TRANSPORTATION CENTER

METHODOLOGY FOR ANALYSIS OF ERRORS IN PREDICTION WITH DISAGGREGATE CHOICE MODELS

by

Frank S. Koppelman
Associate Professor
Department of Civil Engineering

and

The Transportation Center
Northwestern University

NORTHWESTERN UNIVERSITY EVANSTON • ILLINOIS

METHODOLOGY FOR ANALYSIS OF ERRORS IN PREDICTION WITH DISAGGREGATE CHOICE MODELS

by

Frank S. Koppelman
Associate Professor
Department of Civil Engineering

and

The Transportation Center
Northwestern University

Submitted for Publication at

Transportation Research Board
Washington, D.C.
June, 1976

ACKNOWLEDGEMENTS

This research was supported by the U.S. Department of Transportation, Office of the Secretary, under contract DOT-OS-50001 for "Development of an Aggregate Model of Urbanized Area Travel Behavior." The empirical data used was provided by the Metropolitan Washington Council of Governments.

I wish to acknowledge the helpful comments and suggestions of Professors Moshe Ben-Akiva, and Steven Lerman of M.I.T.

METHODOLOGY FOR ANALYSIS OF ERRORS IN PREDICTION WITH DISAGGREGATE CHOICE MODELS

by

Frank S. Koppelman

Abstract

Predictions of future travel behavior and of the performance of alternative transportation systems is needed by transportation planners and decision makers to make judgments about the desirability of alternative transportation plans. The usefulness of predictions, and consequently of prediction methods, depends on their accuracy. The purpose of this paper is to develop a methodology for the analysis of errors in prediction with disaggregate choice models.

The paper describes the process by which disaggregate choice models are formulated and used for prediction. The sources of error in the model formulation and prediction process are identified. The interaction and propagation of these errors to the final prediction is analyzed.

A set of error measures is proposed for evaluating the performance of alternative prediction models. A strategy is developed for analysis of the source of different components of the total error. An empirical analysis of errors in the prediction of mode choice to work illustrates the use of this approach to evaluate the accuracy of a set of prediction models, identify major sources of error in prediction and suggest steps which can be taken to improve these prediction models.

METHODOLOGY FOR ANALYSIS OF ERRORS IN PREDICTION WITH DISAGGREGATE CHOICE MODELS

by

Frank S. Koppelman

Introduction

Transportation planners and decision makers use predictions of expected travel behavior and transportation system performance to evaluate alternative transportation plans. The usefulness of predictions is directly dependent on their accuracy. Thus, an appropriate measure of the quality of prediction methods is the expected magnitude of errors in predictions which they produce. This paper identifies the primary sources of error in the travel demand prediction process, describes the way in which these errors contribute to total error in prediction, proposes a set of measures which may be used to evaluate the expected error of alternative prediction procedures and describes a strategy to identify the portions to total error attributable to different model components.

The paper is organized in five sections. The first section describes the model formulation and prediction process and its application to an aggregated prediction model. The second section identifies the sources of error in model formulation and prediction. The third describes the interaction and propagation of errors from different sources. The fourth section develops a set of error measures which can be used to measure the prediction accuracy of alternative model structures. The fifth section describes and demonstrates a method for analyzing the error in prediction attributable to different model components.

Model Formulation and Prediction Process

Predictions of future travel behavior are based on hypotheses about the factors which influence travel behavior and the structure of those influences. Possible hypotheses cover the range from simple "no change" and "time trend" predictions to relationships which describe the causal influence on travel behavior of changes in a wide range of socio-economic and transportation service characteristics.

The model formulation and prediction process carries the hypotheses through the steps of model specification, data collection, estimation of model parameters, and prediction of future travel behavior. The model structure used to represent the travel behavior process in the following discussion is a disaggregate model of individual choice behavior (McFadden, 1968; CRA, 1972) which is explicitly aggregated (Koppelman, 1976) to obtain group predictions.

The aggregated prediction model consists of three components:

- a disaggregate choice model
- a representation of the distribution of explanatory variables, and
- an aggregation procedure which operates on the two other components to obtain the required aggregate prediction.

The disaggregate choice model relates the probability of choosing an alternative out of a set of available alternatives to the estimated utility of each alternative for the individual decision maker. The utility of an alternative is defined in terms of the characteristics of the decision maker and the attributes of the alternative. The choice model may assume a variety of functional forms which are derived from the underlying assumptions about the individual's choice process (CRA, 1972).

The distribution of independent variables describes the presence in the aggregate prediction group of individuals with different socio-economic characteristics or facing different transportation service characteristics. That is, the distribution represents the frequency of occurrence in the prediction group of different values of the socio-economic and travel service variables which influence individual travel choice decisions.

The aggregation procedure operates on the disaggregate choice model and the distribution of independent variables to produce aggregate predictions. The theoretically consistent aggregation procedure determines the share of the prediction group expected to choose an alternative by averaging choice probabilities for all individuals in the prediction group. Because this approach requires prediction of the explanatory variables of each individual in the prediction group, a variety of alternative procedures with less extensive input data requirements have been proposed. These include (Koppelman, 1976):

- enumeration procedures which estimate expected shares by averaging the choice probabilities for a sample of the prediction group,
- summation/integration procedures which weight disaggregate choice probability estimates for different values of explanatory variables by the frequency of occurrence of these variables in the prediction group,
- statistical differentials which predict aggregate shares in terms of the moments of the distribution of explanatory variables,
- classification procedures which predict the expected choice shares for individual classes using average values of variables for each class and determine overall choice shares as a weighted average of the individual class choice shares, and

- the naive procedure which predicts the expected choice share by using average variable values for the entire prediction group.

The model formulation and prediction process describes the development and use in prediction of the aggregated prediction model. This process consists of the following steps (Figure 1).

- Specification of Disaggregate Travel Choice Mode based on the hypothesis that travel behavior represents an individual's choice response to the stimulus of a set of available alternative (CRA, 1972; Ben-Akiva, 1973). The specification includes selection of a functional form of the model and selection of variables to be included.
- Collection of data on individual Choice Behavior includes the characteristics of the individual, the choice available and the alternative selected by the individual. The data collected are determined by the variables included in the model specification. However, cost or other constraints of data collection may require modification, verification and estimation.
- The Distribution of Choice Influencing Variables is separately predicted or determined by policy selection to represent the characteristics of the prediction group and the alternatives which are available to them.
- The Aggregation Procedure applies the choice model to demographic and transportation service characteristics to predict aggregate travel behavior.

Sources of Error in Aggregate Prediction

Errors are introduced in each stage of the model formulation and prediction process (Figure 1). These errors are associated with the three major components of the aggregated prediction model structure described above.

Errors in the disaggregate choice model are the result of misspecification of the utility function and errors in the measurement of the independent variables. Manski (1973) classifies these errors in to four categories:

- Omitted structure - variables which should have been included in the utility function are excluded.
- Cross sectional preference variation - members of the sample group on which the choice function is calibrated have different parameters in their utility function.
- Instrumental variables - variables which should be included in the utility function have been replaced by other variables, and
- Imperfect information - the reported value of a variable is incorrect.

Errors associated with application of a model calibrated on one data set (collected in one area during one time period) for prediction in a different time or place are included in the category of specification errors. These errors of transferability are due to either omitted structure (failure to include variables which describe the difference between the two situations) or to cross sectional preference variations (different preferences exist in the two different populations). Transferability is an important characteristic of disaggregate models. Transferability depends on the ability to fully specify the relevant utility functions and the quality of the data used.

Atherton and Ben-Akiva (1976) showed that a work trip mode choice model calibrated with Washington, D.C. data was not significantly different from a similar model estimated on data collected in New Bedford, Massachusetts and San Francisco.

Errors in the predicted distribution of independent variables are due to similar errors in specification and estimation of the models used to predict these distributions. These errors may also include random errors and bias errors. For the purpose of this paper, the models used to predict these distributions are considered to be independent of the travel modeling process. That is, the predictions of explanatory variables contain errors which are outside the control of the transportation analyst.

Errors in aggregation result from the use of approximate aggregation procedures to replace the theoretically consistent but impractical complete enumeration procedure described earlier. Errors in aggregation are deterministic. The errors introduced by approximate aggregate procedures are structural errors which cause bias in the predictions obtained.

Interaction and Propagation of Errors to Aggregate Prediction

Errors in the choice model and errors in variables interact to produce errors in the prediction of individual choice probabilities. These errors are propagated through the aggregation procedure to produce errors in aggregate prediction. The aggregation procedure also introduces error directly into the aggregate prediction (Figure 2).

The interaction and propagation of errors in estimated choice model parameters and predicted variables is determined by the formulation of the choice model and the aggregation procedures. This process is described for the binary choice logit model. Koppelman (1975) presents a general analysis of error propagation for the general multiple choice model structure. The binary logit choice model is represented by:

$$P_t = \frac{e^{X_t' b}}{1 + e^{X_t' b}} \qquad (1)$$

where P_t is the probability of individual t choosing an alternative, and

$X_t' b$ represents the net utility of one alternative over the other as a linear additive function of variables, X_t, weighted by choice model parameters, b.

The first step in the propagation of errors is from the errors in model parameters to the error in individual choice probabilities. This propagation of random errors may be expressed approximately (Kendall and Stuart, 1969; Tukey, 1957) by:

$$EV(P_t) = \left[P_t(1-P_t)\right]^2 \left[X_t' \, A \, X_t + b \, Y_t \, b'\right] \qquad (2)$$

where $EV(P_t)$ is the error variance in the probability estimate for individual t,

A is the variance-covariance matrix for errors in parameters, and

Y_t is the variance-covariance matrix for errors in variables for individual t.

The error variance in predicted choice probabilities for pairs of individuals are correlated since they have common errors in parameters (due to use of the same choice model) and may have correlated error in the prediction of variables due to use of a common prediction process. This relationship can be expressed in terms of the error covariance between pairs of predictions which is:

$$EC(P_t,P_{t'}) = \left[P_t(1-P_t)\right] \left[P_{t'}(1-P_t)\right] \left[X_t' AX_{t'} + bY_{tt'}b'\right] \quad (3)$$

where $EC(\ )$ is the error convariance in the probability estimate for individuals t and t', and

$Y_{tt'}$ is the covariance matrix for measurement errors in variables for individuals t and t'.

This information is used to estimate the error variance in share prediction due to errors in parameters and variables for the complete enumeration and naive procedures. The share prediction by the complete enumeration procedure is:

$$S = \frac{1}{T} \sum_t P_t \quad (4)$$

where S is the aggregate share of the group choosing the alternative,

T is the number of members of the group, and

$\sum_t$ indicates summation over all members of the group

The error variance in the aggregate share prediction by complete enumeration is a function of the error variance in the estimates of individual choice probabilities and the error covariance of choice probabilities for each pair of individuals in the prediction group:

$$EV(S) = \frac{1}{T^2} \left[\sum_t EV(P_t) + \sum_t \sum_{t' \neq t} EC(P_t, P_{t'}) \right] \tag{5}$$

When individuals in the prediction group are relatively homogeneous with respect to variable values and error in variables, the error variance in aggregate shares may be expressed in terms of error variance in parameters, error variance in variables and error covariance in variables for pairs of individuals by:

$$EV(S) = \left[P(1-P) \right]^2 \left[\bar{X}' \, A \, \bar{X} + \frac{1}{T} b\bar{Y}_t b' + \frac{T-1}{T} b\bar{Y}_{tt'} b' \right] \tag{6}$$

where P is the probability estimate of the average individual in the prediction group,

$\bar{X}$ is the average variable vector for the prediction group,

$\bar{Y}_t$ is the error variance in variables for a representative individual, and

$\bar{Y}_{tt'}$ is the error covariance in variables for a representative pair of individuals.

This expression illustrates the way in which error variance in parameters, error variance for individual variables and error covariance in variables for pairs of individuals effect error variance in share prediction. It also illustrates the effect of prediction group size on the relative importance of these different sources of error. The effect of errors in parameters is independent of group size. The effect of errors in variables os reduced with increases in group size. The effect of error covariance in variable estimates increases with increasing prediction group size.

The naive aggregation procedure is equivalent to using the individual choice model structure (equation 1) to estimate choice shares based on average variable values in the prediction group. That is:

$$S_N = \frac{e^{\overline{X}'b}}{1 + e^{\overline{X}'b}} \qquad (7)$$

where S_N is the predicted aggregate share by the naive method, and

$\overline{X}$ is a vector of average variable values.

The corresponding variance in share prediction by the naive procedure is

$$EV(S_N) = \left[P(1-P)\right]^2 \left[\overline{X}' A \, \overline{X} + b \, \overline{Y} b'\right] \qquad (8)$$

where $\overline{Y}$, the error variance in average variable values is given by:

$$\overline{Y} = \frac{1}{T^2} \left[\sum_t Y_t + \sum_t \sum_{t' \neq t} Y_{tt'}\right] \qquad (9)$$

When the error variance in individual variables are similar and the error covariance for pairs of individuals are similar, equation 9 can be substituted into equation 8 and simplified to give:

$$EV(S_N) = \left[P(1-P)\right]^2 \left[\overline{X}' A \, X + \frac{1}{T} b \, \overline{Y}_t b' + \frac{T-1}{T} b \, \overline{Y}_{tt'} b'\right] \qquad (10)$$

The equality between equations 6 and 10 indicates that the propagation of errors through these aggregation procedures is similar when there is a high degree of homogeneity in variable values and errors in variable values for the prediction group. Under similar conditions the propagation of random errors in parameters and variables by other aggregation procedures are also similar to those for the enumeration procedure.

The propagation of bias errors can be analyzed in a similar manner. The propagation of these errors to individual choice probabilities is:

$$B(P_t) = P_t(1-P_t)\; B(X'b)$$

$$= P_t(1-P_t) \left[\sum_k X_k\; B(b_k) + \sum_k B(X_k)\; b_k\right] \qquad (11)$$

where B( ) is bias of the term in brackets.

The bias in share prediction by the enumeration procedure due to bias in parameters and variables is:

$$B(S) = \frac{1}{T} \sum_t B(P_t)$$

$$= \frac{1}{T} \sum_t P_t(1-P_t) \left[\sum_k X_{kt}\; B(b_k) + \sum_k B(X_{kt})\; b_k\right] \qquad (12)$$

The bias from different sources may be additive or offsetting depending on the direction of the biases and the sign of the corresponding variable for bias in parameters or the corresponding parameter for bias in variables. When members of the prediction group have similar variable values and biases, the bias equation may be simplified to:

$$B(S) = P(1-P) \left[\sum_k \overline{X}_k\; B(b_k) + \sum_k B(\overline{X}_k)\; b_k\right] \qquad (13)$$

The bias error due to bias in parameters and variables for the naive method is identical to that given in equation 13 for the enumeration method with relatively homogeneous groups. Under similar conditions the propagation of bias error in parameters and variables by different aggregation procedures also is similar to that for the enumeration procedure.

Thus, for relatively homogeneous prediction groups, the effect of errors in parameters and variables on error in share predictions is essentially independent of the aggregation procedure used. However, as within group variance increases, differences in error propagation also increase. The magnitude of differences in error propagation is much smaller than the magnitude of the propagated errors themselves except when the prediction group is very diverse and is located at or near the region of maximum curvature in the choice function.

Errors of Approximate Aggregation Procedures

Approximate aggregation procedures create errors in aggregate prediction in two ways. First, as already described, the propagation of parameter and variable errors to share prediction may be differentially affected by different aggregation procedures. Second, approximate aggregation procedures introduce structural bias into the aggregate prediction. The magnitude and direction of the structural bias depends on the type of aggregation procedure used, the distribution of independent variables in the prediction group and the curvature of the choice function at the point of prediction. These are the same factors which determine differences in error propagation. The structural aggregation bias may appear to have a random component due to unobserved differences in location on the choice function (which determines the curvature of the choice function), and the shape and variance of the distribution of variables in the prediction group. A detailed description of the aggregation bias introduced by different aggregation procedures is given by Koppelman (1975).

The variation in the magnitude of aggregation bias and differences in error propagation with changes in the prediction situation (location on the choice curve, distribution of independent variables, etc.) indicates the need to characterize the prediction situation in order to evaluate the probable magnitude of aggregation bias (Koppelman, 1976).

Error Measures for Evaluation of Prediction Models

The accuracy of different prediction models can be expressed in terms of the expected error of predictions made using the prediction model. The purpose of this section is to identify an error measure which describes the expected error in a prediction model for different prediction situations.

Two decisions must be made in the development of a suitable error measure. The first is to decide how to express the error which occurs in a single prediction. The second is to decide how to aggregate the errors from single predictions to some average or expected error for a group of predictions using a common prediction methodology.

The error measure chosen to describe the error in each prediction, the basic error measure, is defined by:

$$BEM_w = \left(\frac{P_w - A_w}{P_w} \right) \qquad (14)$$

where BEM_w is the basic error measure in prediction per unit of prediction for element w,

P_w is the predicted value for element w,

A_w is the actual value for element w.

This error measure expresses the magnitude of the error as a proportion of the magnitude of prediction. It is free of the dimensions of prediction and thus allows comparability among errors for predictions espressed in different terms.

The overall error measures, based on the use of a quadratic loss function, implies that (1) the importance of an error is proportional to the square of its magnitude and (2) positive and negative errors are treated alike. The resultant measure is the root mean square error (Theil, 1966) which is defined by:

$$RMSE = \left[\frac{1}{N} \sum_{w} BEM_w^2\right]^{1/2} \tag{15}$$

where RMSE is the root mean square error, and

N is the number of predictions over which the measure is determined.

The individual error measures can be weighted to reflect their relative importance. A useful characteristic of this error measure is that it can be disaggregated into average error, AE, and standard deviation of the error, SDE, which are defined by:

$$AE = \frac{1}{N} \sum_{w} BEM_w \tag{16}$$

and

$$SDE = \left[\frac{1}{N_w} \sum_{w} (BEM_w - AE)^2\right]^{1/2} \tag{17}$$

The relationship among these error measures is:

$$RMSE^2 = AE^2 + SDE^2. \tag{18}$$

The separation of the average and standard error portion of the expected error is important as there different errors indicate different deficiencies in the model formulation and prediction process.

Analysis of Errors in Prediction

The preceding section described an error measure which expresses the expected accuracy of a prediction model. The sources of inaccuracy, their interaction and their propagation to the final prediction were described earlier. The purpose of this section is to develop and demonstrate a method for identifying the portion of total error which is contributed by each of the components of the aggregated prediction model in an applied prediction context.

Identification of the contribution to total error of the different components of the model structure provides an indication of which components need to be improved or replaced. This analysis also puts the errors contributed by each model component in perspective with respect to errors constributed by other components. The disaggregation of errors requires an analysis procedure which identifies the separate components of error according to their source and whether they are due to average errors or standard deviation errors.

The analytic approach is to make multiple aggregate predictions of choice shares with a single disaggregate choice model and set of predicted choice variables but using different aggregation procedures. The prediction error resulting from use of the enumeration procedure, which includes no aggregation error, is determined by comparison of the predicted choice shares against the observed choice shares adjusted for error in observed shares

(Cole, 1969). The additional error due to aggregation bias is determined by comparison of the predicted shares by the selected aggregation procedure against the predicted shares by the complete enumeration procedure.

The comparison between sets of predictions and observed shares is conducted in two prediction contexts to allow analysis of error due to transferability as well as errors in aggregation and non-transfer model errors. First, predictions are made of travel choice shares for the data set on which the choice model is estimated. Considering the choice model to be well specified, the only model errors are stochastic variation errors in parameter estimates. The input variables used are obtained from the observed data set and are considered to be accurate. This is equivalent to making "ex post" predictions which are suitable for the analysis of the performance of the prediction model (Klein, 1968). Different aggregation procedures can be used in conjunction with this choice model and data base. The error in the choice model is obtained by comparison of predictions by the enumeration method to observed shares. These predictions include no aggregation error. Aggregation error can be obtained by comparison of predictions by an approximate aggregation procedure with the corresponding predictions by the enumeration procedure. Combined error in prediction is the square root of the sum of squared model errors and aggregation errors.

Second, predictions are made for a different data set from that on which the models are estimated. In this case, errors of model specification which affect transferability are included. The error in the choice model including specification error affecting transferability is obtained by comparison of predictions by the enumeration method to observed shares. The model error affecting transferability can be isolated from other model error based on the assumption that non-transfer model error is the same as the model error for prediction with the estimation data set (except for adjustment for differences in the average size of prediction groups as indicated in equation 6). Aggregation error and combined error can be analyzed as described for the estimation data set.

Figure 3, Comparative Prediction Test, identifies the types of error included in the different sets of predictions. Analysis of differences between sets of predictions and between individual prediction sets and observed shares will be used to identify errors from:

- The choice mode,
- Error which affects transferability, and
- Aggregation error.

The method of error analysis is illustrated by an empirical study of mode choice prediction for work trips to the CBD in the Washington, D.C. metropolitan area. Two subsets of data were created. The first includes 874 work trips from 17 districts in the District of Columbia and Maryland. The second includes 486 work trips from 12 districts in Virginia. A three-mode logit choice model (drive-alone, shared ride, transit ride) was specified and estimated using the first data set. The model specification and parameter estimates are shown in Table 1.

Aggregate share predictions were made for each of the districts in both data sets by the enumeration, naive and classification procedures. The enumeration and naive procedures were described earlier. They are, respectively, the average of the individual choice probabilities (equation 4) and the probability of choice for average socioeconomic characteristics and level of service attributes (equation 7). The classification procedure consists of classifying each prediction group in terms of choice set availability (individuals without drivers license or with no car available to the household do not have the drive alone alternative), using the naive procedure to predict choice share for each class and taking the weighted average of choice shares for the classes as the group share prediction. Errors in the prediction for each district are summarized in terms of average error, standard deviation of error and root mean square error according to equation 15 to 17.

Table 2, Errors in Prediction by District for Estimation Data Set, presents the error measures obtained. Average error, standard deviation of error and root mean square error are given for

- model error, determined by comparison between observed shares and predicted shares by the enumeration procedure,
- aggregation error, obtained by comparison of predicted shares by each method and predicted shares by the enumeration procedure, and
- combined error, which is obtained by combining model error and aggregation error.

This table indicates

1. Aggregation error for the naive and classification procedures is small compared to model error,
2. Aggregation error by the classification procedure is substantially smaller than by the naive procedure, and
3. The effect of aggregation error on combined error is substantially smaller than the aggregation error itself.

Table 3, Errors in Prediction by District for Alternative Data Set, presents error measures for prediction of mode shares for a geographically distinct (in terms of home base location) data set. These error measures are the same as those used in Table 2 except that model error is disaggregated into:

- non-transfer model error, obtained by adjusting model error estimated in the previous case (no transfer error) for differences in size of the prediction group, and
- transfer model error, obtained by adjusting model error for non-transfer model error.

This table indicates:

1. Model error and observed share error is larger than for the estimation data set
   a. Increased non-transfer model error (due to smaller prediction sample size) accounts for part of this increase,
   b. Transfer error accounts for a part of this increase.
2. Transfer model error and non-transfer model error are of similar magnitude (based on root mean square error) and total model error is substantially larger than for prediction with the estimated data set (24% vs. 16%).
3. Transfer model error includes an average component as well as a random component,
4. Aggregation error for the naive and classification procedures is similar in magnitude as for prediction with the estimation data set, and
5. The effect of aggregation error on combined error is substantially smaller than the aggregation error itself.

The overall results obtained indicate that aggregation error by the naive and classification by choice set procedures is small compared to model error. In addition, total error in prediction using these aggregation procedures is similar in magnitude to the error in observed shares based on samples of 40 to 50 observations per aggregate group. (Expected observed share errors based on samples of 40 to 50 observations per prediction group are about 20-25% of prediction values.

These results indicate that

- aggregate share prediction based on disaggregate choice models is relatively accurate as compared to sampling,
- errors due to model specification may be more important than errors due to aggregation.

These results further suggest that continuing emphasis be placed on prediction with disaggregate models and that particular effort be addressed toward improving the specification of the underlying choice model. Furthermore, the preceding analysis demonstrates the feasibility of using the proposed methodology to analyze prediction errors to evaluate alternative prediction methods and to diagnose sources of prediction error.

Summary

The purpose of this paper is to develop an approach to the analysis of errors in prediction. The sources of error in prediction are identified as coming from different elements of the model formulation and prediction process. The types of errors generated in each of these elements are described.

The process by which errors enter, interact with one another and are propagated to the final prediction is analyzed. The analysis indicates that the propagation of random and bias errors in model parameters and explanatory variable values to errors in aggregate share prediction are relatively independent of the method of aggregation used.

A method of analysis is proposed for use in identifying the sources of total error in prediction by use of pairwise comparisons among predictions by different methods and between these predictions and observed shares in the data set. An empirical application of the method of analysis is used to demonstrate the feasibility of using this approach to evaluate alternative prediction methods and to diagnose areas of potential improvement in the prediction process.

REFERENCES

Atherton, Terry J. and Moshe E. Ben-Akiva, "Transferability and Updating of Disaggregate Travel Demand Models", 55th Annual Meeting, Transportation Research Board, 1976.

Ben-Akiva, Moshe E., Structure of Passenger Travel Demand Models, Ph.D. Thesis, Department of Civil Engineering, M.I.T., 1973.

Charles River Associates, A Disaggregated Behavioral Model of Urban Travel Demands, prepared for U.S.D.O.T., FHWA, 1972.

Cole, Rosanne, "Data Errors and Forecasting Accuracy," Economic Forecasts and Expectations, (Mincer, J., ed.), Columbia University Press, 1969.

Kendall, Maurice G. and Alan Stuart, The Advanced Theory of Statistics, Vol. I of III, Third Edition, Griffin and Co., London, 1969.

Klein, L.R., An Essay on the Theory of Economic Prediction, The Academic Bookstore, Helsinki, Finland, 1968.

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

Koppelman, Frank S., Travel Prediction with Models of Individual Choice Behavior, CTS Report No. 75-7, Center for Transportation Studies, M.I.T., 1975.

Manski, Charles F., The Analysis of Qualitative Choice, Ph.D. Dissertation, Department of Economics, M.I.T., Cambridge, Mass., June, 1973.

McFadden, Danial, "the Revealed Preferences of a Government Bureaucracy," Technical Report No. 17, Institute of International Studies, University of California, Berkeley, 1968.

Theil, Henri, Applied Economic Forecasting, North Holland Publishing Company, Amsterdam, 1966.

Tukey, J.W., "The Propagation of Errors, Fluctuations and Tolerances: Basic Generalized Formulas," Tech. Report No. 10, Statistical Techniques Group, Department of Mathematics, Princeton University, 1957.

FIGURE 1

MODEL FORMULATION AND PREDICTION

WITH AGGREGATED DISAGGREGATE MODEL

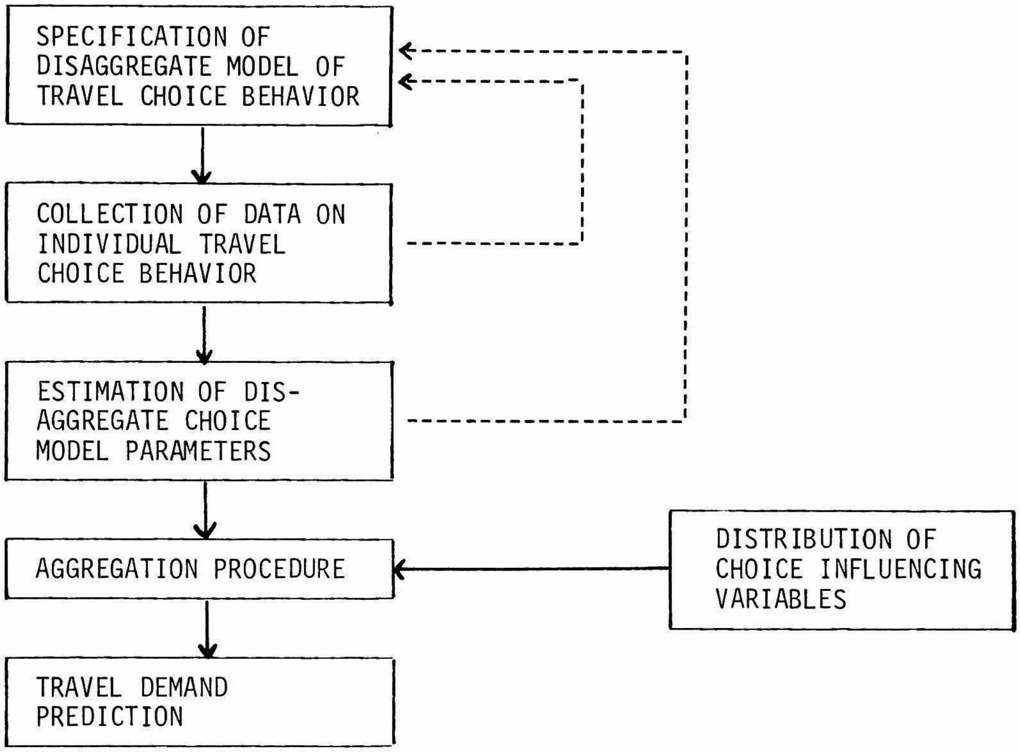

FIGURE 2

INTERACTION AND PROPAGATION OF ERROR

TO AGGREGATE PREDICTIONS

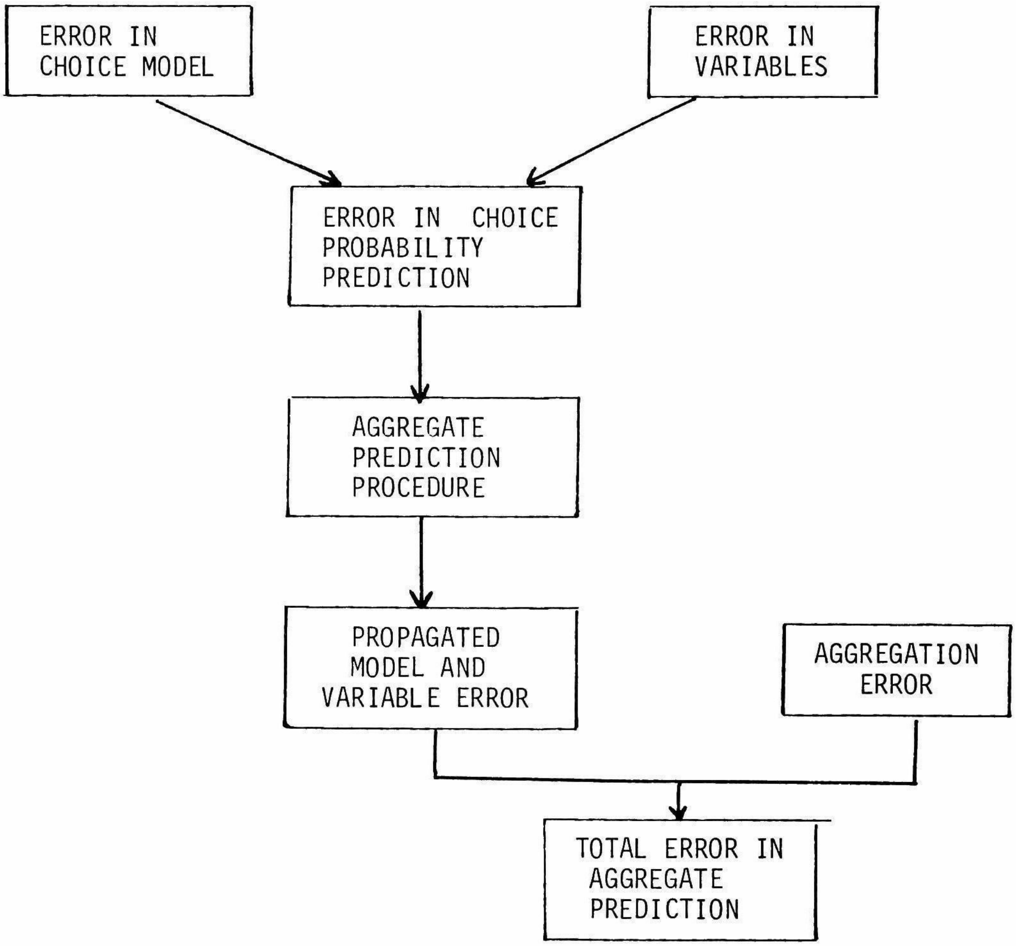

FIGURE 3

COMPARATIVE PREDICTION TESTS

	PERFECT AGGREGATION PROCEDURE	APPROXIMATE AGGREGATION PROCEDURES
CALIBRATION DATA SET	ERROR IN CHOICE MODEL	ERROR IN CHOICE MODEL AND AGGREGATION BIAS
ALTERNATIVE DATA SETS	ERROR IN CHOICE MODEL AND TRANSFER ERROR	ERROR IN CHOICE MODEL, TRANSFER ERROR AND AGGREGATION BIAS

TABLE 1

MODEL SPECIFICATION AND ESTIMATION

Variable	Symbol	Estimated Coefficient	Standard Error
1. Drive Alone dummy	D_d	-2.62	0.36
2. Shared Ride dummy	D_s	-2.36	0.27
3. Autos per licensed driver (Drive Alone)	$AALD_d$	3.64	0.38
4. Autos per licensed driver (Shared Ride)	$AALD_s$	1.51	0.24
5. Out-of-vehicle cost/income	OPTC/INC	-0.028	0.012
6. Total travel time	TTT	-0.024	0.005
7. Out-of-vehicle time/ distance	OVTT/DIST	-0.077	0.055
8. Government worker (Shared Ride)	GW_s	0.77	0.16
9. Number of workers in house-hold (Shared Ride)	$NWORK_s$	0.24	0.10

Note: Estimation based in 874 observations. 621 observations included all choice alternatives. 253 observations included only the shared ride and transit alternatives.

TABLE 2

PERCENT ERRORS IN PREDICTION BY DISTRICT FOR ESTIMATION DATA SET

PREDICTION PROCEDURE	ERROR TYPE	CATEGORY OF ERROR		
		MODEL ERROR	AGGREGATION ERROR	COMBINED ERROR
ENUMERATION	Average	0	0	0
	Standard Deviation	15.9	0	15.9
	Root Mean Square	15.9	0	15.9
NAIVE	Average	0	6.2	6.2
	Standard Deviation	15.9	4.7	16.6
	Root Mean Square	15.9	7.7	17.7
CLASSIFICATION	Average	0	1.3	1.3
	Standard Deviation	15.9	3.0	16.2
	Root Mean Square	15.9	3.3	16.2

TABLE 3

PERCENT ERRORS IN PREDICTION BY DISTRICTS FOR

ALTERNATIVE DATA SET

PREDICTION PROCEDURE	ERROR TYPE	CATEGORY OF ERROR				
		MODEL ERROR	NON-TRANSFER MODEL ERROR	TRANSFER MODERL ERROR	AGGREGATION ERROR	COMBINED ERROR
ENUMERATION	Average	13.0	0	13.0	0	13.0
	Standard Deviation	20.6	16.8	11.9	0	20.6
	Root Mean Square	24.4	16.8	17.7	0	24.4
NAIVE	Average	13.0	0	13.0	6.9	14.7
	Standard Deviation	20.6	16.8	11.9	4.6	21.1
	Root Mean Square	24.4	16.8	17.7	8.2	25.7
CLASSIFICATION	Average	13.0	0	13.0	1.4	13.1
	Standard Deviation	20.6	16.8	11.9	3.2	20.9
	Root Mean Square	24.4	16.8	17.7	3.5	24.7