Guidelines for Aggregate Travel Prediction Using Disaggregate Choice Models

Frank S. Koppelman, The Transportation Center, Northwestern University

This paper describes procedures for aggregating disaggregate choice models after estimation of the choice model parameters to obtain an aggregated model structure. This structure consists of (a) a disaggregate choice model, (b) a representation of the distribution of explanatory variables, and (c) an aggregation procedure. A taxonomy of aggregation procedures that classifies models according to their structural characteristics is developed. Errors in prediction by use of alternative aggregation procedures are empirically estimated. Analysis of these errors leads to conclusions about the performance of different aggregation procedures. These conclusions suggest the following guidelines for aggregate travel prediction using disaggregate choice models: (a) Disaggregate choice models may be most effectively used for prediction at high levels of aggregation appropriate to policy analysis; (b) enumeration procedures should be used whenever adequate sample data are available, especially at high levels of aggregation; (c) when sample data are not available, classification procedures should be based on the most important class distinction, which will be differences in choice set when such differences exist; (d) when data are not available to predict class specific variable values, predictions by the naive procedure should be adjusted for differences in choice set when such differences exist; (e) the specification of the underlying disaggregate choice model should be developed and evaluated with particular care in the grouping of individuals with structurally different choice sets; and (f) incremental prediction should be used for prediction of the expected impacts of policy changes whenever an existing set of choice shares is available to use as a basis for adjustment.

A central component of transportation planning and policy analysis is the prediction of the future performances and impacts on the transportation system for each of the available plan or policy alternatives. These predictions should provide adequate information about the effects of each alternative so that an informed choice can be made among them and should distinguish among the effects of different policies with respect to travel flows, system performance, and external impacts. The evaluation and selection process requires that these predictions be at a level of aggregation that is relevant to the alternatives under study and the precision with which they have been formulated.

In contrast with the need for aggregate predictions based on aggregate descriptions of transportation plan and policy options, travel behavior theory is postulated at the level of the behavioral unit—usually an individual or household. Group behavior, which is the object of prediction, is the aggregation of numerous travel choices of individual behavioral units. This aggregation is implicit in the use of models that are estimated on mean value aggregate data. When disaggregate choice models are used the aggregation is accomplished by use of an explicit aggregation procedure. The aggregate prediction will be sensitive to both the structure of the disaggregate choice model and the distribution of independent variables in the population or prediction group (1).

There are two approaches to the development of aggregate prediction models that are consistent with underlying disaggregate behavior. The first approach is to aggregate the model to obtain a consistent structure that can be estimated with aggregate data (Figure 1, part A) and then to use this model to make aggregate predictions. When the underlying choice model is nonlinear and the aggregate groups are not homogeneous, a consistent aggregate function will include parameters of both the choice model and the distribution of independent variables. The requirements on the structure of the choice model and the distribution of variables necessary to develop a consistent estimable aggregate relationship are extremely restrictive. The only successful example of this approach (2) is limited to use with the binary probit choice function and requires the assumption of multivariate normally distributed variables. The estimation is based on information about the distribution of variables in aggregate groups as well as on their mean values. The second approach is to estimate a disaggregate choice model using disaggregate data and then aggregate the estimated choice model when it is used for prediction (Figure 1, part B). The advantage of this approach is that it makes no assumptions about the future distribution of independent variables prior to the model estimation: Assumptions about the distribution of independent variables can be deferred until the time of prediction and may be varied for each prediction situation.

It is conceptually simple to make aggregate predictions of travel behavior when the characteristics of individual trip makers and the alternatives available to them are known or can be predicted. In this case, predictions of the expected choice behavior can be made for each individual and summed or averaged to obtain the aggregate travel predictions (3). Generally, however, this disaggregate information is not available and aggregate predictions must be based on a less than completely detailed set of independent variables.

This paper reports the results of a study of the accuracy of predictions based on disaggregate choice models using less than completely detailed data. It describes the structure of aggregated prediction models, develops a taxonomy of aggregation procedures, describes the error in prediction by alternative aggregation procedures in different prediction situations, and develops tentative conclusions about the performances of the different aggregation methods in different prediction situations. It then proposes a set of guidelines based on these conclusions for prediction with disaggregate models.

STRUCTURE OF AGGREGATED PREDICTION MODELS

Any model that predicts the travel behavior of groups of behavioral units is an aggregate prediction model. Such models predict travel demand at some level of aggregation based on input variables that are also aggregate.

An aggregated prediction model explicitly incorporates disaggregate behavioral relationships in a structure that describes the causal relationship between socioeconomic and travel service characteristics on the one hand and aggregate travel behavior on the other. A model structure for aggregate prediction based on disaggregate travel choice relationships has three components. These are (a) a disaggregate choice model, (b) a representation of the distribution of explanatory variables, and (c) an aggregation procedure that operates on the two other components to obtain the required aggregate prediction. This structure of an aggregated prediction model makes explicit two important advantages of aggregated prediction models over aggregate models based on correlative analysis of aggregate data. These are sensitivity to changes in individual behavior due to changes in travel service or other environmental attributes, including policy control variables, and sensitivity to changes in the distribution of the characteristics of the population that makes up the prediction group.

The disaggregate choice model relates the probability of choosing one alternative out of a set of available alternatives to the relative use of each alternative to the individual decision maker. The utility of an alternative is defined as a function of the characteristics of the individual and the attributes of the alternatives available to him. The choice model may have a wide range of functional forms that are derived from the underlying assumptions about the choice process of the individual (4, 5).

The distribution of independent variables describes the presence in the prediction group of individuals with different socioeconomic characteristics or facing different transportation service attributes. That is, the distribution represents the frequency of occurrence in the prediction group of different values of the socioeconomic and travel service variables that influence individual travel choice decisions. These distributions may be represented in a variety of ways and with different degrees of detail. The various methods for representing the distribution of variables are given below.

1. Enumeration represents the distribution of variables by actual or estimated values of the variables for individuals. Complete enumeration provides variable values for every member of the aggregate prediction group. Partial enumeration provides variable values for a subset of the prediction group.
2. Density functions represent the distribution of variables by the frequency of different variable values in the prediction group. These distributions are based on theoretical or empirical analyses or both, which describe the structure of the distributions and their parameters.
3. Distribution moments represent the distribution of variables in terms of moments and cross moments, which provide information about the spread and shape of individual distributions and their interactions.
4. Classification represents the distribution of variable values in terms of the proportion that is assigned to each of several relatively homogeneous subgroups.

The different methods of representing the distribution of independent variables provide similar information about the actual distribution in different forms. The representations may, to some degree, be transformed to alternative representations. Some transformations imply a loss of information while others imply an increase in information by the use of externally available information or simplifying assumptions. Enumeration, if based on a large enough sample, can be used to estimate parameters of a density function or distribution moment and can be used as a basis for classification. Density functions can be used to determine distribution moments or as a basis for classification. Distribution moments may be used to identify density functions directly or by augmenting them with an assumed distributional form. Classifications, unless they are extremely fine, cannot be used to generate any of the other distributional representations. The process of transformation may be used to modify an initial representation to an alternative representation that is required as input to a selected aggregation procedure.

The aggregation procedure operates on the disaggregate choice model and the distribution of independent variables to produce aggregate predictions. The theoretically consistent aggregation procedure is to estimate the choice probabilities for each individual and then average these choice probabilities to obtain the expected share choosing each alternative. The extreme data requirement of this procedure, prediction of all variable values for every member of the prediction group, motivates a search for aggregation methods that have less extensive input data requirements. A variety of alternative aggregation procedures have been proposed for this purpose. These procedures can be grouped in five categories:

1. Procedures of enumeration are procedures that represent the explicit theoretical relationship between aggregate and disaggregate demand. The expected share choosing an alternative is the average of the individual choice probabilities for that alternative. Complete enumeration averages the choice probabilities for all individuals in the prediction group; sample enumeration averages the choice probabilities for a sample or subset of individuals in the prediction group.
2. Procedures of summation or integration weight conditional disaggregate choice probability estimates by the probability density function for the independent variables. This is done by integration or summation over the multivariate distribution of explanatory variables in the prediction group. The computational requirements of procedures in this group are high if the integration or summation must be applied over a large number of variables. This group of procedures may use distributions (density functions) determined from theoretical and empirical analyses, or assumed distributions that offer computational or other advantages. The most advantageous distributional assumption is that the variables are multivariate normally distributed (2, 6).
3. Procedures of statistical differentials express aggregate shares as a function of the moments of the utility distribution. The aggregate function is obtained by lin-

Figure 1. Alternative procedures to obtain aggregate predictions based on disaggregate choice models.

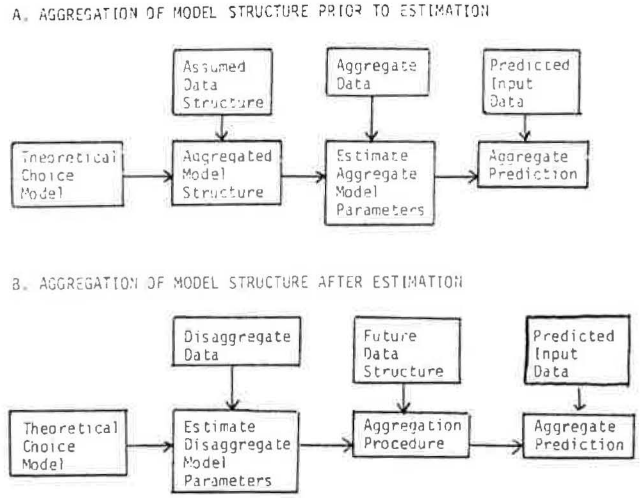
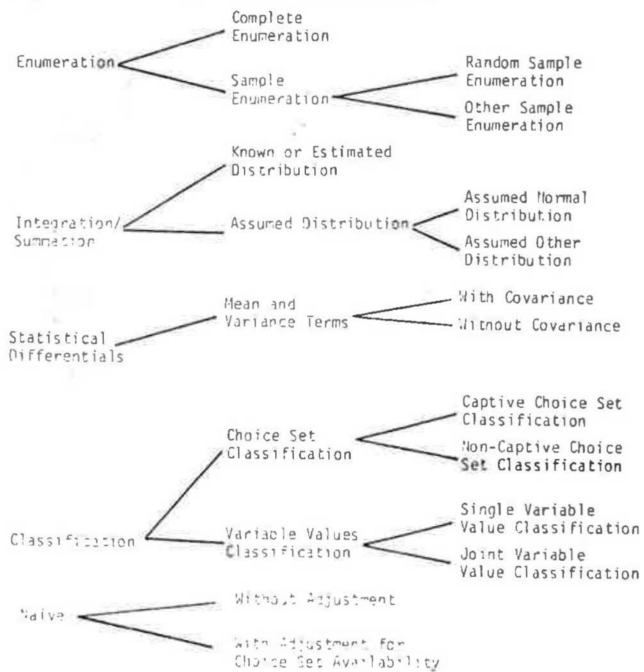


Figure 2. Taxonomy of aggregation procedures.



earizing the disaggregate choice function by the use of a Taylor series expansion and then taking expectation across the aggregate prediction group (7). If the utility function of the alternatives is a linear function of the independent variables, the share estimate is given by a series of terms that includes the nth order derivative multiplied by the nth distributional moments and divided by n factorial. The practical issues associated with estimating higher order moments and the instability of the series when the distribution is highly dispersed (8) suggest that the series be terminated after the second, or variance, term.

4. Procedures of classification assign members of the aggregate group to identifiable classes, use the average variable values for each class to predict aggregate choice shares for each class, and compute overall aggregate share as the weighted average of the class shares. Procedures in this group are differentiated by the basis of classification and the number of classes used. Alternative possibilities include classification by differences in choice set and classification by differences in variable values.

5. Naive procedures use the mean value of the choice influencing variables in the disaggregate choice function. These procedures implicitly assume that each individual acts as if he is described by the average values of the prediction group. The naive procedure is a special case of summation or integration procedures (the distribution is assumed to be concentrated at a point), statistical differentials procedures (truncating the series after the first term), or classification procedures (using one class only). It is useful to treat this procedure separately because (a) the data requirements are the same as those for models calibrated with aggregate data, (b) it is computationally and conceptually simple, and (c) it is the method most likely to be used in the absence of recognition of the aggregation problem. Predictions by naive procedures can be adjusted to account for differences in choice set availability when such differences exist.

The different types of aggregation procedures may be described in terms of the taxonomy represented in Figure 2. The major classes of aggregation procedures are divided into differentiable subgroups of procedures. These subgroups are defined by their important characteristics.

INFLUENCE OF PREDICTION ENVIRONMENT ON CHOICE OF AGGREGATION PROCEDURE

Choice of an aggregation procedure for use in a specific prediction environment depends on the expected magnitude of the aggregation error of alternative procedures and the contribution of the aggregation error to the overall error in prediction from all sources. The aggregation error of alternative procedures depends on the prediction situation, particularly the distribution of explanatory variables in the prediction group, the level of aggregation, the differences in choice set availability among group members, and the choice probabilities for the average member of the prediction group (9). The contribution of the aggregation error to the total error in prediction from all sources depends on the magnitude of the aggregation error relative to the magnitude of error from other sources. The prediction error from other sources is determined by decisions made in the model formulation and prediction process (3). These decisions include the specification of the disaggregate choice model including the functional form and variables to be included, the selection of data to be used in model estimation, the method selected to represent and predict the distribution of explanatory variables, and the selection of the aggregation procedures to be used. The characteristics of the prediction situation and the decisions made in the model development and prediction process influence both the aggregation error and the errors from other sources.

ANALYSIS OF ERRORS IN PREDICTION WITH DISAGGREGATE CHOICE MODELS

Errors in prediction with disaggregate choice models arise from each of the components of the aggregated model structure. For the purpose of the following discussion, these errors in prediction are separated into two categories:

1. Model and variable error, which includes errors in the specification of the choice model, errors in parameter estimation, and errors in the input variables; and
2. Aggregation error, which includes errors due

to the use of approximate aggregation procedures.

Empirical analysis of these errors was based on prediction of mode shares to work in the Washington metropolitan area for the drive alone, shared ride, and transit modes. The predictions were made for aggregate groups identified as 45 districts with an average sample size of 47, 10 super-districts with an average sample size of 213, and four rings with an average sample size of 533. The disaggregate mode choice model predicts the probability of drive alone, shared ride, and transit choices for the first work trip of breadwinners working in the CBD. The overall mode shares in the prediction sample were 38, 30, and 32 percent for the drive alone, shared ride, and transit alternatives respectively. The model specification and parameter estimates based on 824 work trips, of which 621 had all choices available and 253 did not have the drive alone alternative available due to lack of a car or driver's license, are given below.

Variable	Symbol	Estimated Coefficient	Standard Error
Drive alone dummy	D_d	-2.62	0.36
Shared ride dummy	D_s	-2.36	0.27
Automobiles per licensed driver (drive alone)	$AALD_d$	3.64	0.38
Automobiles per licensed driver (shared ride)	$AALD_s$	1.51	0.24
Out-of-vehicle cost per income	OPTC/INC	-0.028	0.012
Total travel time	TTT	-0.024	0.005
Out-of-vehicle time per distance	OVTT/DIST	-0.077	0.055
Government worker (shared ride)	GW_s	0.77	0.16
Number of workers in household (shared ride)	$NWORK_s$	0.24	0.10

The aggregation procedures used include the naive procedure, with adjustment for choice set availability, the statistical differentials procedure, with mean and variance terms, and classification by differences in choice set availability and automobile ownership.

Model and variable error in prediction was estimated by comparison of share predictions by the enumeration procedure against observed choice shares in the data set. Aggregation error was determined by comparison of share predictions by a selected aggregation procedure against share predictions by the enumeration procedure. Both types of error are reported as a percentage of the magnitude of prediction. Prediction errors were obtained for each prediction group (districts, super-districts, or rings) and each choice share. These errors are summarized in terms of average error, standard deviation of the error, and root mean square error (3). The combined error (model and variable error and aggregation error) is the square root of the squared model and variable error and the squared aggregation error. (This formulation implies independence between the errors from these two sources. This is a reasonable assumption in the absence of structural interdependence.) It is most sensitive to changes in the magnitude of error from the source that contributes the larger error.

COMPARISON OF AGGREGATION PROCEDURES

The naive procedure produced aggregation errors of approximately 10 percent of predicted values. Adjustment of the naive procedure by choice set availability reduced aggregation error to about 8.5 percent. Classification by choice set and automobile availability reduced the aggregation error to about 3 percent. The statistical differentials procedure resulted in higher aggregation error than the naive procedure. [Under certain conditions the statistical differentials produced can be expected to increase rather than decrease the aggregation error (9).] The errors are as follows:

Prediction Group	Error of Different Aggregation Procedures (%) Naive	Naive With Adjustment	Statistical Differentials	Classification by Choice Set
Districts (45)	10.5	8.1	12.9	3.3
Super-districts (10)	9.8	8.5	11.8	2.9
Rings (4)	9.6	8.4	12.2	2.8

These results suggest the following conclusions:

1. When differences in choice set availability exist, these differences should be used as a basis for adjusting predictions by the naive procedure or as a basis for classification.
2. The statistical differentials procedure should be used only after verifying that it actually reduces aggregation error in prediction.

LEVEL OF AGGREGATION

Increasing levels of aggregation are expected, a priori, to have two effects on prediction error. First, model and variable error is expected to decline. This reduction of error results from averaging the expected choice probability over the larger sample of observations in the more aggregate prediction group. If the sample predictions are independent, the expected error in the estimate of choice shares would be inversely proportional to the square root of the number of observations in the prediction group. Predictions made with a common disaggregate choice model are not independent (3). Therefore, the effect of increasing the prediction group size should be less than proportional to the prediction group size. Second, aggregation error is expected to be larger for prediction groups with greater geographical dispersion as these groups are also expected to have greater dispersion of explanatory variable values.

The first expectation is supported by the results obtained. The model and variable error for each mode, and for all modes combined, declined with increasing level of aggregation.

Prediction Group	Model and Variable Error (%) Drive Alone	Shared Ride	Transit Ride	All Modes
All districts (45)	19.8	35.4	28.7	27.8
Super-districts (10)	12.3	24.9	21.6	19.7
Rings (4)	5.4	18.0	20.1	15.4

This decline in errors was less than proportional to the square root of prediction group size. On the other hand, the second expectation is not supported by the results, as the aggregation error for the three different aggregation procedures was not strongly or consistently related to the level of aggregation. This is a consequence of the fact that increasing levels of aggregation do not lead to significant increases in intragroup variation in socioeconomic characteristics (10) or in intragroup differences in mode service characteristics. The net effect of these results, however, is that the combined error in prediction declines with increasing levels of aggregation as given below. As the size of the prediction group increases, the portion of the combined error attributable to aggregation error increases and the differences in prediction errors between aggregation procedures are amplified.

Prediction Group	Combined Error of Different Aggregation Procedures (%) Naive	Naive With Adjustment	Statistical Differentials	Classification	Enumeration
Districts (45)	29.7	29.0	30.7	28.0	27.8
Super-districts (10)	22.0	21.4	23.0	19.9	19.7
Rings (4)	18.1	17.5	19.7	15.7	15.4

These results lead to the following conclusions:

3. Predictions made with disaggregate choice models based on a fixed data set increase in accuracy as the level of aggregation increases.
4. The relative improvement in prediction error from the use of more precise aggregation procedures increases as the level of aggregation increases.

DIFFERENCES IN CLASSIFICATION STRUCTURE

The classification procedure used classifies the population according to both choice set availability and intrahousehold automobile availability. This classification was compared to classification solely by choice set availability and classification solely by intrahousehold automobile availability to identify the influence of different classification schemes on aggregation error. The aggregation errors for all three classification procedures at three levels of aggregation are given below.

Prediction Group	Aggregation Error of Different Classification Procedures (%) Choice Set and Automobile Availability	Automobile Availability Alone	Choice Set Alone
Districts	3.3	9.9	5.2
Super-districts	2.9	8.4	5.2
Rings	2.8	7.8	5.3

The most complete classification, that by choice set and automobile availability, has the least aggregation error. Classification by choice set alone has substantially lower aggregation error than classification by automobile availability alone. These results lead to two further conclusions:

5. Increasing refinement in classification leads to reduced aggregation error.
6. Classification by differences in choice set availability, when appropriate, results in lower levels of aggregation error than classification by variable values.

CHOICE MODEL SPECIFICATION

The large magnitude of model and variable error relative to aggregation error indicates the need to improve the choice model specification. The model and variable errors reported for the different levels of aggregation are substantially higher for the prediction of shared ride and transit shares than for the prediction of drive alone mode shares. The sources of these errors can be investigated by disaggregating them into average error and standard deviation of error for each mode as given below. The large errors in the prediction of shared ride and transit ride shares compared to drive alone shares are due to both higher average errors and greater standard deviation of errors. The opposite signs of the average error for shared ride and transit and the higher magnitude of average error and variability indicate that the model is deficient in predicting shares between these two alternatives. This deficiency may be due to inherent difficulties in specifying the utility function for these alternatives (due to large variations in excluded characteristics such as comfort and convenience or greater heterogeneity of preferences for different types of group ride alternatives). It may also be due to too great reliance on the independence of irrelevant alternatives axiom to estimate model parameters for individuals with different choice sets. The errors are disaggregated as follows:

Error Type	Model and Variable Error Disaggregated by Error Type Drive Alone	Shared Ride	Transit Ride	All Modes
Average error	+1.6	-7.1	+4.1	4.5
Standard deviation of error	19.7	35.3	28.4	27.3
Root mean square error	19.8	35.4	28.7	27.8

These results suggest two additional conclusions:

7. Analysis of error in prediction provides a basis for reevaluation and modification of the disaggregate choice model.
8. Individuals with structurally different choice sets should not be combined for estimation without testing the effect of this grouping on estimation and prediction errors.

PREDICTING CHANGED CHOICE SHARES

The discussion to this point has been concerned with the prediction of existing choice shares. We now turn our attention to prediction of choice shares after a change in travel service characteristics. The changes considered are (a) provide shared ride incentives to all trip makers (change 1), (b) reduce transit fares to zero (change 2), and (c) reduce transit times by one-half (change 3). The expected effect of these changes on mode share for the entire data sample is given below.

Prediction Situation	Mode Choice Shares Drive Alone	Shared Ride	Transit Ride
Change 1	0.35	0.36	0.30
Change 2	0.37	0.26	0.37
Change 3	0.35	0.23	0.42
Base	0.39	0.28	0.33

The new mode shares may be predicted directly by modifying the variables to reflect the policy changes. Alternatively, the new mode shares may be obtained by predicting the incremental change in shares resulting from a change in policy and using the predicted change to modify the observed choice shares. The aggregation error by the incremental prediction procedure is substantially lower than that by the direct procedure for all aggregation methods and for all of the policy changes. This result suggests the following conclusion:

9. Incremental prediction should be used to predict aggregate choice shares after policy change whenever predictions can be made for an observed set of choice shares to provide a base for adjustment.

SUMMARY

The preceding discussion describes a framework for the use of disaggregate choice models for the prediction of aggregate choice shares and proposes a taxonomy of procedures for making aggregate predictions based on disaggregate choice models. The aggregation error and the effect on the combined error of different aggregation procedures are empirically estimated for a variety of prediction situations. The results are summarized in a set of conclusions about the expected performances of selected aggregation procedures in specific situations. These conclusions suggest the following guidelines for aggregate prediction using disaggregate choice models.

1. Disaggregate choice models may be most effectively used for prediction at high levels of aggregation appropriate to policy analysis.
2. Enumeration procedures should be used whenever adequate sample data are available, especially at high levels of aggregation.
3. When sample data are not available, classification procedures used should be based on the most important class distribution (which will be differences in choice set when such differences exist).
4. When data are not available to predict class specific variable values, predictions by the naive procedure should be adjusted for differences in choice set when such differences exist.
5. The specification of the underlying disaggregate choice model should be developed and evaluated with particular care in the grouping of individuals having structurally different choice sets.
6. Incremental prediction should be used for prediction of the expected impacts of policy changes whenever an existing set of choice shares is available to use as a basis for adjustment.

ACKNOWLEDGMENTS

This research was supported by the Office of the Secretary, U.S. Department of Transportation. The empirical data used were provided by the Metropolitan Washington Council of Governments.

I wish to acknowledge the helpful comments and suggestions of Moshe Ben-Akiva, Marvin Manheim, Paul Roberts, and Steven Lerman of Massachusetts Institute of Technology, Daniel Brand of Harvard University, and Gerald Kraft of Charles River Associates.

REFERENCES

1. H. A. J. Green. Aggregation in Economic Analysis. Princeton Univ. Press, Princeton, N.J., 1964.
2. D. McFadden and F. Reid. Aggregate Travel Demand Forecasting From Disaggregated Behavioral Models. TRB, Transportation Research Record 534, 1975, pp. 24-37.
3. F. S. Koppelman. Prediction With Disaggregate Models: The Aggregation Issue. TRB, Transportation Research Record 527, 1975, pp. 73-80.
4. Charles River Associates. A Disaggregated Behavioral Model of Urban Travel Demand. Federal Highway Administration, U.S. Department of Transportation, 1972.
5. M. Ben-Akiva. Structure of Passenger Travel Demand Models. Department of Civil Engineering, M.I.T., Cambridge, PhD thesis, June 1973.
6. R. B. Westin. Prediction From Binary Choice Models. Journal of Econometrics, April 1974.
7. A. Talvitie. Aggregate Travel Demand Analysis With Disaggregate Travel Demand Models. Proc., Transportation Research Forum, 1973.
8. N. L. Johnson and S. Kotz. Distributions in Statistics, Vol. 1, Discrete Distributions. John Wiley, New York, 1969.
9. F. S. Koppelman. Travel Prediction With Models of Individual Choice Behavior. M.I.T. Center for Transportation Studies, Rept. No. 75-7, June 1975.
10. C. R. Fleet and S. R. Robertson. Trip Generation in the Transportation Planning Process. HRB, Highway Research Record 240, 1968, pp. 11-31.