Aggregate forecasting with dissaggregate travel demand models using normally available data

by

FRANK S. KOPPELMAN
Northwestern University Evanston, Illinois

and

MOSHE E. BEN-AKIVA
Massachusetts Institute of Technology Cambridge, Massachusetts

INTRODUCTION

Disaggregate travel choice models have been extensively developed in recent years (CRA, 1972; PMM, 1973; Ben-Akiva, 1973; Richards and Ben-Akiva, 1974; Lerman and Ben-Akiva, 1975; and others). Their development has contributed to the refinement of theories of travel behavior through improved identification of explanatory variables and their relative influence on travel choice behavior. This paper is concerned with the application of disaggregate models to obtain predictions of aggregate travel flows required for transportation systems analysis and planning. In this context prediction is the final step in a sequence of activities which include model development, estimation and prediction (Figure 1). The flow diagram in Figure 1 calls for the use of a prediction procedure which obtains the desired aggregate predictions based on predicted input data and the estimated disaggregate model. The objective of this paper is to identify prediction procedures for use with disaggregate models which (1) use commonly available data or additional data which can be obtained easily, (2) are computationally inexpensive, and (3) provide relatively accurate travel predictions.

Behavioral travel demand theory is postulated at the level ot the decisionmaking or behavioral unit, usually an individual or household. Disaggregate travel choice models are estimated at this level of analysis. The resultant disaggregate model takes the form:

$$P_t = f(X_t, \theta) \tag{1}$$

where P_t is the probability that individual t selects a given alternative,

X_t is a vector of independent variables for individual t,

θ is a vector of model parameters, and

$f(\ )$ denotes the functional form of the demand relationship.

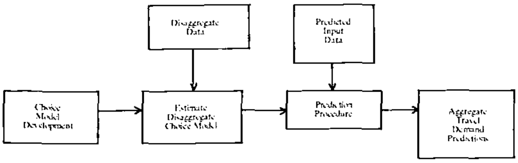

Figure 1 - Model development, estimation and prediction

Aggregate travel demand is the sum of the travel choices of numerous behavioral units making independent decisions. Thus, the aggregate share of travelers choosing an alternative is obtained by averaging the disaggregate choice probabilities over the relevant population so that

$$\begin{aligned}\bar{D} &= \frac{1}{T}\sum_{t=1}^{T} P_t \\ &= \frac{1}{T}\sum_{t=1}^{T} f(X_t, \theta)\end{aligned} \tag{2}$$

where $\bar{D}$ is the share of the population choosing the alternative, and

T is the number of individuals in the population.

This formulation implies that the independent variables for each individual, X_t, must be known or predicted for every member of the population in order to predict aggregate demand. The difficulty of satisfying this requirement motivates the search for alternative aggregation procedures.

An alternative situation exists when the distribution of independent variables is known for the population. The joint distribution of independent variables can be expressed as

$$h(X, \bar{X}, \alpha) \tag{3}$$

where $\bar{X}$ is the vector of mean values for the independent variables, and

α is a vector of other parameters describing the distribution of X.

In this case, aggregate demand is obtained by integrating the individual demand function over the distribution of independent variables so that:

$$\bar{D} = \int_X f(X, \theta)\ h(X, \bar{X}, \alpha)\ dx \tag{4}$$

which can be expressed as an aggregate demand function by [1]

$$\bar{D} = F(\bar{X}, \theta, \alpha) \tag{5}$$

This formulation implies that aggregate demand is determined by the distribution of the independent variables as well as the coefficients of the disaggregate demand function.

This definition of aggregate demand shows that, in general, substitution of the mean values of the independent variables in the disaggregate demand function will not equal the aggregate demand function. That is,

$$\bar{D} \neq f(\bar{X}, \theta) \tag{6}$$

since, in general,

$$F(\bar{X}, \theta, \alpha) \neq f(\bar{X}, \theta) \tag{7}$$

The magnitude of the error resulting from substitution of aggregate values in the disaggregate model depends on both the form of the demand function and the shape of the distribution of independent variables. There are two situations for which the substitution does not result in any aggregation error. These are:

- when the aggregate group is homegeneous with respect to the values of the independent variables - that is, every individual has identical values of the variables; or,
- when the disaggregate demand function is linear in the independent variables. [2]

In these two situations only:

$$\bar{D} = f(\bar{X}, \theta) \tag{8}$$

and only information on mean values is required for prediction. However, these two situations are rare in travel demand applications. Thus, in general, the use of average population variable values in the disaggregate choice model (as is normally done in the use of conventional aggregate models) will produce aggregate errors. [3]

Complete elimination of this aggregation error requires detailed information on the distribution of individual variables. Such information is generally either not available or difficult and expensive to obtain. Thus, the basic problem in aggregate prediction is to find a procedure which will reduce this aggregation error using readily available or easily obtained information on the distribution of independent variables.

ALTERNATIVE AGGREGATION PROCEDURES

We can identify a variety of aggregation procedures with different information requirements, different levels of computational complexity and different levels of expected aggregation error. Koppelman (1976a) proposed a taxonomy of aggregation procedures depicted in Figure 2 and described below.

Procedures of enumeration are based on the theoretical relationship between aggregate and disaggregate demand defined in equation 2. Aggregate demand is calculated by averaging individual demands. Complete enumeration is accomplished by averaging individual demand predictions for the entire population and requires values of independent variables for the entire population. Sample enumeration estimates aggregate demand by averaging predictions for a sample of the population only. Sample enumeration requires data only for the sample used, but introduces random variations due to the nature of the sampling process.

Procedures of summation/integration weight the disaggregate demand estimates by the probability density function for the indepentent variables. This is done by integration when the density function is continuous (equation 4), or by summation when the distribution is discrete. The aggregation error of these procedures depends primarily on the accuracy of the distributional representation and secondarily on the computational procedure used. These procedures may be differentiated by the use of estimated or assumed distributions.

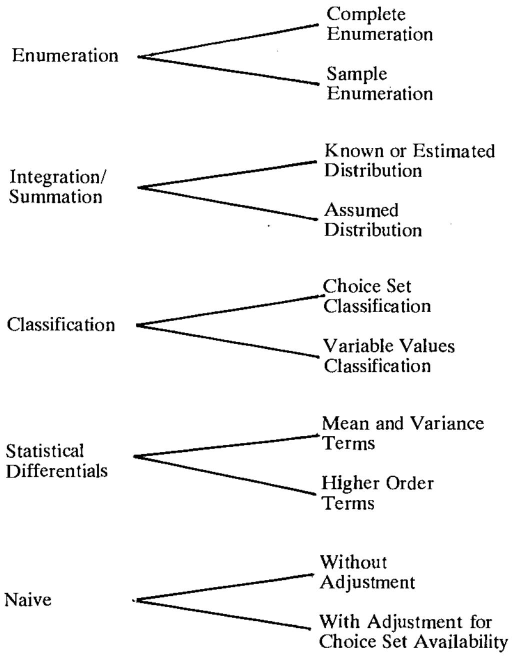

Figure 2 - Taxonomy of aggregation procedures

Procedures of statistical differentials express aggregate demand as a function of the moments of the distribution of independent variables. The aggregate function is obtained by linearizing the disaggregate function by a Taylor series expansion about the mean values and taking expectations (Talvitie, 1973). In practice, the resultant series is always truncated so that aggregate demand is expressed in terms of the means, variances and covariances of the distribution of independent variables.

Procedures of classification assign members of the population to relatively homegeneous groups, use group-average variable values to predict demand for each group using the naive procedure (described below), and compute aggregate demand as the weighted average of group demands. Classification procedures are differentiated by the basis for classification and the number of classes used. An important classification basis in travel choice models is the set of available alternatives (or choice set), which may differ among population subgroups.

The naive procedure uses the mean value of the independent variables in the disaggregate demand function to predict aggregate demand. The naive procedure is a special case of summation/integration procedures (when the distribution is degenerate), statistical differentials procedures (when the series is truncated after the first term) and classification procedures (when only one class is used). It is useful to consider this procedure separately for four reasons. First, the data requirements are the same as those for conventional aggregate models. Second, it is computationally and conceptually simple. Third, it is the method most likely to be used when the aggregation problem (equation 6) is not recognized. Fourth, it establishes a benchmark against which the aggregation error of other procedures may be compared.

The naive procedure can be modified by information on choice set availability when such information exists.

It is useful to identify a smaller set of aggregation procedures which are most likely to satisfy the dual objectives of small aggregation error and minimal requirements for additional information. Sample enumeration is particularly useful for short-range analysis when a prior sample exists. In this procedure, a sample of the population is used for prediction after modifying the variables for each observation to reflect the policy being analyzed (reduce transit fare by X%, impose a parking surcharge of Y dollars, etc.). This procedure has been used in recent studies by CSI (1976), Ben-Akiva and Atherton (1977), and Small (1977). The advantage of this procedure is that it accurately reflects individual or household characteristics when the prediction timeframe is sufficiently short to assume that households remain unchanged except for the effects of the policy imposed.

The sample enumeration is approximated for long-range prediction or when a prior sample does not exist by a Monte Carlo integration procedure which we refer to as the *pseudo-sample enumeration procedure*. In this procedure a pseudo sample is synthesized by taking random drawings from the expected future distribution of independent variables. The development of pseudo samples can be simplified by assuming that the distribution of variables are independent and/or that some variables have no distribution. Procedures for generating a pseudo sample of households using census data or other readily available data have been developed by Duguay *et al.* (1976), and Chatterjee *et al.* (1977), and applied to an area-wide analysis of urban travel by Koppelman *et. al.* (1976).

Naive and classification procedures are suitable for many applications because they require information which is commonly available in existing studies and do not require changes in the basis structure of conventional travel forecasting software. If only the means of the variables are available, the naive procedure must be used. When classifications are available for important variables such as household size, income or automobile availability, these classifications can be used. Classification by the availability of different alternatives, which may be determined by access distance to public transportation alternatives or vehicle ownership for private alternatives, is very effective in reducing aggregation error. Information about the portion of the population which has access to different alternatives also can be used to adjust naive predictions as described in the next section.

Classification using values of level-of-service variables (distance to a transit station) was used in a Dual Mode case study (CSI, 1975). An urban transportation model system developed for the San Francisco Metropolitan Transportation Commission uses a classification scheme based on number of workers in the household, income and automobile ownership (Ruiter and Ben-Akiva, 1977).

Statistical differentials procedures may be used to improve on naive predictions when the spread or variance of the distribution is relatively small. However, when the distribution is wide, statistical differentials with mean and variance terms often produces aggregation errors greater than those of the naive procedure.

Thus, the aggregation procedures most appropriate for general application should be selected from (1) sample enumeration, (2) pseudo-sample enumeration or Monte Carlo integration, (3) classification by alternative availability or variable values, and (4) the naive procedure with or without adjustment for choice-set availability. Each of these procedures provides aggregate predictions based on information which either is commonly available in current forecasting studies or can be developed with reasonable effort.

USE OF CHOICE SET INFORMATION

The choice set adjusted naive procedure modifies prediction by the naive procedure to account for differences in alternative availability. Different adjustment procedures can be used depending on the available information. When the analyst knows the portion of the population which has each set of choices available, the share choosing each alternative within each choice set group is:

$$S_{ig} = \frac{S_i}{\sum_{j \epsilon A_g} S_j} \quad \text{if } i \epsilon A_g \text{ (if alternative } i \text{ is available to group } g\text{),} \tag{9a}$$

$$S_{ig} = 0 \quad \text{otherwise.} \tag{9b}$$

where S_{ig} is the share of group g predicged to choose alternative i,

S_i is the naive prediction for choice share i, and

A_g is the set of alternatives available to group g.

That is, the choice shares for alternatives which are not available are set to zero and the choice shares for the remaining alternatives are increased proportionately. The population choice shares are the weighted (by size of choice set group) average of choice shares for each group:

$$S_i{*} = \sum_g R_g \; S_{ig}, \tag{10}$$

where $S_i{*}$ is the adjusted share prediction, and

R_g is the proportion of the population which has choice set g available.

Consider, for example, a share prediction for three alternatives, two of which are available to the entire population. If R_1 is the proportion of the population which has all three alternatives available and S_1, S_2 and S_3 are the naive predictions, the adjusted shares are:

$$S_1{*} = R_1 S_1 \tag{11a}$$

$$S_2{*} = R_1 S_2 + (1-R_1) \frac{S_2}{S_2+S_3} = S_2\left[\frac{1-R_1S_1}{1-S_1}\right] \tag{11b}$$

$$S_3{*} = R_1 S_3 + (1-R_1) \frac{S_3}{S_2+S_3} = S_3\left[\frac{1-R_1S_1}{1-S_1}\right] \tag{11c}$$

A different situation exists when the analyst knows the portion of the population which has each alternative available but does not know how availability of different alternatives is related. For example, the analyst may know that twenty percent of the population does not have an automobile available and sixty percent does not have commuter rail available, but he might not know which, if any, part of the population does not have either alternative available. In this case the naive prediction is adjusted by:

$$S_i{*} = S_i R_i \left[\frac{1-S_i}{1-S_i R_i}\right] \cdot \frac{1}{K} \tag{12a}$$

$$K = \sum_j \left[\frac{1-S_j}{1-S_j R_j}\right] S_j R_j \tag{12b}$$

where R_i is the proportion of the population with alternative i available, and

K is a normalization factor which assures that the sum of shares equals one.

When only one alternative has partial availability, as in the above example, both adjustment procedures produce identical results.

Classification assigns the population to groups according to the actual choice set available or according to the values of selected variables (the selection of the method of classification in discussed later). Average values of the explanatory variables are determined for each of the population groups. The naive procedure is used to predict choice shares for each group. The overall estimate of choice shares is obtained by weighting the choice shares for each group according to its size:

$$S_i = \sum_g R_g S_{ig}. \quad (13)$$

where S_i is the predicted share,

R_g is the size of group g, and

S_{ig} is the predicted share for group g.

Classification procedures can be simplified by using population averages for some variables and group average values for other variables based on the differences which they contribute to the choice utility estimates. When only population average variables are used, this procedure is equivalent to the choice set adjusted naive procedure.

EMPIRICAL STUDIES OF AGGREGATION ERROR

Aggregation procedures, other than complete enumeration, introduce aggregation error to aggregate demand predictions. The magnitude of aggregation error depends on the prediction situation. However, Koppelman (1975) has shown that the relative performance of different aggregation procedures is relatively stable over a wide range of prediction situations. The purpose of this section is to describe the results of some empirical analyses of aggregation error. The results of these analyses are indicative of the performance of these aggregation procedures in different situations.

The analysis is based on mode share predictions for the work trip to the Central Business District (CBD) from residence districts in the Washington, D.C. metropolitan area. [4] A disaggregate logit choice model was estimated to predict choice probabilities for drive alone, shared ride and transit. The drive-alone alternative is available only to individuals who have a driver's license and reside in a household which has one or more automobiles. The other alternatives are available to all individuals. The choice model includes variables which describe the household (in terms of income, number of automobiles and number of workers), the level of service by different modes (in terms of travel time and cost), and the existence of incentives (special parking privileges, etc.) to use the shared-ride alternative.

Aggregation errors are determined by comparing the predictions by a selected aggregation procedure to corresponding predictions by the sample enumeration procedure. The aggregation procedures considered are:

- the naive procedure
- the naive procedure with adjustment for choice set availability,
- classification by choice set *and* automobile availability,
- classification by choice set availability *only,* and
- classification by automobile availability *only.*

The aggregation error for each prediction (each mode for each district) is expressed relative to the magnitude of the prediction by

$$E_{md} = \frac{\hat{N}_{md} - N_{md}}{\hat{N}_{md}} \quad 14)$$

where E_{md} is the relative aggregation error for the prediction of mode m in district d,

N_{md} is the observed number choosing mode m in district d, and

$\hat{N}_{md}$ is the predicted number choosing mode m in district d.

The aggregation errors for each prediction are summarized first by mode in terms of average error, standard deviation of error and root-mean-square error, all weighted by the magnitude of prediction, as follows:

$$AE_m = \sum_d E_{md} \cdot \frac{\hat{N}_{md}}{N_m} \quad (15a)$$

$$SDE_m = \left[\sum_d (E_{md} - AE_m)^2 \cdot \frac{\hat{N}_{md}}{N_m}\right]^{\frac{1}{2}} \quad (15b)$$

$$RMSE_m = \left[\sum_d E_{md}^2 \cdot \frac{\hat{N}_{md}}{N_m}\right]^{\frac{1}{2}} \quad (15c)$$

$$N_m = \sum_d \hat{N}_{md} \quad (15d)$$

where AE_m is the average error for mode m,

SDE_m is the standard deviation of error for mode m, and

$RMSE_m$ is the root-mean-square error for mode m.

These error measures can be further summarized across modes to obtain overall measures of error by:

$$AE = \left[\sum_m \frac{N_m}{N} (AE_m)^2\right]^{\frac{1}{2}} \quad (16a)$$

$$SDE = \left[\sum_m \frac{N_m}{N} (SDE_m)^2\right]^{\frac{1}{2}} \quad (16b)$$

$$RMSE = \left[\sum_m \frac{N_m}{N} (RMSE_m)^2\right]^{\frac{1}{2}} \quad (16c)$$

$$N = \sum_m N_m \quad (16d)$$

The average, standard deviation and root-mean-square error terms by mode or overall are related by

$$RMSE^2 = AE^2 + SDE^2. \quad (17)$$

That is, the root-mean-square error, which is a total error measure, can be disaggregated into average and variational components. [5]

This summarization makes it possible to represent the aggregation error in multiple predictions with a relatively small amount of information. The root-mean-square error for all modes provides a single overall index of aggregation error. Disaggregation into average and standard deviation error or by modes or both provides more detailed information on the types of error and their source. This summary of aggregation error for the naive procedure is given in Table 1. The overall expected aggregation error is 10.5 percent of predicted values. The portion of this error associated with average error is 6.2 percent, and that associated with standard deviation around average error is 8.5 percent. The overall error can also be analyzed in terms of the errors associated with the prediction of each of the modal shares. The observed aggregation errors are substantially larger for the transit-ride alternative than for either the drive-alone or shared-ride alternatives.

Table 1 - Aggregation error by naive procedure summarized across 45 districts

	MODE			
	Drive Alone	Shared Ride	Transit Ride	All Modes
Average error	6.0	0.6	–8.7	6.2
Standard Deviation of Error	7.4	6.6	11.1	8.5
Root Mean Square Error	9.5	6.6	14.1	10.5

Similar summaries of aggregation error by the other procedures have also been developed but will not be reported here. Instead, we compare the aggregation error for all the procedures in terms of the average, standard deviation and root-mean-sqaure error summarized across modes (Table 2). Table 2 indicated the degree to which aggregation error by the naive procedure is reduced by the use of additional information.

The use of alternative information in the naive procedure with adjustment reduces aggregation error by almost 25 percent. This reduction is due to a substantial decrease in the standard deviation of error and a small increase in the average aggregation error. Classification by choice set alone or choice set and automobile avialability alone results in a nominal reduction in aggregation error. The rank order of aggregation error for each procedure is given in Table 3. The corresponding subjective ranking of degree of information and effort required to use each procedure is also given in Table 3. This information provides the basis for selecting procedures which use increasing information to reduce aggregation error. The additional effort of classification using choice set availability compared to adjustment of the naive procedure results from the need to obtain class-specific variable values which are not required for the naive procedure with adjustment.

The rankings for magnitude of aggregation error and information and effort required are generally in reverse order. However, the naive procedure with adjustment for choice set availability is preferable to classification based on automobile availability with respect to both criteria. Thus, we obtain a reduced set of procedures for use when sample enumeration is not feasible. These are (1) the naive procedure, (2) naive procedure with choice set adjustment, (3) classification by choice set availability, and (4) classification by choice set and automobile availability in increasing order of effort and decreasing order of aggregation error.

Table 2 - Aggregation error for five procedures summarized across 45 districts and three modes

	PROCEDURES				
	Naive		Classification		
Error Measure	Without Adjustment	With Adjustment	Choice Set & Auto Avail.	Choice Set	Auto Avail.
Average Error	6.2	6.5	1.2	2.0	7.3
Standard Deviation of Error	8.5	4.8	3.0	4.8	6.6
Root Mean Square Error	10.5	8.1	3.3	5.2	9.9

Table 3 - Comparison of aggregation procedures

	PROCEDURES				
	Naive		Classification		
	Without Adjustment	With Adjustment	Choice Set & Auto Avail.	Choice Set	Auto Avail.
Rank by Minimum Aggregation Error	5	3	1	2	4
Rank by Minimum Information and Effort Required	1	2	5	4	3

This analysis of aggregation error is based on a single set of mode share predictions for 45 groups of individuals (each group living in a common residential district). The representativeness of these results was tested by analyzing the aggregation errors, first for three geographically-defined sets of districts, and second for two different policy changes. In each case the general magnitudes of aggregation error for each procedure were similar and the ordering of different aggregation procedures was unchanged (Koppelman, 1975). Thus, these aggregation errors should be indicative of the errors which might be expected in a variety of situations.

It is useful to compare the aggregation errors to other sources of error in prediction. Using the same prediction situation, errors from other sources were estimated to be 27.8% of the magnitude of prediction. (Note: Errors from other sources are estimated by comparing predictions by the enumeration procedure against the observed shares in the data). Thus, the aggregation error, even by the naive procedure, is small compared to error from other sources.

The empirical analyses indicate that the magnitude of aggregation error by the naive procedure is relatively small compared to errors from other sources. Reduction in these aggregation errors can be obtained by use of alternative procedures based on information which is commonly available or which can be generated with moderate effort. The naive procedure with adjustment, which requires information on availability of alternatives only, reduces aggregation error by about 25%. Classification procedures based on choice set availability, which requires information on the availability of alternatives and class-specific average variable values, reduce aggregation error by more than 50%.

These results indicate the importance of considering differences in choice set availability when they exist.[6] This information can be used either to adjust predictions initially made by the naive procedure or as a basis for classification. Consideration of modal availability will obtain substantial reductions in prediction error. When differences in choice set availability do not exist but wide differences in explanatory variables occur, classification by variable values or pseudo-sample procedures should be used.

PREDICTING CLASSIFICATION GROUPS AND ALTERNATIVE AVAILABILITY

The classification procedures and the naive procedure with adjustment for alternative availability require information about the size of different prediction groups and, for classification procedures, class-specific average variable values. This information is not available in many prediction situations, but often can be generated with a moderate amount of effort. This section describes how this information can be developed.

The criterion for selecting a classification structure is to identify a classification which will obtain significant reduction in aggregation error with a minimal increase in required input data and computational effort. This criterion can best be satisfied by classifying the population according to differences in choice set availability when such differences exist. That is, when different individuals make choices from different sets of alternatives, differences in the choice set provide a powerful basis for classification. Choice set classifications exist for many choice situations. These include mode choice limited by automobile availability or access to public transit, automobile ownership limited by household income and number of drivers, shopping location limited by travel time, etc. Classifications may also be based on differences in the value which members of the population have for selected independent variables. The maximum reduction in aggregation error is obtained by classifying according to the variable which contributes the largest variance to the dependent variable. Recent mode choice studies identify automobile availability, the number of automobiles per licensed driver, as the most important variable for classification. The most powerful classifier for an automobile ownership model is expected to be household income. Powerful classifiers for other models can be obtained by examination of the data for prediction groups and the coefficients of estimated choice models.

Alternative availability information is required for both the naive procedure adjusted for alternative availability and classification by choice set. In the mode choice situation, availability of private alternatives such as drive alone is based on availability in the household of the required vehicle. Availability of public transportation alternatives is based on access to a transit stop.

Prediction of the proportion of households which do not own a vehicle required for a specific mode requires development of a distributional respresentation which relates levels of ownership to the average vehicle ownership of the population. That is, the objective is to transform commonly available information such as the average number of automobiles per household to an estimate of the proportion of households which do not have any automobiles. One method of obtaining the required distribution is to project existing distributions to the future time period. This can be done by grouping districts which have similar average values for vehicle ownership levels as a function of average vehicle owwhich fall in each level of ownership category with particular emphasis on the proportion of households which do not onw any such vehicles. A table of relationships between average values and percent zero ownership can be produced and used to transform predicted average vehicle ownership values to the proportion of households without a vehicle available (Chatterjee *et. al.*, 1977). An alternative approach is to develop a structural model of the distribution of vehicle ownership to predict ownership levels as a function of average vehicle ownership in the district. A more sophisticated procedure for predicting automobile availability, which also has the advantage of producing information on the degree of competition for automobiles within households, is a two-step process which predicts the distribution of household size in terms of number of adults and, conditionally on this distribution, predicts the number of automobiles in each household.

Availability of public transportation is determined by access to the nearest public transportation stop. Estimates of maximum acceptable walk access time are obtained by examining existing data to determine the distance at which choice of the transit mode drops to zero or near zero in a variety of areas. This analysis of reasonable access distance should be undertaken separately for different linehaul transit modes, and should reflect use of different access modes. Past observations suggest that walk access to local transportation service is generally limited to ten to fifteen minutes walk time.

Predicting average values of independent variables for different classification groups requires considerable additional effort. This effort can be reduced by use of simplifying assumptions and simple relationships between class average values and overall average values observed in existing data sets. Variables which have little influence on the dependent variable or do not vary greatly between classification groups can be used at their overall average values. Other class-specific average variable values can be related to overall average variable values by adjustment factors developed from existing data. For example, when classification by drive-alone alternative availability is used, class values of income may be x% below average for the no-auto group and y% above average for the auto-available group. In the special case where classification is based on a selected variable value, the class average value can be derived directly from the classification process. That is, if classification is based on income, class values of income can be taken as the mid-point of the class range.

The pseudo-sample aggregation process requires generation of sample variable values for a set of "individuals". These sample values are obtained by Monte Carlo sampling from the distribution of variables. The procedure requires knowledge of the joint distribution of all variables in the prediction groups. This requirement can be simplified in a variety of ways. Variables which do not have important influence, and which are relatively invariant across the population can be valued at the population average. Variables can be assumed to be independently distributed unless there is strong reason to use joint or conditional distributions. Distributions for the remaining independent variables can be represented by relatively simple density functions. Income can be distributed according to a standard gamma distribution with only one parameter which can be determined from average income. Travel time can be represented by a uniform distribution with mean equal to the predicted average and range proportional to the size of the district. Alternatively, future distributions can be obtained by adjusting empirically estimated existing distributions to fit predicted mean values. For example, the existing distribution may be shifted to increase each value by the increase in the mean values, or the existing distribution may be factored by a constant radio equal to the proportional increase in the mean values.

DISCUSSION AND CONCLUSIONS

The use of disaggregate models has developed rapidly and extensively in recent years. Early research into the development of disaggregate models focused on improved understanding of travel choice behavior. More recent work has been directed at using disaggregate models in the analysis of practical planning issues.

An important advantage of the disaggregate modelling approach over the conventional aggregate models is the reduction in data required for model development. The large-scale home-interview surveys needed to develop aggregate models are not required. Furthermore, disaggregate models can be estimated without increased expense for the development of disaggregate level-of-service data. [7] Thus, substantial savings can be obtained in developing the data base needed for model development.

Even when data are already collected, it is preferable to use disaggregate choice models because of their improved statistical efficiency, transferability, behavioral structure and policy sensitivity. Enhanced behavioral structure and policy sensitivity provide the basis for improved understanding of travel behavior and improved evaluation of transport policy.

The transferability of disaggregate models holds the potential for major savings in data collection and model development costs. These savings result from the ability to "import" a model developed in a different geographic area, subject only to validation testing and adjustment or updating of selected model parameters. Such updating can be accomplished by use of limited data on local-area travel behavior (Atherton and Ben-Akiva, 1976). Corresponding aggregate models cannot be readily transferred between geographic areas because aggregate model parameters implicitly represent the distribution of independent variables in the estimation data. Thus, the aggregate model will not be valid in other areas or future time periods unless the underlying distribution of independent variables is unchanged over time or space.

These advangtages of disaggregate models can be obtained without incurring substantially increased cost or effort in prediction. This paper demonstrates the feasibility of making aggregate predictions using disaggregate models. The aggregation error using the naive procedure, which requires the same data and effort as conventional aggregate models, is likely to be smaller than errors from other sources. Furthermore, this aggregation error can be reduced substantially by use of limited information which can be obtained with little additional effort.

Analysis of alternative aggregation procedures demonstrates the improvements in predictive accuracy which can be obtained by use of information about difference in choice set availability when such differences exist. Alternative procedures for use of this additional information are described and evaluated. Methods to predict this additional information are also described. This material combined with what is already known about the estimation of disaggregate travel choice models provides a basis for application in a wide variety of policy and planning situations.

ACKNOWLEDGEMENTS

This paper is derived from a report prepared for the Project Bureau for Integral Traffic and Transportation Studies of the Netherlands Ministery of Transport by the authors, working for Cambridge Systematics, Inc. of Cambridge, Massachusetts, U.S.A., under sub-contract to Buro Goudappel en Coffeng B.V. of Deventer, The Netherlands. Earlier research in this area was supported by the U.S. Department of Transportation, Office of the Secretary, under contract DOT-OS-50001 for "Deve'opment of An Aggregate Model of Urbanized Area Travel Behavior".

REFERENCES

Atherton, T.J., and M. Ben-Akiva, "Transferability and Updating of Disaggregate Travel Demand Models", **Transportation Research Record,** 1976 (forthcoming).

Ben-Akiva, M.E., **Structure of Passenger Travel Demand Models,** Ph.D. dissertation, Massachusetts Institute of Technology (Civil Engineering), Cambridge, Massachusetts, 1973.

Ben-Akiva, M.E., and T.J. Atherton, "Choice Model Predictions of Carpool Demand: Methods and Results", **Transportation Research Record,** 1977 (forthcoming).

Cambridge Systematics, Inc., **Carpooling Incentives: Analysis of Transportation and Energy Impacts,** Federal Energy Administration, Washington, D.C., 1976.

Cambridge Systematics, Inc., **A Study Design for Dual Mode Transit Planning Case Studies,** Urban Mass Transportation Administration, U.S. Dept. of Transportation, 1975.

Charles River Associates (CRA), **A Disaggregate Behavioral Model of Urban Travel Demand,** Federal Highway Administration, U.S. Department of Transportation, Washington, D.C., 1972.

Chatterjee, A., S. Khasnabis and L.J. Slade, "Household Stratification Models for Travel Estimation", **Transportation Engineering Journal,** Vol. 103, American Society of Civil Engineers, New York, 1977.

Duguay, G., W. Jung and D. McFadden, "SYNSAM: A Methodology for Synthesizing Household Transportation Survey Data", Travel Demand Forecasting Project, Institute of Transportation Studies, University of California, Berkeley, 1976.

Green, H.A.J., **Aggregation in Economic Analysis,** Princeton University Press, Princeton, N.J., 1964.

Kanafani, A., "An Aggregative Model of Trip Making", **Transportation Research,** Vol. 6, 1972.

Koppelman, F.S., "Prediction with Disaggregate Models: The Aggregation Issue", **Transportation Research Record,** Number 527, Transportation Research Board, Washington, D.C., 1974.

Koppelman, F.S., **Travel Prediction with Models of Individual Choice Behavior,** Ph.D. dissertation, Massachusetts Institute of Technology (Civil Engineering), Cambridge, Massachusetts, 1975.

Koppelman, F.S., "Guidelines for Aggregate Travel Prediction Using Disaggregate Choice Models", **55th Annual Meeting,** Transportation Research Board, Washington, D.C., 1976 (a).

Koppelman, F.S., "Methodology for Analysis of Errors in Prediction with Disaggregate Choice Models", **55th Annual Meeting,** Transportation Research Board, Washington, D.C., 1976 (b).

Koppelman, F.S., M.E. Ben-Akiva and T. Watanatada, **Development of an Aggregate Model of Urbanized Area Travel Behavior,** Phase I Report, Assistant Secretary for Policy, Plans and International Affairs and Federal Highway Administration, U.S. Dept. of Transportation, 1976.

Lerman, S.R., and M.E. Ben-Akiva, "A Disaggregate Behavioral Model of Automobile Ownership", Transportation Research Record 569, Transportation Research Board, Washington, D.C., 1975.

Peat, Marwick, Mitchell & Co. (PMM), **Implementation of the N-Dimensional Logit Model: Final Report,** prepared for the Comprehensive Planning Organization, San Diego County, California, 1973.

Richards, M.G., and M.E. Ben-Akiva, **A Disaggregate Travel Demand Model,** D.C. Heath, Lexington, Massachusetts, 1975.

Ruiter, E.R., and M. Ben-Akiva, "A System of Disaggregate Travel Demand Models: Structure, Component Models, and

Application Procedures", Cambridge Systematics, Inc., Cambridge, Mass., 1977.

Small, K.A., "Priority Lanes on Urban Radial Freeways: An Economic Simulation Model", **Transportation Research Record,** Transportation Research Board, Washington, D.C., 1977 (forthcoming).

Talvitie, A.P., "Aggregate Travel Demand Analysis with Disaggregate or Aggregate Travel Demand Models", **Proceedings,** Transportation Research Forum, 1973.

Train, K., "Work Trip Mode Split Models: An Empirical Exploration of Estimate Sensitivity to Model and Data Specification", Working Paper No. 7602, Urban Travel Demand Forecasting Project, Institute of Transportation Studies, University of California, Berkeley, Calif., 1976.

FOOTNOTES

[1] Kanafani (1974) provides an example of an aggregate demand function derived using equation (4).

[2] This result is shown in detail in Green (1974) and Koppelman (1974)

[3] Conventional aggregate models are subject to aggregation error both in the estimation of parameters and the prediction of future demand (Koppelman, 1975).

[4] The detailed study is reported by Koppelman (1975).

[5] See Koppelman, 1976b, for a discussion of error analysis in prediction.

[6] The importance of modal availability is further illustrated in a study of the impact on work mode shares pursuant to the establishment of an auto-restricted zone (ARZ) in the Washington, D.C. central business district. The change in drive-alone mode shares was predicted by the sample enumeration procedure as −2.2%. Corresponding direct incremental share predictions (see Ben-Akiva and Atherton, 1977) were −4.7% by the naive procedure and −2.5% by classification based on modal availability. Thus, the use of modal availability information substantially reduced aggregation error in this case.

[7] Train (1976) showed that disaggregate choice models estimated with aggregate (zonal) level-of-service data are not significantly different from those estimated with disaggregate data.