inary Res. Rept. 41, 1973.
9. M. A. Johnson. Psychological Variables and Choices Between Auto and Transit Travel: A Critical Research Review. Institute of Transportation and Traffic Engineering, Univ. of California, Berkeley, Working Paper 7509, 1975.
10. I. P. Levin. Averaging Processes in Ratings and Choices Based on Numerical Information. Memory and Cognition, Vol. 2, 1974, pp. 786-790.
11. I. P. Levin. Information Integration in Transportation Decisions. In Human Judgment and Decision Processes: Applications in Problem Settings (M. F. Kaplan and S. Schwartz, eds.), Academic, New York, 1975.
12. I. P. Levin. The Development of Attitudinal Modeling Approaches in Transportation Research. In Proc., 3rd International Conference on Behavioral Travel Modeling (D. A. Hensher and P. R. Stopher, eds.), Pergamon, New York, in press.
13. I. P. Levin and M. J. Gray. Analysis of Human Judgment in Transportation. In Human Judgment and Spatial Behavior (R. Reider and J. J. Louviere, eds.), Great Plains/Rocky Mountains Geographical Journal, Special Issue, in press.
14. J. B. Margolin, M. R. Misch, and R. Dobson. Incentives and Disincentives to Ride-Sharing Behavior: A Progress Report. TRB, Transportation Research Record 592, 1976, pp. 41-44.
15. R. B. Meyer and I. P. Levin. Functional Determinants of Mode-Choice Behavior. Institute of Urban and Regional Research, Univ. of Iowa, Iowa City, Technical Rept. 88, 1977.
16. K. L. Norman and J. J. Louviere. Integration of Attributes in Bus Transportation: Two Modeling Approaches. Journal of Applied Psychology, Vol. 59, 1974, pp. 753-758.
17. R. B. Ross. Measuring the Effects of Soft Variables on Travel Behavior. Traffic Quarterly, Vol. 29, 1975, pp. 333-346.
18. P. R. Stopher. On the Application of Psychological Measurement Techniques to Travel Demand Estimation. Paper presented at the Meeting of the American Psychological Association, New Orleans, 1974.

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

Comparative Analysis of Determinants of Modal Choices by Central Business District Workers

Ricardo Dobson* and Mary Lynn Tischer, Federal Highway Administration, U.S. Department of Transportation

The role of individuals' perceptions and preferences in traveler decision making is a growing and active area of theoretical and empirical research. This study was designed (a) to quantify the relation between perceived system attributes and modal choice, (b) to compare the magnitude of this relation to that of the alternative relations of sociodemographic and network time and cost data and modal choice, and (c) to determine whether the linkage between perceived system attributes and modal choice is dependent on the relations of sociodemographic and network data to modal choice. The sample was composed of Los Angeles central business district workers who live within approximately 3.2 km (2 miles) of a freeway that feeds radially downtown. Models were calibrated for three dependent-variable criteria; these were monthly differences in use for (a) automobile versus bus, (b) automobile versus car pool, and (c) bus versus car pool. The multiple coefficients of determination for modal choice as a function of perceived system attributes were statistically significant at the 0.001 level for all dependent-variable criteria. The coefficients ranged from 0.265 to 0.125, but the analogous coefficients for sociodemographic or planning data ranged from a low of 0.004 to a high of only 0.054. The effects of perceived system attributes on the dependent variables were not diminished by the other types of independent variables. Tests of significance for the individual components of combined models with these types of data showed the perceived-system-attribute data to be significant at beyond the 0.001 level in all cases. However, sociodemographic and network data appear to be influenced by the addition of perceived-system-attribute data to the degree of becoming nonsignificant in some cases. The overall conclusion is that perceived system attributes can be a statistically significant correlate of modal choice over and above any influence by network or sociodemographic variables or both.

The potential relevance of individuals' perceptions of and preferences for transportation modes and their service attributes to traveler decision making makes these topics active areas of theoretical and empirical research. Golob (1) and subsequently Golob and Dobson (2) have reviewed consumer-behavior models derived from market research and psychology and discussed their applicability to urban travel issues. An organizational structure has emerged from a conference on behavioral travel-demand that specifies the relation of perceived system attributes to traveler behavior (3). The perceptual models developed here are consistent with this organizational structure.

From an empirical perspective, the role of perceptions and preferences in traveler decision making is mixed. Hartgen (4) found that attitudes accounted for only 10 to 20 percent of the explained variation in traveler modal choices while situational variables, e.g., income and automobile ownership, accounted for 80 to 90 percent of the explained variance. A similar finding was reported by Dobson and Kehoe (5), who did not find a statistically significant correspondence between view of system attributes and actual modal choices, although they did find substantial correlations between perceptual measurements of transportation system attributes and anticipated satisfactions with innovative urban transportation modes. Thus, neither of these investigations offers strong support for perceptions or preferences as determinants of modal choice.

On the other hand, at least three other analyses support the concept that perceptions and preferences are determinants of modal choice. Dobson and Tischer (6) found 63 percent correct assignments to mode for a model based solely on perceived system attributes for which the chance probability of correct assignments was only 33 percent. Nicolaidis (7) reported higher correlations between comfort and modal choice than between either perceived time or cost differences and modal choice. Spear (8) found that the correlation ratio for a logit function fit was appreciably enhanced by the addition of a generalized convenience variable to a modal-choice model already specified in terms of perceived time and cost differences.

This study was designed to advance the level of understanding with respect to the potential relevance of individuals' perceptions and preferences to traveler decision making. Three questions are specifically addressed with respect to a sample of Los Angeles central business district (CBD) workers. These are as follows:

1. Is there a correspondence between perceived system attributes and modal choice?
2. What is the magnitude of the relationship between perceived system attributes and modal choice relative to those of network time and cost or of sociodemographic data and modal choice?
3. What is the nature of the interaction between perceived system attributes, network time and cost, and sociodemographic variables with respect to modal choice?

STUDY DESIGN

Sample

The sample was comprised of 889 individuals who work in the Los Angeles CBD and live within 3.2 km (2 miles) of a freeway that feeds radially to the CBD. It was selected from census tracts surrounding freeways and having a high incidence of downtown workers. The households were chosen from a telephone directory according to a random sampling procedure. However, only commuters who worked in the CBD were interviewed. If a household contained more than one CBD worker, the person taking the lesser used mode was interviewed.

Data Sets

The following types of data were collected from the home-interview survey: (a) frequency of modal use, (b) perceived system attributes, and (c) sociodemographic data regarding the individual and the household. An additional data set composed of network time, distance, and cost was derived at the aggregate level for the sample and is presented. These aggregate data were assigned to individual travelers depending on their origin and destination (O-D) zones.

Individuals were asked how frequently (per month) they used each of three modes—the single-occupant automobile, the bus, and the car pool—to travel to work. The frequency of use of each mode was subtracted from that of the other modes to obtain comparisons of modal use. These constructed variables—frequency of automobile use minus that of bus use, frequency of automobile use minus that of car-pool use, and frequency of bus use minus that of car-pool use—are used throughout the analysis as the dependent variables.

Beliefs about attributes of the three modes—the single-occupant automobile, the bus, and the car pool—were collected by using a semantic differential format in which each attribute was rated on a one to seven scale. The respondent was asked to describe the following 19 attributes for each of the modes.

1. Worry about being harmed by others versus do not worry about being harmed by others,
2. Easy to get where I am going after I leave the vehicle versus not easy to get where I am going after I leave the vehicle,
3. Is not crowded versus is crowded,
4. Usually do not have to wait a long time for vehicle versus usually have to wait a long time for vehicle,
5. Do not feel relaxed in this vehicle versus feel relaxed in this vehicle,
6. Am not exposed to weather versus am exposed to weather,
7. Can avoid waiting in lines in traffic versus cannot avoid waiting in lines in traffic,
8. Can come and go on my own schedule versus cannot come and go on my own schedule,
9. Very little extra time spent waiting for others or walking to or from vehicle versus much extra time spent waiting for others or walking to or from vehicle,
10. Would not cost much for parking versus would cost a lot for parking,
11. Comfortable versus not comfortable,
12. Not convenient versus convenient,
13. Not expensive versus expensive,
14. Not enough space for packages versus enough space for packages,
15. Easy to use versus not easy to use,
16. Cannot rely on it versus can rely on it,
17. Usually arrive at work on time versus usually do not arrive at work on time,
18. A slow way to travel during rush hour versus not a slow way to travel during rush hour, and
19. Can feel safe from vehicle accidents versus cannot feel safe from vehicle accidents.

(Hereafter, the terms belief and perceived system attributes will be used interchangeably.)

The following sociodemographic information was collected from each individual: marital status, number of people in the household, type of dwelling, income, age, race, sex, presence of children under 18, and number of automobiles in the household. Each of the variables was effect coded (9) and divided into category variables.

The automobile-network-time and distance data were derived from studies made available by the California Department of Transportation. The creation of the data was based on the following:

1. The 1967 Los Angeles Regional Transportation Study of the O-D in the five-county area,
2. A 1975 network update, and
3. Minimum time path assignments of the 1967 O-D patterns to the 1975 network.

The zones selected for analysis were those represented by one or more respondents. The automobile-network-time data were used to calculate an automobile impedance according to a formula that weighted out-of-vehicle time 2.5 times as heavily as in-vehicle time.

Automobile costs represent a combination of parking cost and a distance-based rate for gasoline, maintenance tires, and oil costs, as well as travel-related depreciation. Transit time data were developed in a fashion analogous to that for the automobile. Transit costs refer specifically to the bus fare.

Car-pool impedance was calculated by adjusting the automobile access time to account for the picking up of passengers, taking the line haul and egress to be the same as for the automobile. Car-pool cost was obtained by dividing automobile cost by mean car-pool occupancy.

Both the perceived system attributes and the sociodemographic data were subjected to data-reducing techniques that are described more fully below.

Hypothesis Testing of General Linear Forms

The primary focus of this paper is the analysis of the relative explanatory abilities of each of the data sets, separately and in conjunction with one another. By using a regression framework, it is possible to partition the regression sum of squares (R^2) into various components so that the effect of each independent variable (X_i) can be tested. One can thus obtain the independent explanatory power of a particular variable, or as in this case, set of variables, in terms of the percentage of variance accounted for in the dependent variable (Y) and its relative statistical significance (10, 11). The additional contribution of each independent variable in explaining the dependent variable can be determined by calculating R^2 with and without the particular variable.

This general analysis of the variance of the components requires three steps. First, a regression is performed for all independent variables, as in the following equation:

$$Y = a_0 + a_1 X_1 + a_2 X_2 \quad (1)$$

where a_0 is the intercept term and a_1 and a_2 are the slope coefficients. Then, a second regression is performed that omits the variable of interest, for example, X_2.

$$Y = b_0 + b_1 X_1 \quad (2)$$

where b_0 is the intercept term and b_1 is the slope coefficient. The R^2 of Equation 2 is then subtracted from the R^2 of Equation 1, leaving the amount of variance in Y that is explained by X_2, above and beyond that explained by X_1.

It is possible, however, that the explanatory variables are not completely independent from one another such that the effect of X_1 on the dependent variable depends on its combination with X_2. Significant interaction suggests that a given change in X_1 will produce different changes in Y for different values of X_2.

Testing for interaction follows the procedure described above. Two regression equations are necessary, but a new component—the interaction term—is added, as shown in Equation 3.

$$Y = c_0 + c_1 X_1 + c_2 X_2 + c_3 X_1 X_2 \quad (3)$$

where c_0 is the intercept term and c_1, c_2, and c_3 are the slope coefficients. The second regression omits the variable of interest, which in this case is the interaction term.

$$Y = d_0 + d_1 X_1 + d_2 X_2 \quad (4)$$

where d_0 is the intercept term and d_1 and d_2 are the slope coefficients. When the R^2 of Equation 4 is subtracted from the R^2 of Equation 3, the effect of the interaction between X_1 and X_2 is obtained.

The significance of each of the effects can be determined by using the following formula (9):

$$F = [(R^2_{y.12\ldots k_1} - R^2_{y.12\ldots k_2})/(k_1 - k_2)] \div [(1 - R^2_{y.12\ldots k_1})/(N - k_1 - 1)] \quad (5)$$

where

$R^2_{y.12\ldots k_1}$ = squared multiple correlation coefficient for the regression of Y on k_1 variables (the full model),

$R^2_{y.12\ldots k_2}$ = squared multiple correlation coefficient for the regression of Y on k_2 variables (the model with deleted terms),

k_1 = number of variables in the full model,

k_2 = number of variables in the model with deleted terms, and

N = sample size.

Regression analysis assumes a linear relation, but the functional form of the relation between planning data and modal choice has been most often specified in logit or probit form. One of the major problems of both logit and probit formulations, however, is the lack of standard measures for assessing the goodness of fit of the models or their statistical significance (12). There are no widely accepted statistics associated with the logit or probit forms available that allow a precise analysis of the effects of different data sets. Regression analysis and the related analysis of the variance of the components are the best such techniques, even though they require the assumption of linearity.

Because the concern of this analysis is not with model building, but rather with the comparison of independent effects of sets of explanatory variables, the form of the model is not central to the study design. Furthermore, it was assumed that any underrepresentation of the relation resulting from the imposition of the linear form would occur across data sets. And because the concern is with the comparability of data, the underrepresentation would not have any effect on the results.

These assumptions, however, were tested. The results of the linearity tests of the models and the analyses for bias showed only occasional deviations from linearity that were not biased.

RESULTS

This section of the paper describes the processing of the three types of data examined in the study and the model specifications of their relations to modal choices. The premodeling analysis included factor analysis and clustering of the sociodemographic data and factor analysis of the perceived-system-attribute data. The model specification activities involved calibrations of regression models, tests of hypotheses based on a general procedure for analysis of the variance of the components, and tests for data nonlinearity and bias in the model estimates.

Premodeling Analysis of Perceived System Attributes

For each transportation mode—single-occupant automobile, bus, and car pool—19 belief ratings were obtained. To reduce the number of variables quantifying the perceived system attributes and to take advantage of natural correlations among the variables, a principal-component factor analysis with a varimax rotation was computed for the beliefs about each mode. The factor scores were then estimated for the calibration of modal-choice models that include the perceived system attributes. These premodeling computations were implemented through use of the Statistical Package for the Social Sciences computer program (13).

Table 1 gives the labels and definitions of the factors that are based on the perceived-system-attribute structure with respect to single-occupant automobiles, buses, and car pools. The criterion for the selection of the factors to be retained for analysis was a principal-component eigenvalue greater than 1. The rotated factor-loading matrices were examined for an obvious gap between high and low factor loadings. The boundary criterion was set at 44 and gave readily interpretable dimensions. Attributes with factor loadings at or above the lower bound were used to define a factor.

Table 1 shows that the factor structures for the different modes are relatively similar. For example, all modes have factors that include traffic wait, cost, comfort, and convenience. Although the definitions of the factors differ slightly across modes, there is a general pattern. Furthermore, the number of factors is relatively constant across modes. There are five factors each for the bus and single-occupant-automobile modes, and there are four factors for the car-pool mode.

Table 1. Factor definitions of beliefs about modes.

Mode	Factor Label	Belief Variable	Factor Loading
Single-occupant automobile	F_1 (comfort and convenience)	Comfort	0.70
		Convenience	0.68
		Ease of use	0.63
		Arrive on time	0.59
		Ease to destination	0.56
		Relaxing	0.53
		Flexible schedule	0.54
	F_2 (cost)	Cost	0.76
		Parking cost	0.72
	F_3 (safety)	Vehicular safety	0.54
		Personal safety	0.74
	F_4 (traffic wait)	Travel time	0.75
		Waiting time in traffic	0.69
	F_5 (extra time, crowding, and weather)	Crowding	0.73
		Waiting for vehicle	0.74
		Weather	0.59
		Extra time	0.60
Bus	F_1 (comfort)	Comfort	0.71
		Space for packages	0.65
		Crowding	0.50
		Relaxing	0.66
	F_2 (convenience)	Convenience	0.74
		Ease of use	0.80
		Reliability	0.60
		Arrive on time	0.71
		Ease to destination	0.64
		Wait for vehicle	0.61
	F_3 (cost and safety)	Cost	0.45
		Vehicular safety	0.59
		Personal safety	0.51
		Parking cost	0.66
	F_4 (traffic wait)	Travel time	0.78
		Waiting in traffic	0.55
	F_5 (extra time and weather)	Weather	0.73
		Flexible schedule	0.54
		Extra time	0.52
Car pool	F_1 (comfort)	Comfort	0.56
		Space for packages	0.53
		Reliability	0.56
		Vehicular safety	0.46
		Personal safety	0.53
		Relaxing	0.66
	F_2 (convenience)	Convenience	0.48
		Ease of use	0.56
		Arrive on time	0.53
		Ease to destination	0.62
		Crowding	0.61
		Waiting time	0.75
		Weather	0.50
		Extra time	0.57
	F_3 (cost)	Cost	0.73
		Parking cost	0.74
	F_4 (traffic wait and flexible schedule)	Travel time	0.47
		Waiting in traffic	0.81
		Flexible schedule	0.45

Premodeling Analysis of Sociodemographic Data

The category-coded sociodemographic data were analyzed by a principal-component factor analysis with a varimax rotation. The common-factor criterion used for the perceived-system-attribute data suggested a six-factor solution. However, after a review of solutions with three through six factors, a four-factor solution was selected on grounds of interpretability. The sets of high factor loadings used to define factors were chosen on the same basis as those used for the perceived-system-attribute data. The factor scores were then computed and clustered to find homogenous travel markets via an average linkage clustering algorithm by using the Biomedical Computer Program P2M (14). Blashfield (15) notes that this clustering approach has been shown to be effective in at least four empirical studies of alternative clustering methods.

The cluster analysis resulted in a total of 11 groups that formed at an amalgamation distance of 1.644 or less. Further amalgamations of these 11 groups produced increases in the amalgamation distance that were substantially and consistently larger than the earlier increases. Five of the groups have cluster sizes of at least 71 workers, but the remaining 6 groups have cluster sizes of 30 or fewer. Therefore, the centroids of the sociodemographic-factor-score space for the 5 largest groups were used as the nuclei of a 5-group solution by assigning respondents from the other 6 groups to their nearest neighbor in the 5-group solution. With one exception, no large centroid coordinate was modified by more than 5 percent. Even for the exception, the new centroid coordinate did not result in a different interpretation of characteristics of the respondents. This centroid position, cluster three on the first factor score, changed from 0.83 to 0.61.

Table 2 summarizes the results of the cluster analysis for the five-group solution after the addition of the six smaller groups. Assigning all of the sample to the five clusters changes the bases for the interpretation of only one cluster solution. The first factor of cluster three was decreased significantly and altered the interpretation of that segment. The large factor-score centroid coordinates are used as indicators of the descriptive characteristics of a group.

Table 2 shows that the groups occupy disparate positions in the factor-score space. Cluster four is characterized by three-or-more-person households that reside

Table 2. Summary of sociodemographic cluster analysis.

			Sociodemographic-Factor-Score Centroids			
Cluster	Sample Size	Descriptive Characteristics	Household Size and Type of Residence	Income	Age	Automobile Ownership
1	125	Households having three or more automobiles	-0.40	-0.16	-0.17	-2.25[a]
2	108	Households having income >$30 000	-0.50	-2.01[a]	-0.37	0.52
3	138	Households having one or more workers that are at least 55 years old	0.61	0.27	-1.79[a]	0.43
4	254	Households having three or more persons, including a child under 18 years old, residing in a single-family house	-1.06[a]	0.44	0.24	0.47
5	264	Households having one or two persons residing in an apartment or households having one or more workers that are less than 55 years old	0.92	0.34	0.63	0.18

[a] Used as indicators of the descriptive characteristics of a group.

Table 3. Multiple coefficients of determination of models calibrated with a single type of data.

Type of Data	Frequency of Use					
	Automobile - Bus		Automobile - Car Pool		Bus - Car Pool	
	Coefficient	p	Coefficient	p	Coefficient	p
NT (time and cost)	0.033	<0.001	0.004	>0.10	0.016	0.001 to 0.01
SD (sociodemographic)	0.054	<0.001	0.030	<0.001	0.012	0.01 to 0.05
B (modal beliefs)	0.265	<0.001	0.124	<0.001	0.178	<0.001

Table 4. Multiple coefficients of determination of models calibrated with two or three types of data.

Types of Data	Frequency of Use					
	Automobile - Bus		Automobile - Car Pool		Bus - Car Pool	
	Coefficient	p	Coefficient	p	Coefficient	p
NT + SD	0.077	<0.001	0.032	<0.001	0.027	<0.001
NT + B	0.269	<0.001	0.126	<0.001	0.181	<0.001
SD + B	0.279	<0.001	0.140	<0.001	0.181	<0.001
NT + SD + B	0.282	<0.001	0.142	<0.001	0.184	<0.001

Table 5. Multiple coefficients of determination of models calibrated with two or three types of data and interaction terms.

Types of Data	Frequency of Use		
	Automobile - Bus	Automobile - Car Pool	Bus - Car Pool
NT + SD + (NT * SD)	0.094	0.033	0.042
NT + B + (NT * B)	0.274	0.127	0.189
SD + B + (SD * B)	0.284	0.150	0.196
NT + SD + B + (SD * B)	0.288	0.153	0.200
NT + SD + B + (NT * SD)	0.293	0.143	0.194
NT + SD + B + (NT * B)	0.286	0.143	0.192
NT + SD + B + (SD * B) + (NT * SD)	0.299	0.154	0.209
NT + SD + B + (SD * B) + (NT * B)	0.293	0.154	0.208
NT + SD + B + (NT * SD) + (NT * B)	0.296	0.143	0.201
NT + SD + B + (SD * B) + (NT * SD) + (NT * B)	0.303	0.154	0.216
NT + SD + B + (SD * B) + (NT * SD) + (NT * B) + (NT * SD * B)	0.332	0.166	0.249

Note: All multiple coefficients of determination are statistically significant beyond the 0.001 level.

in single-family homes, and cluster five is characterized by one and two-person households that live in apartments. Clusters three and five differ in that cluster three represents households with at least one worker who is at least 55 years old, but cluster five represents households with a worker who is less than 55 years of age. Clusters one and two are relatively extreme groups that are characterized by households with three or more automobiles or an income of at least $30 000 respectively.

A similar clustering attempt was made for the perceived-system-attribute data, but initial efforts did not result in interesting cluster patterns.

Model-Specification Statistics

The three different dependent-variable criteria were (a) the frequency of automobile use minus that of bus use, (b) the frequency of automobile use minus that of car-pool use, and (c) the frequency of bus use minus that of car-pool use. Time and cost data were treated as the difference between the modes for these variables. The five sociodemographic clusters given in Table 2 were effect coded. These were four categories of sociodemographic variables. These variables were coded with values of 1, 0, and -1 depending on the cluster to which a respondent was assigned. These same sociodemographic-category variables were used for all three dependent-variable criteria.

To prevent the development of models with an excessive number of terms, only the scores for two perceived-system-attribute factors were used for each mode. These were F1 and F2 for the bus mode, F1 and F2 for the car-pool mode, and F1 and F5 for the single-occupant-automobile mode. These factors were chosen for inclusion because they accounted for the largest percentage of variance in the common-factor solution for beliefs about each mode. Only those factor scores that pertained to the dependent variable were included in a model. Therefore, the perceived-system-attribute terms for an automobile-minus-bus model consisted of the F1 and F5 scores for the automobile and the F1 and F2 scores for the bus.

Models Calibrated With a Single Type of Data

Table 3 gives the multiple coefficients of determination and the associated statistical significance levels for models calibrated with a single type of data for each of the three dependent-variable criteria. The results for models calibrated exclusively with beliefs about modes have consistently and substantially larger multiple coefficients of determination than do the results for models calibrated with time and cost and sociodemographic data. The multiple coefficient of determination for the perceived system attributes is statistically significant beyond the 0.001 level for all three dependent-variable criteria. Furthermore, the signs of the regression coefficients in the belief models are always in the correct direction. For example, in the automobile-minus-car-pool model, car-pool factors F1 and F2 have negative signs, but single-occupant-automobile factors F1 and F5 have positive signs. This shows that the higher use of the single-occupant automobile relative to that of the car pool is associated with a more positive perception of driving alone than of riding or driving in an automobile with others.

When models calibrated with time and cost data are compared with models calibrated with sociodemographic data, it appears that sociodemographic data are marginally more highly correlated with modal use. The multiple coefficients of determination for the sociodemographic data are always statistically significant by at least the 0.05 level. The time and cost data are not statistically significant for the automobile-minus-car-pool model, because the automobile and car-pool impedances are the same except for a constant added for the picking up of passengers. However, the signs of the regression coefficients for the time and cost data were not always correct. For example, in the automobile-minus-bus model, the cost term had a statistically significant positive sign. (Automobile cost is composed of several elements, some of which the individual may not consider to be work-trip related. It is possible that the individual considers only the daily out-of-pocket costs, such as parking fees, when evaluating the use of the automobile for the work trip. In any case, several attempts made to lower the objective measure of automobile cost had the effect of changing the sign or making the cost-coefficient statistically insignificant. However, the multiple coefficient of determination for the planning data remained unchanged.) The signs of the regression coefficients for the sociodemographic data do not appear to be counter intuitive.

Models Based on Two or Three Types of Data With No Interaction Terms

Table 4 gives the multiple coefficients of determination and the associated statistical significance levels for models based on two or three types of data with no explicit interaction terms. The four specific sets of types of data that are considered include two-way models of network time and cost, sociodemographic, and belief data and a three-way model that includes all three types of data.

A number of important details can be observed in Table 4. All the models are statistically significant beyond the 0.001 level. This result implies that any two-way or three-way model in the table can reliably estimate a trend of modal use with respect to its independent-variable set. However, the models based solely on time, cost, and sociodemographic data have multiple coefficients of determination that are approximately three to six times smaller than those of the models based on beliefs and either network time and cost or sociodemographic data. Therefore, perceived system attributes again show a strong correspondence with modal choice. Finally, the three-way model represents only a very minor improvement in the percentage of variance accounted for in the dependent variable in comparison to either of the two-way models that include beliefs.

Models Based on Two or Three Types of Data With Interaction Terms

Table 5 gives the multiple coefficients of determination for models calibrated with two or three types of data that have explicit terms for interaction among the data types. These interaction terms are formed by the product of the corresponding single type-of-data terms. There are a total of 11 models with explicit interaction terms. Those models that systemically incorporate two-way interactions before the final inclusion of a three-way interaction set of terms were designed to facilitate hypothesis testing based on analysis of the variance of the components as discussed by Appelbaum and Cramer (16), Cohen (17), Dobson (18), Kerlinger and Pedhazur (9), and Overall and Klett (11).

There is only one model described in Table 5 that does not include any belief data. This model, which is given in the first row, has multiple coefficients of determination that are substantially lower than those of the models that are based partly on belief data. This relation highlights the potency of belief data in accounting for the variation of modal use. In general, the addition of more or higher level interaction terms increases the percentage of variation accounted for by a model. For example, the last model described, which includes all possible two-way interaction terms and the set of three-way interaction terms, has larger multiple coefficients of determination than any other across the three dependent-variable criteria. Interaction terms enhance the size of the multiple coefficient of determination, but they vastly complicate the models. A model specified by NT + SD has 6 terms plus a constant, but one specified by NT + SD + (NT * SD) has 14 terms plus a constant, and the last model in Table 5 contains 75 terms, including the constant!

Hypothesis Testing Based on a General Procedure for Analysis of the Variance of the Components

While the summary statistics given in Tables 3, 4, and 5 are informative, they do not show the relative statistical significance of the different types of data. The general procedure for the analysis variance of the components described in the discussion of Equations 1 through 5 permits testing for the relative statistical significance of alternative classes of data based on their multiple coefficients of determination.

The procedure for assessing the statistical significance of the interaction terms proceeded in a hierarchical fashion. R^2 [NT + SD + B + (SD * B) + (NT * SD) + (NT * B) + (NT * SD * B)] was compared with R^2 [NT + SD + B + (SD * B) + (NT * SD) + (NT * B)]. The three-way interaction set of terms (NT * SD * B) was not found to be statistically significant. The statistical significance of the two-way interaction denoted by (NT * SD) was tested through the comparison of R^2 [NT + SD + B + (SD * B) + (NT * B) + (NT * SD)] with R^2 [NT + SD + B + (SD * B) + (NT * B)]. It was not statistically significant and neither were any of the other two-way sets of interaction terms. These findings of nonsignificance were invariant across the dependent-variable criteria. As a consequence of the lack of statistical significance for the interaction terms, the statistical significance of NT, SD, and B was tested through the set of models without interaction terms that are summarized in Tables 3 and 4.

Table 6. Tests of significance for network, sociodemographic, and belief data.

	Frequency of Use								
	Automobile - Bus			Automobile - Car Pool			Bus - Car Pool		
Type of Data	Significance	Variables Held Constant	Model	Significance	Variables Held Constant	Model	Significance	Variables Held Constant	Model
B	$p < 0.001$	NT	B + NT	$p < 0.001$	NT	B + NT	$p < 0.001$	NT	B + NT
	$p < 0.001$	SD	B + SD	$p < 0.001$	SD	B + SD	$p < 0.001$	SD	B + SD
	$p < 0.001$	NT + SD	B + NT + SD	$p < 0.001$	NT + SD	B + NT + SD	$p < 0.001$	NT + SD	B + NT + SD
SD	$p < 0.001$	NT	SD + NT	$p < 0.001$	NT	SD + NT	$p < 0.05$	NT	SD + NT
	$p < 0.01$	B	SD + B	$p < 0.01$	B	SD + B	NS	B	SD + B
	$p < 0.001$	NT + B	SD + NT + B	$p < 0.01$	NT + B	SD + NT + B	NS	NT + B	SD + NT + B
NT	$p < 0.001$	SD	NT + SD	NS	SD	NT + SD	$p < 0.01$	SD	NT + SD
	NS	B	NT + B	NS	B	NT + B	NS	B	NT + B
	NS	SD + B	NT + SD + B	NS	SD + B	NT + SD + B	NS	SD + B	NT + SD + B

Table 6 summarizes the results for the tests of significance for the network, sociodemographic and belief data sets when one or both of the other types of data are held constant. The belief data are shown to be uniformly statistically significant at beyond the 0.001 level, no matter which combination of variables they are compared to or what dependent variable is being considered. This is the only type of variable to demonstrate such a strong and unequivocal relation with modal use.

The sociodemographic data demonstrate a consistently strong relation to modal use beyond the 0.01 level of significance across the automobile-minus-bus use and automobile-minus-car-pool-use dependent-variable criteria. However, it is nonsignificant for two of three tests with respect to the bus-minus-car-pool-use dependent-variable criterion. The network data show the weakest pattern of relation to modal use, being nonsignificant in seven of nine tests.

SUMMARY AND CONCLUSIONS

The principal goals of this paper were (a) to establish whether there is a relation between perceived system attributes and modal choice, (b) to compare the magnitude of any linkage between perceived system attributes and modal choice with those of the linkages between sociodemographic and network data and modal choice, and (c) to establish the degree to which a linkage between perceived system attributes and modal choice is dependent on the linkages of sociodemographic and network data to modal choice. It was clearly shown that perceived system attributes or beliefs about modes are strongly associated with modal choices. Furthermore, the correlation between beliefs and modal choice is substantially larger than the correlations between either sociodemographic or network data and modal choice. Finally, the linkage between perceived system attributes and modal choice cannot be accounted for by relations of sociodemographic or network data or both and modal choice.

There is unquestionably a strong association between perceived system attributes and modal choice. The multiple coefficients of determination for modal choice as a function of beliefs about modes ranged from 0.265 to 0.124, and these coefficients were always statistically significant beyond the 0.001 level. The relation between perceived system attributes and modal choice is not influenced by either sociodemographic or network data. Two-way and three-way interaction terms were found not to be statistically significant, and the statistical significance of the perceived-system-attribute terms is not diminished beyond the 0.001 level when they are combined with sociodemographic or network data in two-way and three-way models. Network data do not independently contribute to the estimation of modal choice when combined with belief data. In other words, belief variables account for variance in modal choice above and beyond time and cost data, but the reverse is not true. The sociodemographic data overlap somewhat less with the belief data, but there is still a tendency for diminution of the sociodemographic effect when it is combined with the belief set.

The models and hypothesis-testing features of the empirical analyses reported here are based on the assumption of linear relations between modal choice and various sets of predictor variables. For the majority of linearity tests that were computed (but are not reported here), the nonlinearity assumption can be rejected. However, a pattern of significant nonlinear deviations emerges for the automobiles versus bus comparisons. Even these significant deviations, nevertheless, do not result in substantial differences between R^2 and eta square, a nonlinear analog to R^2. Therefore, the nonlinear characteristics were not substantively important for the data sets studied here.

The sample used in this research—CBD workers—is potentially atypical and sets a basis for restricting the generality of conclusions. The modal split is 56.2 percent single-occupant automobile, 25 percent bus, and 18.8 percent car pool. The CBD restriction introduces an enhanced level of availability of non-single-occupant-automobile modes with respect to a more random areawide sample. Therefore, beliefs that favor buses or car pools can be more easily translated into actual modal use patterns. For corridor planning analyses and planning studies of short-term low-cost improvements, this situation may be more general than is commonly believed.

No single model is recommended by the research reported here. If the effect of travel time and cost needs to be measured, it is generally possible to calibrate statistically significant models. Models based on network data combined with sociodemographic data were always found to be statistically significant. However, the sign for the travel-cost variable was incorrect. The addition of belief data to a model with network or sociodemographic data or both substantially increases the percentage of variance accounted for relative to that of the old model. While curvilinear models, such as those based on a logit formulation, are preferable on logical grounds for modal-choice analysis, a generalized curvilinear model was only slightly superior to linear representations on empirical, statistical grounds.

The results reported here are apparently at variance with those previously found by Hartgen (4) and Dobson and Kehoe (5), neither of whom found a strong linkage between perceived system attributes and modal choices. However, there are substantial differences between this study design and theirs. Among the most important is that the earlier analyses used data from an areawide random survey while this study design is restricted to CBD workers, for whom it is appreciably easier, relative to areawide workers, to translate favorable beliefs about buses and car pools into actual modal choices.

On the other hand, these results support and extend the findings of Dobson and Tischer (6), Nicolaidis (7), and Spear (8), whose studies found attitudes to be significantly correlated with modal choices. Spear restricted his sample to CBD workers who were not automobile or transit captives. Nicolaidis conducted his study in a college town. Dobson and Tischer used a sample that was closely related to the one studied here. The earlier Dobson and Tischer study is extended by the consideration of more than just belief data to permit the evaluation of perceived system attributes with respect to other possible causal factors for modal choice.

ACKNOWLEDGMENTS

The California Department of Transportation (Caltrans) and the Southern California Rapid Transit District (SCRTD) kindly provided the data that permitted the calibration of models with network time and cost information for single-occupant automobiles and buses. Norman P. Roy of Caltrans and Patrick Conway of SCRTD were particularly helpful in making information available. The advice of David S. Gendell, Thomas A. Hillegass, and Samuel L. Zimmerman, all of the U.S. Department of Transportation (U.S. DOT), was of assistance in the development of the network coding standards and the treatment of the network data for incorporation in modal-choice models. The respondent data were collected through a contract with a market research firm. Finally, William E. McGarvey and Mark T. Stahr, both of U.S. DOT, are gratefully acknowledged for assistance in the running of computer programs and the preliminary analysis of results. We accept full responsibility for the accuracy of the findings and the conclusions derived from the findings.

REFERENCES

1. T. F. Golob. Attitudinal Models. In Urban Travel Demand Forecasting (D. Brand and M. L. Manheim, eds.), HRB, Special Rept. 143, 1973, pp. 130-145.
2. T. F. Golob and R. Dobson. Assessment of Preferences and Perceptions Toward Attributes of Transportation Alternatives. In Behavioral Demand Modeling and Valuation of Travel Time (P. R. Stopher and A. H. Meyburg, eds.), TRB, Special Rept. 149, 1974, pp. 58-81.
3. R. Dobson. Uses and Limitations of Attitudinal Modeling. In Proc., 2nd International Conference on Behavioral Travel Demand (P. R. Stopher and A. H. Meyburg, eds.), Elsevier, New York, 1976.
4. D. T. Hartgen. Attitudinal and Situational Variables Influencing Urban Mode Choice: Some Empirical Findings. Transportation, Vol. 3, 1974, pp. 377-392.
5. R. Dobson and J. F. Kehoe. Disaggregated Behavioral Views of Transportation Attributes. TRB, Transportation Research Record 527, 1974, pp. 1-15.
6. R. Dobson and M. L. Tischer. Beliefs About Buses, Car Pools, and Single-Occupant Autos: A Market Segmentation Approach. Proc., Transportation Research Forum, Vol. 17, 1976, pp. 200-209.
7. G. C. Nicolaidis. Quantification of the Comfort Variable. Transportation Research, Vol. 9, 1975, pp. 55-66.
8. B. D. Spear. A Generalized Attribute Variable for Models of Mode Choice Behavior. TRB, Transportation Research Record 592, 1976, pp. 6-11.
9. F. N. Kerlinger and E. J. Pedhazur. Multiple Regression in Behavioral Research. Holt, Rinehart, and Winston, New York, 1973.
10. J. Johnston. Econometric Methods. McGraw-Hill, New York, 1972.
11. J. E. Overall and C. J. Klett. Applied Multivariate Analysis. McGraw-Hill, New York, 1972.
12. P. R. Stopher. Goodness-of-Fit Measures for Probabilistic Travel Demand Models. Transportation, Vol. 4, 1975, pp. 67-83.
13. N. H. Nie, C. H. Hull, J. Jenkins, K. Steinbrenner, and D. H. Bent. Statistical Package for the Social Sciences. McGraw-Hill, New York, 1975.
14. W. J. Dixon, ed. Biomedical Computer Programs. Univ. of California Press, Berkeley, 1975.
15. R. K. Blashfield. Mixture Model Tests of Cluster Analysis: Accuracy of Four Agglomerative Hierarchical Methods. Psychological Bulletin, Vol. 83, May 1976, pp. 377-388.
16. M. I. Appelbaum and E. M. Cramer. Some Problems in the Nonorthogonal Analysis of Variance. Psychological Bulletin, Vol. 81, 1974, pp. 335-343.
17. J. Cohen. Multiple Regression as a General Data-Analytic System. Psychological Bulletin, Vol. 70, 1968, pp. 426-443.
18. 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.

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

**Mr. Dobson is now with Charles River Associates, Cambridge, Massachusetts.*

Development of Market Segments of Destination Choice

Peter R. Stopher, Transportation Center, Northwestern University

This paper discusses the development of individual-choice models of the destination choice of nongrocery shopping locations. Two key features of the approach are the use of perceptual data for characterizing alternative destinations and an attempt to segment the population before the model building on the basis of homogeneity of perceptions of destinations. Data were obtained about the perceptions of shoppers of several shopping locations and on their preferences for various attributes of the shopping locations. The attributes were selected as those that make up the image of a shopping location independent of the transportation system. Several techniques are discussed for segmenting the population by perception, all of which are based on analysis of the psychological distance between shopping locations. Given the special properties of psychological distances, two forms of analysis were undertaken. First, correlations were computed for the set of interpoint distances for each socioeconomic group identified in the data. High correlations indicate similarity of perceptual space, while low correlations indicate lack of similarity. Second, the group interpoint distances were used as inputs to an individual scaling process that attempts to fit the perceptions into a common perceptual space by stretching or compressing the axes of the space by obtaining weights on each axis for each observation. In this case, market segmentation was sought through a hierarchic, fusion clustering process on the axis weights for each socioeconomic group. The results of these analyses converge well. Length of residence and age were found to be important segmentation variables. Sex and income were not found to be very powerful segmentation variables, but occupation may be worth study as a basis for segmentation.

In the last decade, there have been numerous developments in the formulation, refinement, and operation of travel-demand models at the level of the individual trip maker (1), by using behavioral constructs from psychology and microeconomics. For various reasons, most of this work has taken place in the subchoice of travel mode, primarily that for the work trip. From time to time, extensions of the behavioral approach beyond the modal-choice process have been proposed, but little progress had been made in such extensions until recently.

A major problem in achieving such extensions is the characterization of the elements of utility of other subchoices. In modal-choice models, utility was characterized initially in terms of the physical attributes of