PREDICTIONS OF INTERCITY MODAL CHOICE FROM DISAGGREGATE, BEHAVIORAL, STOCHASTIC MODELS

Peter L. Watson, Transportation Center, Northwestern University

This paper reports the results of some empirical prediction tests on disaggregate, behavioral, stochastic models of modal choice. It is discovered that the expected-number calculation (summation of probabilities) yields modal-choice predictions with a high degree of accuracy, although a caveat is in order regarding the transferability of such models: They should be transferred with care and only to similar situations. The stability of the models is evidenced by the fact that they can be broken down by trip purpose and reaggregated without loss of predictive power. Finally, the method of obtaining predictions by classifying probabilities is considered and rejected.

•IN RECENT years, work in the area of modal choice has tended to concentrate on the development of disaggregate, behavioral, and stochastic models. (Modal choice should be interpreted as including route choice.) Those developments have led to a model type that has a number of advantages over the traditional zonal-based models. First, the building of models at the disaggregate level with the individual as the basic decision-making unit means that problems of zonal homogeneity are avoided and that better use is made of the data collected. Second, because the models are based on the observable behavior of the individual traveler, they are more realistic. Third, the introduction of the stochastic element means that modelers are now concerned with the probability that a given mode will be chosen rather than with the modal split of a set of travelers from a zone. [A more detailed and comprehensive statement of the advantages of disaggregate, behavioral, and stochastic models is given by Stopher and Lisco (1).] The result of these modifications is that a generation of models has evolved that perform most successfully in terms of their ability to describe and explain modal choices. The models both reflect acceptable hypotheses about traveler behavior and achieve acceptable levels of significance in a statistical sense.

However, if the new generation of models is to be a useful addition to the planner's set of tools, it must be demonstrated that the new models have acceptable prediction properties. A comparison of the prediction capabilities of both aggregate and disaggregate models would be desirable, but documented evidence on the aggregate models is not readily available. Moreover, the aggregate nature of the data used in zonal-based models means that they cannot be used to produce disaggregate predictions in the same situation. Until the 2 types of model can be tested on the same data set, one-sided statements of prediction capability must suffice.

It is the aim of this paper to investigate the capabilities of disaggregate, behavioral, and stochastic models of modal choice. An exposition of the prediction methodology is followed by a series of empirical tests that determine predictive performance under different conditions.

The tests were carried out by using data from the Edinburgh-Glasgow area modal-split study, which investigated modal choices in the Forth-Clyde corridor of the central lowlands of Scotland. Details of the study background, data collection, and model

Publication of this paper sponsored by Committee on Social, Economic and Environmental Factors of Transportation.

development are given in another report (2). The travel data represent medium-range intercity trips. The choice modeled is between car and train; each of those modes carried approximately 45 percent of total traffic. The main aim of the study was to examine the transferability of disaggregate models from intracity commuting situations to other trip lengths and purposes; thus, models were developed for 3 trip purposes—journey to and from work (JTW), business (BUS), and social-recreational (SOCREC)—and for total travel. The variables that resulted in the best explanatory models for each trip purpose are given in Table 1. The abbreviations in Table 1 are defined as follows:

TD Rel = time difference relative to total journey time,
CD Rel = cost difference relative to total journey time,
WW Tim = difference in walking-waiting time,
WW Rel = difference in walking-waiting time relative to total journey time,
TJT Ca = total journey time by car,
SUBCOS = cost of access-egress to and from station,
Ju Tra = number of walk, wait, and travel segments in train journey, and
Ju Diff = difference in number of segments by each alternative.

The "relative" versions of time and cost differences and the relative walk-wait time variable reflect the fact that a given time (or cost) saving becomes less important as the length (or cost) of the journey increases. The "journey unit" variables represent a variable developed to reflect the inconvenience of each trip by allocating 1 trip unit to each segment (walking, waiting, and riding) of the trip.

The models were estimated by using logit analysis (3) so that

$$P_t = \frac{e^{G(x)}}{1 + e^{G(x)}} \qquad (1)$$

where P_t is the probability of choosing the train, and G(x) is a linear combination of the explanatory variables.

METHODOLOGY

Two main methods of obtaining predictions from disaggregate models of modal choice have been advocated. The first proceeds by a summation of predicted choice probabilities, and the second is a classificatory approach. The following discussion of the methods of obtaining predictions is set in terms of a car-train choice, where the dependent variable is the probability that the train will be chosen. Having calibrated the model, one can calculate the probability of choosing the train for each individual traveler, and those probability estimates are used to derive the prediction of the number who will choose the train. The predictions are derived on the assumption that the important result from a prediction point of view is the expected number of travelers who will choose a given mode. Thus, the models are calibrated by using the individual traveler as the basic unit of analysis, and the results are then aggregated to produce results that are meaningful from a planning standpoint. (The appropriateness of this approach will be taken up again in the discussion of the criteria that will be used to evaluate the prediction results.)

Table 1. Model content.

Purpose	Variables
JTW	TD Rel, CD Rel, WW Rel, Ju Dif
BUS	TD Rel, CD Rel, WW Tim, Ju Dif
SOCREC	TJT Ca, SUBCOS, WW Tim, Ju TRA
Total	TD Rel, CD Rel, WW Tim, Ju Dif

Expected-Number Approach

This method derives from the fact that the expected value of a random variable is the sum of the products of the potential outcomes and the probabilities of their occurrence. Binary outcomes, coded as 0 or 1, are the special case of a Bernoulli random variable such that

$$E(p) = 1\,(P_T) + 0\,(1 - P_T) \qquad (2)$$

$E(p)$ = expected value of the probability, P_T = probability of choosing the train, and $1 - P_T$ = probability of not choosing the train (i.e., of choosing the car). Thus, the expected value is equal to the probability of choosing the train. To obtain the expected number of travelers who will choose the train (i.e., the expected number of positive outcomes in the sample) requires that the expected values be summed.

$$E(\#) = \sum_i E(P_{Ti}) = \sum_i P_{Ti} \tag{3}$$

$E(\#)$ = expected number of positive outcomes (train choices), and P_{Ti} = probability that the i th traveler will choose the train.

An example may clarify this point. Suppose that 10 travelers had the following observed probabilities and choices, where 1 = choice of train:

Choice	Predicted Probability
1	0.9
1	0.7
0	0.2
0	0.5
1	0.8
0	0.4
1	0.8
1	0.6
0	0.3
1	0.7

In this example, 6 travelers chose the train, i.e., there were 6 travelers with a probability of choosing the train equal to one; the sum of those probabilities is equal to the number of travelers choosing the train. The prediction of the number of train travelers (the expected number) is obtained by summing the predicted probabilities. That sum is equal to 5.9; therefore, it is predicted that 6 travelers will choose the train.

Classification Approach

This method is based on the premise that the predicted probabilities can be used to assign travelers to the alternative modes. Misclassifications are minimized by allocating travelers about a probability of 0.5. This means that, if the traveler's probability of choosing a train is greater than 0.5, he is assigned to the set of train travelers; if the probability is less than 0.5, he is assigned to the set of car travelers. The numbers assigned to each set then constitute the prediction of the number who will choose each alternative.

Discussion of Methods

To examine the relative performance of these 2 methods requires the establishment of criteria that will be used to judge them. Although the models that yield the predictions of probabilities are disaggregate models, in the sense that the unit of analysis is the individual traveler, testing their predictive abilities at a disaggregate level is inappropriate for a number of reasons. The true probability that a given traveler will choose the train is unknown. It is, therefore, not possible to test the ability of the models to predict (accurately) the probability of a given choice. The known information is the choice that the traveler was observed to make, and it is possible to compare the predicted probability with the observed choice. That is the basis of the classification criterion that has been applied to predictions by classification. However, a consideration of the meaning of the probability of choice reveals that the use of classificatory ability as a test of a model that predicts probabilities is a meaningless test. Consider, for example, a traveler for whom the probability of choosing the train is

0.6. That probability can be interpreted as implying that, for a large number of observations, the traveler would choose the train 60 percent of the time. The classification approach would, however, assign him to the train group and, if he were observed to take the car, would treat it as an erroneous classification.

Thus, neither the expected-number approach nor the classification approach should be judged on its ability to predict either the probabilities or the actions of individual travelers. That is not a serious problem, however, because the use of disaggregate models in a planning context requires that aggregate, not disaggregate, predictions be derived. This leads to an evaluation of both the prediction approaches and the predictive capabilities of the disaggregate models themselves in terms of their abilities to produce accurate aggregate predictions of the number of travelers who will choose a given mode.

PREDICTION TESTS

Although the procedures that are used to derive predictions are very simple, a caveat is in order before the prediction tests are undertaken. The most appropriate test of a predictive model involves 2 data collection efforts: one to calibrate the model and the other, preferably after a system change, to test the predictions. Such an effort is seldom possible. Thus, it is necessary to employ a second-best approach in which both the model and the predictions must be tested on the same data set. The more ideal situation of 2 random samples is approximated from the population by the following procedure. The data set is randomly divided into halves: One half is used to calibrate the models, and the other is used for the prediction tests. The tests described in this paper employ a double-edged version of this procedure in which coefficients are estimated for both halves of the data set, after which the coefficients and data sets are interchanged to obtain 2 sets of predictions, which are summed to provide the final prediction. It is acknowledged that this testing procedure is not ideal, but, in the absence of more extensive data collection facilities, it is the best available.

After the models given in Table 1 are calibrated, the expected-number procedure was used first to obtain predictions of the expected number of train travelers. The results are given in Table 2. It is clear that the differences between the number of travelers who are observed to choose the train and the number who are predicted to choose the train are small. Two tests are proposed that will indicate the magnitude of the prediction errors. The first considers the difference between the predicted and observed numbers of train travelers and expresses the error in prediction as this difference relative to the observed number. In other words, ϵ_1 indicates how close the model comes to providing a perfect prediction of the number of train travelers. The second takes a rather broader view of the prediction process and examines it from the point of view of modal split. Given the nature of the predictions, i.e., that the probability of choosing the train, P_T, is equal to one minus the probability of choosing the car, $1 - P_T$, the models also provide a prediction of the number of travelers who will choose the car, $\sum_i (1 - P_{Ti})$. It is obvious that the difference between predicted and

Table 2. Predictions of number choosing train.

		Set 1		Set 2		Sets 1 and 2		
Approach	Purpose	Predicted	Observed	Predicted	Observed	Predicted	Observed	Number
Expected number	JTW	125	117	123	135	248	252	360
	BUS	245	249	241	235	486	484	878
	SOCREC	241	237	208	211	449	448	1,202
	Total	582	589	601	595	1,183	1,184	2,440
Classification	JTW	138	117	143	135	281	248	360
	BUS	263	249	260	235	523	484	878
	SOCREC	148	237	182	211	330	448	1,202
	Total	574	589	609	595	1,183	1,184	2,440

observed train travelers will be equal (although a different sign) to the difference between predicted and observed car travelers. That being the case, the number of erroneous predictions can be regarded as an indicator of the failure of the model to predict correctly the modal split and may be expressed relative to the sample size to provide an alternative measure of prediction performance ϵ_2. Thus,

$$\epsilon_1 = \frac{\text{OBS - PRED}}{\text{OBS}} \times 100$$

and

$$\epsilon_2 = \frac{\text{OBS - PRED}}{\text{N}} \times 100$$

Applying those error measures to the predictions given in Table 2 yields the following percentages:

Measure	JTW	BUS	SOCREC	Total
ϵ_1	1.59	0.41	0.22	0.08
ϵ_2	1.11	0.23	0.08	0.04

Although it may be argued that ϵ_2 overstates the performance of those models, it is clear that they predict extremely well.

Next, the classification approach was used to obtain predictions of the number of train travelers. These predictions are essentially different from those presented in the preceding section because they derive not from the summation of probabilities but from the use of the estimated probabilities to assign travelers to modes. Thus, the basis of this method is a classification procedure that assigns travelers with an estimated probability of more than 0.5 to the train and those with an estimated probability of less than 0.5 to the car. It should be noted at this point that the estimated probabilities that are used were obtained from the random-division estimations. The resulting predictions are also given in Table 2.

The error percentages associated with the predictions by classification are as follows:

Measure	JTW	BUS	SOCREC	Total
ϵ_1	13.31	8.06	26.34	0.08
ϵ_2	9.17	4.44	9.82	0.04

In evaluations of the predictive performance of these models, both absolute and comparative with regard to the alternative methods of obtaining the predictions, it is important to note that the tests carried out reflect the ability of the models not simply to replicate the data on which they were calibrated (the nature of the calibration technique means that such a procedure yields perfect replications) but to predict travelers from a second random sample that is taken from the same population as the sample used to calibrate the model.

In the light of this consideration, the predictions obtained from the expected-number approach are remarkably accurate. If only the ϵ_1's are considered, the highest error among the 4 models is 1.59 percent. The classification approach does not perform so well, and the errors associated with the predictions derived by classification are several orders of magnitude greater than those by the expected-number procedure.

Several factors may contribute to the relative failure of the classification approach. It was argued above that classification implies the ability to make precise inferences about an individual's action from his predicted probability. However, probability statements are only meaningful from a prediction point of view when aggregated either for the population of travelers or for multiple trips by the individual traveler. It seems very likely that the failure of the classification method to result in accurate predictions is a result of the inappropriate attempt to match predicted probabilities with specific actions.

PREDICTIONS WITH NONRANDOM DIVISIONS

It has been argued that the random-division procedure is improper when large data sets are used because the new distributions should differ little from the parent distribution and, therefore, the process of randomly dividing the data should result in little change. It is against that background that the non-random-division tests are advocated. The argument that nonrandom divisions should be used to counter the above problems is not very attractive because such divisions are usually based on income (or other socioeconomic characteristics). If one believes that those factors affect modal choice, it is unreasonable to expect that a model built for one income group would predict well for travelers in a different income group. In short, it is argued that a model should be able to predict well in the case of a further sample drawn from the original population, (i.e., a sample with similar distributional characteristics to the one used to calibrate the model); it is not reasonable to expect a model to predict well when presented with a new set of data whose distributional characteristics are different from those of the calibration data, whether the differences are the result of non-random-resampling procedures or even the transfer of the model to a new, but different, situation.

COMPOSITION OF THE PREDICTIONS

Having undertaken the above detour to deal with the question of nonrandom divisions, it is now appropriate to return to the main stream of this paper. The prediction results presented above were obtained for 3 trip purposes—journey to and from work, business, and social-recreational—and for total travel. It should be emphasized at this point that a different model was constructed for each trip purpose, for it was found that the same model did not explain modal choices equally well in each case. Thus, the models used in this paper represent the ones that may be considered best, in the statistical sense, for each trip purpose. In the same way, the model for total travel is the model that best explains the modal choices of all travelers regardless of trip purpose. A comparison of the predictions from the total-trip model with the sum of the predictions from the models for the different trip purposes is as follows:

Item	Sum	Total
Predicted	1,183	1,183
Observed	1,184	1,184

The comparison of the combined results of the different trip purposes reveals that the predictions obtained are identical. That finding has some interesting implications for planners attempting to develop modal-choice models. In many cases, it is of interest to deal with different trip purposes separately. For example, in an investigation of staggered work hours, one might wish to consider the work trip in some detail and to ignore other purposes. Similarly, one who studies modal choice on trips to recreational sites may not wish to consider commuting trips. The results presented above have 2 interesting features in this regard. First, it is possible to isolate specific trip purposes, which may be modeled and used for predictive purposes, without being affected by the omission of the remaining trip purposes. Given the fact that planners are not always concerned with global changes, it is clearly useful to be able to separate the trip purposes. Second, it may further be implied that the models need not necessarily all be built at the same time. That may be an important consideration for the planner who may not be able to mount an effort of sufficient scale to yield models of all types simultaneously.

In more general terms, the evidence presented above indicates that it is possible to build separate disaggregate models of modal choice for different trip purposes. Moreover, all the models predict with a high degree of accuracy. For a planner concerned with subregional problems, this is an important finding. It is also demonstrated that the sum of the predictions from the separate models is equivalent to the prediction from the model for all trip purposes. The intent is not to place great emphasis on the possibility that models built separately may be combined in order to predict overall mode choices; nevertheless, an important feature of these models is that separate models may be built without impairing their joint ability to predict modal choices for all travelers.

CONCLUSIONS

A number of interesting conclusions may be drawn from the above investigation, but first it should be noted that the results are based on the empirical investigations of the Edinburgh-Glasgow area modal-split study data. As such, the results are data specific and require corroboration. It is argued that the sizes of the data sets and the consistency of the results imply generality; this should not, however, be taken as proved.

The first, and most important, conclusion is that disaggregate, behavioral, and stochastic models of modal choice are able to predict modal choices with an extremely high degree of accuracy, in some cases with prediction errors of less than 1 percent. It would have been interesting to compare these error properties with those of aggregate models based on the zone as the unit of analysis. However, the error properties of aggregate modal-choice models are not well documented, and few data sets exist that are amenable to both aggregate and disaggregate analysis. An interesting area for future research would be an explicit comparison based on the dual analysis of a single data set. In the meantime, the disaggregate results must stand alone. There can be no doubt, however, that disaggregate models predict with a high degree of accuracy.

The second conclusion is that the expected-number approach is to be preferred for deriving predictions from disaggregate models of modal choice. The classification approach is conceptually inferior, in the sense that its attempts to match predicted probabilities and observed behavior are inconsistent with the interpretation of the predicted probability. The classification approach is also an inferior predictor, as demonstrated by the fact that it produces predictions of modal-choice behavior whose prediction errors are several orders of magnitude greater than those of expected-number predictions. On both counts, the classification approach must be rejected in favor of the expected-number approach.

The third conclusion concerns the potential ability of disaggregate models to predict in circumstances that are spatially or temporally different from those under which the model was developed. The conclusion derives from the discussion of the random-division testing technique. It is argued that models should only be expected to predict well in new situations when the underlying distributions are similar to those on which the model is based. Thus, an intracity modal-choice model should not be expected to predict transatlantic mode choice; nor would a model developed for one income group necessarily predict well for another income group.

The fourth conclusion is that disaggregate models of modal choice built separately for different trip purposes yield accurate predictions for each trip purpose. In addition, the results of the separate models may be combined to yield accurate predictions of overall modal choice. That finding may have interesting implications for planners interested in subregional analyses.

In more general terms, it may be concluded that the results presented above are extremely promising and augur well for the use of disaggregate, behavioral, stochastic models in a predictive framework. A caveat is in order at this point: The predictions under consideration in this paper are point estimates. It is important that confidence intervals be developed in order that the prediction range may be assessed. It is hoped that future work will be undertaken, both to develop confidence intervals and to confirm the results of this analysis, that will lead to the use of disaggregate modeling techniques as a useful planning tool.

REFERENCES

1. Stopher, P. R., and Lisco, T. E. Modelling Travel Demand: A Disaggregate, Behavioural Approach, Issues and Implications. Proc., Transp. Res. Forum, 1970.
2. Watson, P. L. The Value of Time and Behavioural Models of Mode Choice. Univ. of Edinburgh, PhD thesis, 1972.

3. Watson, P. L. Choice of Estimation Procedure for Binary Models of Travel Choice: Some Statistical and Empirical Evidence. Northwestern Univ., Evanston, Ill., 1972.
4. Westin, R. M. Prediction From Binary Choice Models. Paper presented to Econometric Soc., Toronto, 1972.