COMPARISON OF THE MODEL STRUCTURE AND PREDICTIVE POWER OF AGGREGATE AND DISAGGREGATE MODELS OF INTERCITY MODE CHOICE

Peter L. Watson, Northwestern University

This paper reports the construction of both disaggregate and aggregate models of intercity mode choice; data from the Edinburgh-Glasgow Area Modal-Split Study were used. The models are then compared in terms of their structure (i.e., the variables included) and their ability to predict modal split. The disaggregate models provide a better statistical explanation of mode-choice behavior. Moreover, the failure of the variable representing relative travel time to reach a satisfactory level of statistical significance in the aggregate models indicates that the alleged behavioral nature of aggregate models is not supported by the empirical evidence. In addition, the predictions of modal split derived from the aggregate models are inferior to those obtained from the disaggregate models. Several tests show that the errors associated with the aggregate models are several times as large as those associated with the disaggregate models. Disaggregate models have extremely desirable performance characteristics. It is, therefore, time to make serious effort to incorporate them into the transportation planning process.

• THE RECENT developments in the use of disaggregate, behavioral, stochastic models in the area of transport mode choice have been most encouraging in terms of both the explanatory and the predictive powers of the models. From a transport planning point of view, however, the more important feature is the ability of the models to predict the behavior of travelers. It has been shown that the disaggregate models of mode choice can be used to predict aggregate mode-choice behavior with a high degree of accuracy (7) and also that models calibrated in one situation may be transferred spatially to yield accurate predictions of behavior in other situations (10). These results, however, suffer from one serious drawback: Although they demonstrate the predictive efficiency of disaggregate models, comparisons with the predictive abilities of the aggregate or zonal mode-choice models commonly used by urban transportation planners are difficult. This difficulty is the result of 2 main factors. First, evidence on the errors associated with predictions from zonal mode-choice models is not readily available. Although it has been claimed that the errors may run as high as 300 percent, no evidence has been produced to support such claims. Second, because zonal models use the aggregate data, the aggregate data sets cannot be used to construct disaggregate models for the purposes of comparison. Thus, the controversy over which type of model is better has continued in the absence of either a means (in terms of data) or a method (in terms of error results for zonal models) of making a comparison and, hence, of reaching a conclusion. The analysis reported in this paper offers a way out of this impasse by way of a single data set that is sufficiently versatile to allow the construction of both an aggregate and a disaggregate model. These models are used to produce predictions of modal split, and the errors associated with these predictions may then be compared. Thus, the relative merits of disaggregate and aggregate models of mode choice may be assessed.

This research attempts to compare current practice, as embodied in the urban transportation planning package of models, with practice as recommended by the developers of the disaggregate models. Thus, each type of model is presented in such a way as to represent its use in a planning context.

STUDY BACKGROUND AND DATA

The data used to perform these tests are derived from the Edinburgh-Glasgow Area Modal-Split (EGAMS) Study (8). The data collected in this study (which represent all journey purposes) were originally used to build disaggregate models of mode choice for medium-range, intercity journeys in the Forth-Clyde corridor of the Central Lowlands of Scotland. This model development effort represents the basis for the predictions of modal split used to represent disaggregate models in the comparison reported below.

Each individual observation in the EGAMS Study has its origin and destination coded to the traffic zones developed for use in the Land Use and Transportation Study for South-East Scotland and the Greater Glasgow Transportation Study. The result of this method of coding is that all observations may be allocated to pairs of zones, which are the same zones used in actual zonal analysis of this area. Thus, the zones used in the development of the aggregate models in this paper are not an academic construct; rather they are the zones used in 2 actual transportation studies.

In fact, the number of zones used in this analysis is smaller than the number of zones included in the EGAMS study area. A number of zones were eliminated because no trips either originated or terminated in them. Then the zonal boundaries were redrawn to combine existing zones into larger zones while maintaining regional characteristics and contiguity. Since most of the "empty" zones were in areas that were peripheral to the cities themselves, the reconstruction of the zones emphasizes the intercity nature of the choice being analyzed. Each city was divided into 10 zones; an identifiable central area was surrounded by a group of zones divided approximately by geographical quadrants. Identifying the zone-to-zone pairs that represented an intercity trip yielded a 200-cell trip matrix. Since 42 of those interzonal pairs were empty (i.e., no trips were observed between that pair of zones), 158 zone-to-zone pairs remained for analysis.

The number of observations in each zonal pair ranged from 1 to 101. Although the sample sizes are somewhat smaller than would be encountered in a transportation study (as a result of the fact that the total disaggregate data set is limited), the range of observations and the existence of cells that are either empty or that contain a small number of observations are a realistic representation of the data available to a transportation study.

The mean value for each variable was calculated for each zonal pair, and these means were used as the independent variables in the analysis. The dependent variable was the proportion of travelers in each zonal pair who chose to travel by train. The aggregate models were calibrated by multiple regression analysis; given the fact that the dependent variable in the aggregate case is the proportion of travelers choosing the train, the problems associated with a binary (coded 0 or 1) dependent variable do not arise, i.e., heteroscedasticity and invalid tests of significance. The problems of out-of-range predictions are not eliminated. However, the results presented by Watson (9) indicate that this is not a serious problem. Thus, to use logit analysis for the aggregate model was not considered necessary. [Tests where logistic transformations were made are discussed by Watson and Westin (10).] By contrast, the disaggregate models were calibrated by logit analysis; the data for each individual traveler made up the inputs to the model.

MODEL STRUCTURE

The first stage of the comparison between the disaggregate and the aggregate models takes the form of an examination of the structure of the models. The term "structure" may be interpreted in a number of ways; for example, the use of either multiple regression or logistic analysis intrinsically imposes a structure on the models. In this sense, the structural difference between the disaggregate and aggregate models is self-evident. For comparative purposes, the structure of the models in terms of the variables that they contain is more interesting.

Disaggregate Model

The model calibrated by logistic analysis on the disaggregate data set took the following form:

$$P(T) = F(\text{TD REL, CD REL, JU DIF, WW TIM})$$

where the relative time difference, TD REL, or relative cost difference, CD REL, is the difference in time or cost between the 2 modes relative to the average time or cost of the journey; walking-waiting time, WW TIM, is the time spent walking and waiting during the train journey; and the journey unit difference, JU DIF, is the difference in the number of segments (walking, waiting, and riding) associated with the journey by each mode. The estimated coefficients are as follows:

Variable	Coefficient	t-Value
TD REL	-1.050	-6.84
CD REL	-0.666	-8.95
WW TIM	-0.002	-9.45
JU DIF	-0.132	-5.95

The likelihood ratio value (χ^2, 4 d.f.) = 527.28. All the variables, and the equation as a whole, were statistically significant at the 0.01 level of significance; and all the variables had the hypothesized sign. The structure of this model parallels the models of intracity commuter mode choice (3, 5, 6). The original type of model has been modified by the inclusion of a proxy variable for convenience and by transforming the time and cost variables into a relative form that reflects the nature of the intercity journey.

Aggregate Models

Since the aggregate model is to be used to produce predictions that are comparable with the predictions from the disaggregate model, the first aggregate model tested was made up of the same 4 variables that had yielded the best disaggregate model. The coefficients, t-values, and related statistics are as follows:

Constant and Variables	Coefficient	t-Value
Constant	1.0107	9.89
JU DIF	-0.08607	-4.17
WW TIM	-0.41454	-1.45
TD REL	-0.14862	-1.04
CD REL	-0.32668	-5.19

$R^2 = 0.419$, and $F = 27.59$. Although the R^2 and F statistics for the significance of the equation as a whole are satisfactory, this model cannot be considered adequate as the coefficients of 2 of the independent variables are insignificant. Nevertheless, although this model is not the model that best represents the aggregate data, it will be used to provide predictions for comparative purposes since it corresponds directly with the disaggregate model.

Clearly, to use an inferior model is to do less than justice to the aggregate modeling technique; thus, the aggregate data were analyzed in order to find a better model. Trials with numerous variable combinations and choice hypotheses revealed that it was not possible to use variables representing both time and cost differences in the same equation without the coefficient on the time variable becoming insignificant. The model that was finally judged to be the best model includes the cost difference variable plus 3 variables that were used in earlier disaggregate model tests to represent the level of convenience and accessibility associated with the train journey. The variables are the cost of station access and egress, SUBCOS; the time spent walking and waiting on the train journey, WW TIM; and the difference in the number of journey units (segments) associated with the journey, JU DIF. The coefficients and t-values are as follows:

Constant and Variables	Coefficient	t-Value
Constant	1.038	10.88
CD REL	-0.312	-5.20
SUBCOS	-0.193	-3.010
WW TIM	-0.005	-2.199
JU DIF	-0.080	-3.92

$R^2 = 0.453$, and $F = 31.698$. This model is similar to the model derived in disaggregate tests for the group of social-recreational travelers, in which the inconvenient features of the train journey (SUBCOS, WW TIM, JU TRA) were contrasted with the best features of the automobile, i.e., its speed (TD REL). In this case, the model may be interpreted as meaning that the inconvenient features of the train journey are compared with the least attractive feature of the automobile, i.e., its cost (CD REL). This hypothesis does not seem at all unreasonable; thus, this model will be used as the best aggregate model to obtain the modal-split predictions that will then be compared with the predictions from the disaggregate model.

Comparison of Model Structure

The differences in the statistical significance of the variables tested on the aggregate data set led to an important conclusion. The disaggregate models are based on the behavioral hypothesis that the traveler makes his or her mode-choice decision on the basis of the relative times costs and other characteristics of the modes available to him or her. The statistical tests performed on the models fail to reject this hypothesis. It has been argued (2) that the aggregate models are also behavioral. However, the hypothesis that the proportion of travelers choosing the train depends on the relative characteristics of the modes is not confirmed by the statistical tests. Thus, although the aggregate models may be behavioral in concept, the behavioral hypothesis is rejected by the data and the behavioral nature of the aggregate models must be placed in doubt. Simply to choose variables on the basis of a behavioral hypothesis is not sufficient if the data do not support that hypothesis.

In summary, the data aggregation process conceals the behavioral basis of the mode-choice decision, and the disaggregate approach yields models that are both behaviorally and statistically sound.

PREDICTIVE POWER

The second stage of the comparison involves examining the predictive power of the disaggregate and the aggregate models. In the absence of a second, predictive data set, the prediction tests were carried out by using 2 data sets representing 2 random drawings from the original population; these data sets were obtained by randomly dividing the sample into 2 halves. It is acknowledged that this is a less than optimal testing procedure; it is, however, the best available under the circumstances and can provide useful insights into relative predictive abilities.

The intercity flows are regarded as a corridor, and the tests developed represent the relative accuracy of predicting the modal split for the corridor as a whole. The prediction error is presented in 2 forms. The first, ϵ_1, is defined as the absolute difference between the predicted and actual mode split; it may also be interpreted as the percentage of the sample for whom erroneous predictions of mode split are made. The second considers the error as a percentage of the actual modal split. Clearly, the second form will yield a higher error than the former, the difference being diminished as the actual modal split approaches 100 percent on one of the alternative modes. Both error measures are presented for comparative purposes.

Disaggregate Predictions

The properties of predictions of modal split from disaggregate models have been reported elsewhere (8, 9); therefore, these results are given in a somewhat abbreviated form in Table 1. These results require little explanation or comment. It is evident that the prediction errors associated with disaggregate models are low. The mean errors are presented together with the errors associated with summing the predictions from the 2 data sets to indicate that the errors tend to offset each other so that the total error is smaller than the errors for each data set. Thus, the better overall measure of error is the mean error.

Aggregate Predictions

To provide a broad picture of the errors associated with the aggregate model, 3 prediction tests were performed. The first 2 involve deriving predictions of the number of train travelers in each zone-to-zone pair; these predictions were then summed to yield a corridor prediction. The third test involves a different error concept. The tests were performed by using the models derived above. As it is based on the best statistical relation derivable from the aggregated data, model 2 is referred to in the following discussion as the best aggregate model. Model 1, which contains the best set of variables from the analysis of the disaggregate data, is referred to as the best disaggregate model. However, the terms "best aggregate" and "best disaggregate" refer to the derivation of the variable set in each model. The prediction tests are carried out on models built with the aggregated data.

Random-Split Predictions—The closest replication of the disaggregate model prediction tests involved randomly dividing the zone-to-zone pairs into 2 groups. The first data set contained 83 zone-to-zone pairs; the second contained 75. Each data set was used to calibrate the model: The data sets and coefficients were then interchanged and 2 sets of predictions were obtained. The predictions are given in Table 2.

Directional-Split Predictions—Since the procedure of randomly dividing the data into 2 sets for prediction tests has been criticized, a new method was derived that made use of the origin-destination characteristics of the zone-to-zone pairs to produce the 2 data sets: The first data set represents travel from Edinburgh to Glasgow (81 pairs), and the second data set represents travel from Glasgow to Edinburgh (77 pairs). Such a breakdown also provides some insight into the transferability of zonal models, i.e., the extent to which a model developed in one situation may be used to predict behavior in another. The results for the "best" models are given in Table 3.

Mean Prediction Errors—Since the effect of a transportation system improvement may be highly localized, the error associated with the prediction of modal split for a given zone-to-zone pair must be considered. The mean zone-to-zone prediction errors are given in Table 4 as indications of the errors associated with the prediction of mode split for the average zone. They were obtained by first calculating the absolute error in predicted modal split for each zone-to-zone pair; the mean of these errors was then obtained.

Comparison of Predictive Power

The predictions obtained from the aggregate models have much larger errors associated with them than the predictions from the disaggregate models. The results are particularly strong when one considers the different models and the different methods of obtaining the 2 data sets. Although the results from the random-split method are, as might be expected, better than the results from the directional-split method, the errors are still larger when compared with the errors from the disaggregate models. Moreover, the errors from the aggregate models improve only marginally when the best aggregate rather than the best disaggregate model is used. Moreover, the average prediction errors indicate that the ability to predict at a less than corridor level is suspect. (Regrettably, the data format is insufficiently flexible to allow the predictions from the disaggregate model to be broken down by zonal pairs without excessive data manipulation.) The fact that the errors are consistently high across models

Table 1. Disaggregate model predictions.

Item	Data Set 1	Data Set 2	Both Data Sets
Sample size, number	1,197	1,243	2,440
Predicted train, number	582	601	1,183
Actual train, number	589	595	1,184
ϵ_1, percent	0.59	0.48	0.04
ϵ_2, percent	1.19	1.00	0.08

Note: $\bar{\epsilon}_1$ = 0.535, and $\bar{\epsilon}_2$ = 1.095.

Table 2. Random-split data set predictions by aggregate model.

	Best Aggregate Model			Best Disaggregate Model		
Item	Data Set 1	Data Set 2	Both Data Sets	Data Set 1	Data Set 2	Both Data Sets
Sample size, number	902	1,040	1,942	902	1,040	1,942
Predicted train, number	434	595	1,029	435	659	1,094
Actual train, number	363	537	900	363	537	900
ϵ_1, percent	7.88	5.58	6.64	7.98	11.73	9.98
ϵ_2, percent	19.58	10.81	14.33	19.83	22.72	21.55

Note: $\bar{\epsilon}_1$ = 6.73 and $\bar{\epsilon}_2$ = 15.19 for aggregate models; $\bar{\epsilon}_1$ = 9.85 and $\bar{\epsilon}_2$ = 21.27 for disaggregate models.

Table 3. Directional-split data set predictions by aggregate model.

	Best Aggregate Model		Best Disaggregate Model	
Item	Glasgow to Edinburgh	Edinburgh to Glasgow	Glasgow to Edinburgh	Edinburgh to Glasgow
Sample size, number	582	1,360	582	1,360
Predicted train, number	253	931	240	946
Actual train, number	433	467	433	467
ϵ_1, percent	30.93	34.12	33.16	35.22
ϵ_2, percent	41.57	99.36	44.57	102.57

Note: $\bar{\epsilon}_1$ = 32.52 and $\bar{\epsilon}_2$ = 70.46 for aggregate models; $\bar{\epsilon}_1$ = 34.19 and $\bar{\epsilon}_2$ = 73.57 for disaggregate models.

Table 4. Mean zone-to-zone prediction errors by aggregate model.

	Error (percent)		
Model	Data Set 1	Data Set 2	Mean
Random split			
Best aggregate	18.51	25.18	21.85
Best disaggregate	20.48	24.04	22.20
Directional split			
Best aggregate	37.93	36.95	37.43
Best disaggregate	42.80	39.17	40.94

and data divisions is a clear indication that the errors result from more fundamental problems. Although the results cannot be used as direct evidence of this, the fundamental problem is more likely the information loss (1, 4) that is associated with the aggregate models. In the case of the disaggregate models, all the available data are used in the calibration of the models and the derivation of the predictions; in the aggregate approach, much of the information content of the data is lost when the mean values of the zone-to-zone pairs are used to represent the complete range of information. Although this information loss may be greater in this test than in a transportation study (because of the combining of zones), the prediction errors from the aggregate procedure are still extremely large and should cause serious concern.

CONCLUSIONS

The objectives of this paper are very simple: to produce comparable predictions of mode-choice behavior by applying aggregate and disaggregate methods to the same data. The results are unambiguous. The errors associated with the aggregate method are several times as large as those associated with the disaggregate method. Even taking into account the limitations necessarily imposed by the design of the test, these results must be interpreted as a clear demonstration of the predictive superiority of disaggregate models. These results, taken in conjunction with recent results (10) on the ability of disaggregate models to produce accurate predictions using no more data than are required by aggregate models, make it clear that disaggregate models of transport mode choice have made the transition from academic toys to serious transportation planning tools.

ACKNOWLEDGMENT

The data used in this analysis were derived from the Edinburgh-Glasgow Area Modal-Split Study (8). This study was financed by the Great Britain Ministry of Transport (now part of the Department of the Environment) as part of a research program into the value of time. The results and views expressed in this paper are not to be interpreted as the official position of the department. I am indebted to Tom Lisco, Charlotte Chamberlain, and Stein Hansen for comments on an earlier version of this paper. All remaining errors are, of course, my responsibility.

REFERENCES

1. Fleet, C. R., and Robertson, S. R. Trip Generation in the Transportation Planning Process. Highway Research Record 240, 1968.
2. Hartgen, D. T., and Wachs, M. Disaggregate Travel Demand Models for Special Context Planning—A Dissenting View. In Issues in Behavioral Demand Modeling and the Valuation of Travel Time. U.S. Department of Transportation, 1973.
3. Lisco, T. E. The Value of Commuters' Travel Time: A Study in Urban Transportation. Univ. of Chicago, PhD dissertation, 1967.
4. McCarthy, G. M. Multiple Regression Analysis of Household Trip Generation. Highway Research Record 297, 1969, pp. 31-43.
5. Quarmby, D. A. Choice of Travel Mode for the Journey to Work: Some Findings. Journal of Transport Economics and Policy, 1967, pp. 273-314.
6. Stopher, P. R. A Probability Model of Travel Choice for the Work Journey. Highway Research Record 283, 1969, pp. 57-65.
7. Watson, P. L. Predictions of Intercity Modal Choice From Disaggregate, Behavioral, Stochastic Models. Highway Research Record 446, 1973, pp. 28-37.
8. Watson, P. L. The Value of Time: Behavioral Models of Mode Choice. D. C. Heath and Co., Lexington, 1974.
9. Watson, P. L. Choice of Estimation Procedure for Binary Choice Models—Some Statistical and Empirical Evidence. Department of Economics, Northwestern Univ., unpublished manuscript, 1973.
10. Watson, P. L., and Westin, R. B. Transferability of Binary Choice Models: An Empirical Test. Department of Economics, Northwestern Univ., unpublished manuscript, 1973.