice relationships because of the implied universality of the utility factor. It also would permit inclusion of such economic factors as road pricing and other economic policies. Additionally, it would validate extrapolation in areas where calibrating data were sparse by virtue of the theory of the utility function.

The theory of the utility function is of this nature: Given that a variable can be defined that explicitly or implicitly represents the datum on which people base decisions, the distribution of the results of such decisions plotted against values of the variable will approach normality. Thus, if a decision variable represents all perceived travel disutility, it should be a candidate for such a function. If the results of observations of a dependent variable appear to be distributed normally when plotted or examined, it may be assumed that the decision variable is adequate. The following section will elaborate on the testing to prove such a hypothesis.

PROCEDURE

The procedure for developing a modal-split relationship, regardless of its formulation, follows a basic pattern. This pattern, in general, consists of matching existing or observed ridership to corresponding socioeconomic characteristics of the trip-maker and service characteristics of the transit system. A relationship is developed, tested, revised to match observed conditions, and retested. The process is iterated to satisfactory closure tolerances. The final relationship is applied to future-year person travel estimates using future-year socioeconomic and transit system characteristics as independent variables. The independent characteristics used in development of the relationship must have been predicted for the future year.

Development of the Twin Cities modal-split relationship began with preparation of survey-year data. Major effort was concentrated on preparing a representation of the 1958 transit system for computer processing. This effort included coding transit routes and their characteristics for input to the Department of Housing and Urban Development (HUD) Transit Planning Program package. This interrelated set of programs is capable of representing most aspects of a transit system that are important to modal choice and operational analysis. It also permits development of data that are technically and physically consistent with data currently developed for highway systems using the BPR BELMN package. The two sets of data then can be compared. The 1958 Twin Cities highway network was prepared using BELMN and inputs provided by the Minnesota Highway Department. These data were the same as those used by the department in model calibration for the TCATS. The HUD transit programs permit coding of a transit system and gaining access to it in such a manner as to have available for individual analysis such components of transit service as walk, wait, transfer, and run time; number of transfers; and fare. Coding of the 1958 Twin Cities transit system was done from route schedules, thereby using schedule stops, times, headways, and fares.

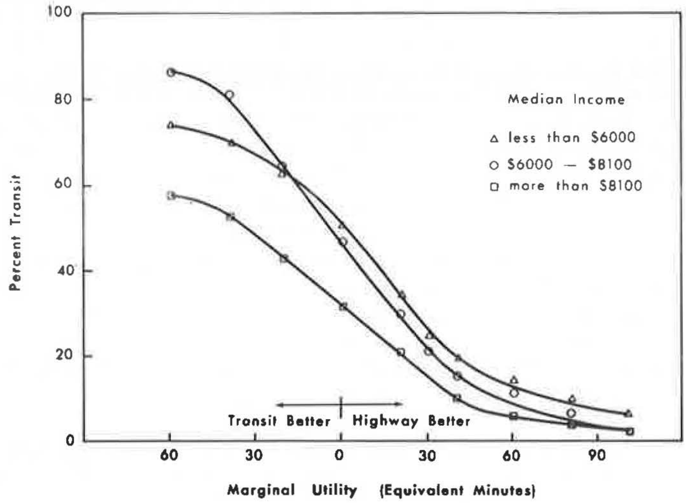

Figure 1. Work trips in Twin Cities, marginal utility, modal-choice model.

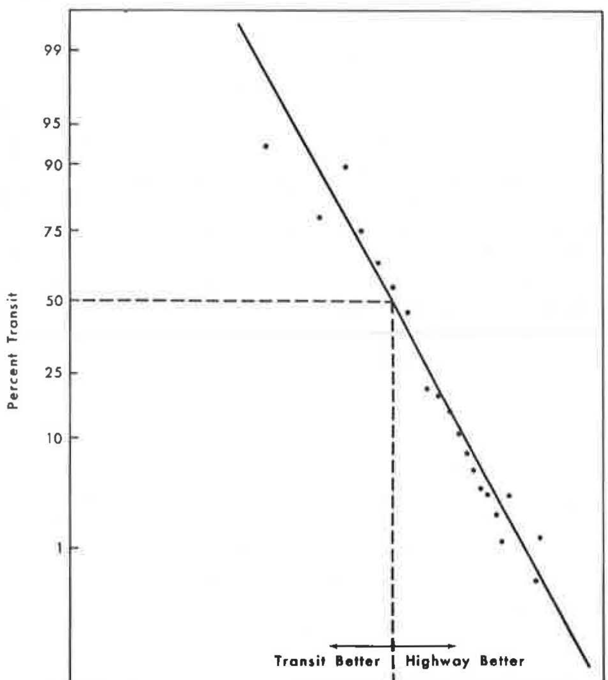

Figure 2. Log-probability relationship; percent transit versus marginal utility work trips.

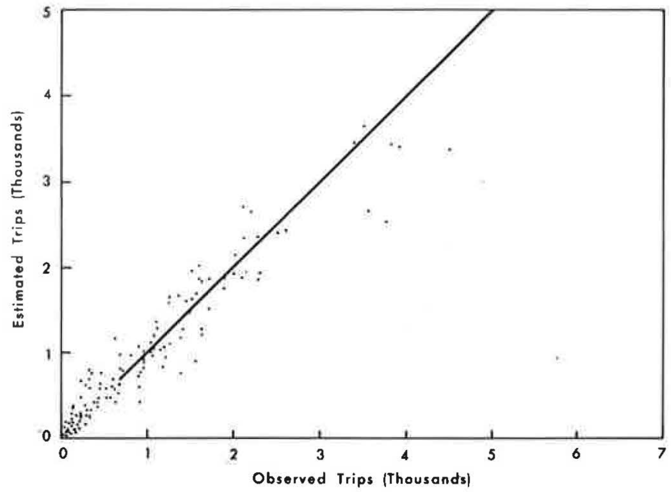

Figure 3. Transit work trips, Twin Cities, 1958 (observed and estimated productions by district).

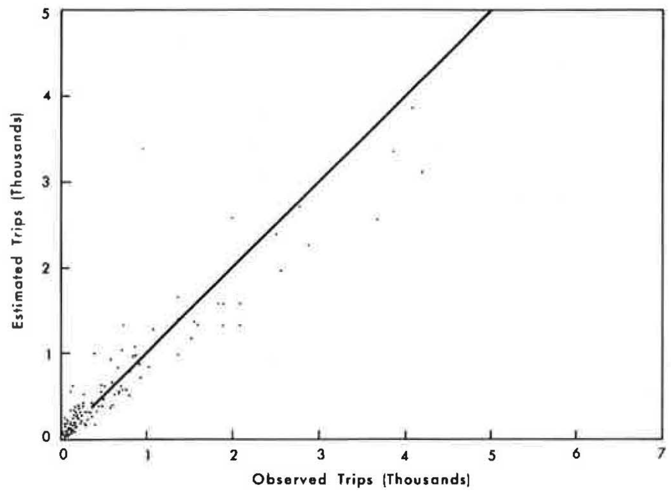

Figure 4. Transit work trips, Twin Cities, 1958 (observed and estimated attractions by district).

work trips, were predicted extremely well. Even production of transit trips was predicted very well. The attraction results imply that parking cost is indeed a good indicator of the attractiveness of a zone for transit travel. The lower R^2 value for productions implies that the income variable used is less reflective perhaps of transit trip production ability than of something else, such as auto ownership. The apparently low evaluation of transit interchange prediction performance is actually quite good. Statistics for the same comparison run considerably lower in other studies. The impact of a good result here is that actual travel movements or patterns were predicted accurately 57 percent of the time. This reflects the validity of the interchange service